

18:4 BERKELEY TECHNOLOGY LAW JOURNAL

2003

Pages
945
to
1360

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 © 2003 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 18

NUMBER 4

FALL 2003

TABLE OF CONTENTS

ARTICLES

ELECTRONIC EVIDENCE COMPLIANCE—A GUIDE FOR INTERNET SERVICE PROVIDERS..	945
By The U.S. Internet Service Provider Association	
THE BUSINESS METHOD PATENT MYTH.....	987
By John R. Allison & Emerson H. Tiller	
ENDING PATENT LAW’S WILLFULNESS GAME	1085
By Mark A. Lemley & Ragesh K. Tangri	
THE INFORMATION SEMICOMMONS.....	1127
By Robert A. Heverly	
A RETURN TO LILLIPUT: THE <i>LICRA V. YAHOO!</i> CASE AND THE REGULATION OF ONLINE CONTENT IN THE WORLD MARKET.....	1191
By Marc H. Greenberg	
CYBERSPACE TECHNOLOGICAL STANDARDIZATION: AN INSTITUTIONAL THEORY RETROSPECTIVE	1259
By Daniel Benoliel	
PATENTS AND PATH DEPENDENCY: A COMMENT ON BURK & LEMLEY	1341
By R. Polk Wagner	

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

LATHAM & WATKINS
San Francisco, CA

FARELLA BRAUN + MARTEL LLP
San Francisco, CA

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

FENWICK & WEST LLP
Palo Alto, CA

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

GRAY CARY WARE & FREIDENRICH,
LLP
Palo Alto, CA

WEIL, GOTSHAL & MANGES LLP
Redwood Shores, CA

HELLER EHRMAN WHITE
& MCAULIFFE LLP
San Francisco, CA

WILSON SONSINI GOODRICH &
ROSATI
Palo Alto, CA

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

ALSCHULER GROSSMAN STEIN & KAHAN LLP <i>Los Angeles, CA</i>	MAYER, BROWN, ROWE & MAW <i>Palo Alto, CA</i>
BINGHAM MCCUTCHEN <i>San Francisco, CA</i>	MCDERMOTT, WILL & EMERY <i>Menlo Park, CA</i>
COVINGTON & BURLING <i>San Francisco, CA</i>	MORRISON & FOERSTER LLP <i>San Francisco, CA</i>
DAY CASEBEER MADRID & BATCHELDER LLP <i>Cupertino, CA</i>	O'MELVENY & MYERS LLP <i>San Francisco, CA</i>
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER LLP <i>Palo Alto, CA</i>	ORRICK, HERRINGTON & SUTCLIFFE LLP <i>San Francisco, CA</i>
FISH & RICHARDSON P.C. <i>Redwood City, CA</i>	PILLSBURY WINTHROP LLP <i>San Francisco, CA</i>
GIBSON, DUNN & CRUTCHER LLP <i>Palo Alto, CA</i>	SHEARMAN & STERLING <i>San Francisco, CA</i>
KIRKLAND & ELLIS <i>San Francisco, CA</i>	TOWNSEND AND TOWNSEND AND CREW LLP <i>San Francisco, CA</i>

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

BAKER & MCKENZIE <i>Palo Alto, CA</i>	KENYON & KENYON <i>San Jose, CA</i>
DEWEY BALLANTINE LLP <i>Palo Alto, CA</i>	KNOBBE MARTENS OLSON & BEAR <i>Newport Beach, CA</i>
FISH & NEAVE <i>Palo Alto, CA</i>	MANATT, PHELPS & PHILLIPS LLP <i>Palo Alto, CA</i>
HOWREY SIMON ARNOLD & WHITE LLP <i>Menlo Park, CA</i>	MUNGER, TOLLES & OLSON <i>San Francisco, CA</i>
IRELL & MANELLA LLP <i>Century City, CA</i>	VAN PELT & YI LLP <i>Cupertino, CA</i>
KEKER & VAN NEST LLP <i>San Francisco, CA</i>	

Comment Competition Prize Sponsor

COOLEY GODWARD LLP
San Francisco, 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.
Interim Dean &
Walter Perry Johnson Professor of Law
Boalt Hall School of Law
Berkeley, California

ROGER BOROVY
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

G. GERVAISE DAVIS III
Davis & Schroeder P.C.
Monterey, California

MARK A. LEMLEY
Elizabeth Josslyn Boalt Professor of Law
& Director of Berkeley Center for Law &
Technology
Boalt Hall School of Law
Berkeley, California

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

DIANE WILKINS SAVAGE
Cooley Godward LLP
Palo Alto, 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 18 BERKELEY TECH. L.J. ____ (2003).

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

2003-2004

Editor-in-Chief

TARRA ZYNDA

Managing Editor
MATTHEW C. STAPLES

Senior Article Editors
WENFANG CHEN
MICHELE GUSTAFSON
RYAN OWENS

Senior Executive Editor
AMALIE M. WEBER

Annual Review Editors
AARON BURSTEIN
WILL THOMAS DEVRIES

Submissions Editors
DAVID ALBAN
JEFF DANLEY

Production Editor
TITI NGUYEN

Symposium Editors
SEÁN PATRICK BUTLER
JOSEPH WIEDMAN

Article Editors

STEPHEN BURDICK
ALEX EATON-SALNERS
BRIAN PAUL GEARING
DANIEL HIGGS

MATTHEW HOLOHAN
JAE HONG LEE
JULIETA LERNER

NICK MARTINI
ELIZABETH MILES
MARC SHARP
RUOYU ROY WANG

Executive Editors

RAEHEL GROOM
TERESA HUANG
STEVEN KAM

ALICE KAO
JOE MARRA

KRISTOFFER MAYFIELD
JOSEPH TELTSER
SVETLANA VAS

Annual Review Advisors

AARON BURSTEIN
WILL THOMAS DEVRIES
KAREN FESSLER

C. ALAN FU
RYAN OWENS
BHANU SADASIVAN

ALBERT SIEBER
MATTHEW C. STAPLES
KATE WILLIAMS

MEMBERSHIP

2003-2004

Submissions Advisory Committee

ERIC BROXMEYER
ANDREA OTTOLIA
JOO-YOUN PARK

SAMUEL ROGOWAY
RICHARD RONALD

SUSAN A. RUSSELL
BRYAN WAHL
FELIX WU

Associate Editors

KARINA ABOLINA
TONY BROWN
MICHELLE CAI
DANIEL CASTRO
CARLOS CERDA

KRISTINA GROENNINGS
CHARLES HUSE
CHERYL KAHN
OLGA V.
KOTLYAREVSKAYA

KENDRA NIELSEN
JOO-YOUN PARK
SHERWIN SIY
FELIX WU
ELIZABETH ZIRKER

Members

DANNY ANDREWS
AMANJIT ARORA
DIANE BARKER
LILY BLUE
JEFFREY BOLES
LARISSA BURFORD
BRIAN W. CARVER*
JANETTE CATRON
CYNTHIA CHEN
JEFF CHENG
PATTY CHENG
YVONNE SHIAO CHIEN
EUGENE CHOO
WES DERRICK
CARRIE DUONG
ANDREA FREEMAN*
ALICIA GAMEZ*
J. RYAN GILFOIL
SERENA GREEN
GALEN HANCOCK
TOMOMI HARKEY

CHERYL KAWEZYA
HUTCHINSON
CAROL JOHNS
ANN L. KIM*
GRACE KIM
CHARLENE KON
ANDREW LEE
ANNA LEE
CHRISTEN LEE
JADE XIAOYU LIN
SUNNY LU
JENNIFER LYNCH
LIWEN MAH
AZRA MEDJEDOVIC*
ESTEBAN MENDOZA
JONATHAN MUKAI
ROBERT MUSSIG
KATIE NOLAN-STEVAUX
GREGORY NOVOTNY
KATHERINE A. OYAMA
AARON PERZANOWSKI

NATHAN PHILLIPS
MIGUEL RUIZ
CHRISTOPHER SANCHEZ
RANGANATH SUDARSHAN
GEMMA SUH
JOCELYN SUM
ADWAIT TALATHI
CHELSEA TANAKA-
DELGADO
NIDHI TANDON
NATALIA THURSTON
ROB VARTABEDIAN
DEREK WALTER
JANE WANG
TARA WHEATLAND*
TINA WONG
JOHNAS WRIGHT
DAVID C. YANG
JUNE YI
LILY ZHANG

*Denotes Recipient of
Outstanding Member Award

ELECTRONIC EVIDENCE COMPLIANCE—A GUIDE FOR INTERNET SERVICE PROVIDERS

Prepared by the U.S. Internet Service Provider Association†

ABSTRACT

This Guide provides general guidelines for Internet service provider compliance with law enforcement and national security evidence gathering authorities. It is not intended to constitute or be a substitute for legal advice provided to individual clients on the basis of particular facts. In light of the law’s complexity, Internet service providers should consult counsel regarding questions about the law.

TABLE OF CONTENTS

1	INTRODUCTION	947
2	WHICH ENTITIES ARE COVERED BY ECPA?.....	949
	2.1 Provider of “Electronic Communication” or “Remote Computing” Services	949
	2.2 What Constitutes an “Electronic Communication”?	950
3	NON-CONTENT CUSTOMER RECORDS.....	951
	3.1 Voluntary Disclosure to Non-Governmental Entities.....	951
	3.2 Disclosure to Government Entities.....	951
	3.2.1 <i>Government-Compelled Disclosure</i>	952
	3.2.1.1 Basic Subscriber Information	952
	3.2.1.2 Transaction and Other Account Records.....	954
	3.2.1.3 Presence of Officer Not Required	955
	3.2.2 <i>Voluntary Disclosure of Non-Content Customer Records to the Government</i>	956
4	PEN REGISTERS AND TRAP AND TRACE DEVICES	956
	4.1 Exceptions to Prohibition Against Installing Pen Registers and Trap and Trace Devices.....	957
	4.2 Law Enforcement’s General Ability to Use Pen Registers and Trap and Trace Devices.....	958
	4.3 Emergency Law Enforcement Use of Pen Registers and Trap and Trace Devices.....	958
	4.4 A Provider’s Obligations to Provide Assistance	958
5	VOLUNTARY AND COMPELLED DISCLOSURE OF STORED ELECTRONIC COMMUNICATIONS	960
	5.1 Voluntary Disclosure of Stored Electronic Communications.....	961
	5.2 Law Enforcement’s Ability to Compel Production of the Content of Stored Electronic Communications.....	963

© 2003 U.S. Internet Service Provider Association

† Refer to Part 13 for contact information.

5.3	A Provider's Obligations to Provide Assistance	964
5.4	Back-up Copies	964
5.5	Civil Requests for E-mail Content	964
6	INTERCEPTION OR DISCLOSURE OF ELECTRONIC COMMUNICATIONS	965
6.1	What Constitutes an "Interception"?	966
6.2	Exceptions to Prohibition Against Intercepting Electronic Communications.....	966
6.3	Exceptions to Prohibition Against Divulging Contents of Electronic Communications.....	967
6.4	Law Enforcement's Ability to Intercept Electronic Communications	968
	6.4.1 <i>Law Enforcement Interception of Electronic Communications in Emergency Situations</i>	968
6.5	A Provider's Obligation to Provide Assistance	969
7	PRESERVATION REQUESTS	969
7.1	Preservation.....	969
7.2	Nondisclosure.....	970
8	NATIONAL SECURITY INVESTIGATIONS	972
8.1	Customer Records	973
	8.1.1 <i>FISA Order for Business Records</i>	973
	8.1.2 <i>FISA Order for Physical Search</i>	973
	8.1.3 <i>Certification for Subscriber Information and Toll Records in Counter-Intelligence Investigations</i>	974
8.2	Pen Registers and Trap and Trace Devices	975
	8.2.1 <i>The Government's Ability to Use FISA Pen Register and Trap and Trace Surveillances</i>	975
	8.2.2 <i>Emergency Law Enforcement Use of Pen Register and Trap and Trace Surveillances</i>	976
	8.2.3 <i>A Provider's Obligations to Provide Assistance</i>	976
8.3	Interception or Disclosure of Communications.....	976
	8.3.1 <i>Government's Ability to Intercept Communications</i>	977
	8.3.1.1 "Roving" FISA Orders	977
	8.3.1.2 Attorney General's Certification to Intercept Communications without a Court Order.....	977
	8.3.2 <i>Government Interception of Communications in Emergency Situations</i>	978
	8.3.3 <i>A Provider's Obligations to Provide Assistance</i>	978
9	REIMBURSEMENT FOR COSTS	978
10	LIABILITY FOR ECPA VIOLATIONS.....	980
	10.1 Criminal Liability	980
	10.2 Civil Liability	980
	10.3 Service Provider Immunity	981
11	INTERNATIONAL JURISDICTIONAL ISSUES.....	981
12	INTERRELATION WITH STATE LAW	983
13	CONTACT INFORMATION	984
14	GLOSSARY	984

1 INTRODUCTION

Internet service providers (“ISPs”) are increasingly being asked to provide assistance to government agencies in both criminal and national security investigations. The types of assistance being requested can take many forms, including:

- a request for *non-content records* (for example, billing records or transactional records);
- a request to *preserve* certain records or information;
- a request to implement a *pen register or trap and trace* surveillance;
- a request for *stored* electronic communications (for example, e-mail messages); or
- a request to *wiretap* a subscriber’s communications.

Requests for assistance by the government are governed by a series of federal surveillance laws.¹ Assistance in criminal investigations² is governed by Title III of the Omnibus Crime Control and Safe Streets Act of 1968—better known as “Title III”³—and the Electronic Communications Privacy Act of 1986—known as “ECPA.”⁴ Assistance in national security investigations⁵ is governed by the Foreign Intelligence Surveillance Act of 1978—better known as “FISA.”⁶

Although a major purpose of these laws is to regulate how the government conducts electronic surveillance, these laws also impose obligations on private parties, including ISPs. This Guide is intended to provide an overview of the laws as they may apply to ISPs, especially after the

1. Most states have adopted surveillance laws that apply to state and local law enforcement agencies within their jurisdiction. These state laws generally follow the federal rules, although states have adopted more restrictive requirements in a few cases. As a result, this Guide focuses on the standard, federal rules.

2. See discussion *infra* Parts 8, 10.

3. Pub. L. No. 90-351, 82 Stat. 212 (1968) (codified as amended at 18 U.S.C. §§ 2510-2520 (2000 & Supp. 2003)).

4. Pub. L. No. 99-508, 100 Stat. 1848 (1986) (codified as amended at 18 U.S.C. §§ 2232, 2510-2521, 2701-2711, 3117, 3121-3126 (2000 & Supp. 2003)).

5. See discussion *infra* Part 8.

6. Pub. L. No. 95-511, 92 Stat. 1783 (1978) (codified as amended at 50 U.S.C. §§ 1801-1811, 1822-1829, 1841-1846 (2000 & Supp. 2003)).

passage of the USA PATRIOT Act⁷ anti-terrorism law in October of 2001 and the Homeland Security Act⁸ in November of 2002. It is not meant to be a substitute for advice from a lawyer familiar with this often obscure area of the law and with the particular facts of an individual case. If in doubt, consult your own attorney.

The court's interpretation of these provisions has been disjointed due in large part to the complexity of the statutes. While the recent amendments introduced by the USA PATRIOT Act and the Homeland Security Act clarify how Title III and ECPA apply to ISPs, it may still be beneficial to interpret the laws' implications to the Internet by drawing analogies to the "plain old telephone system" and the early precursors to the Internet.

As summarized in the following chart, the surveillance laws make two general distinctions in criminal investigations. First, they distinguish between 1) historical information and 2) information acquired in real time. Second, they distinguish between 1) non-content records (such as subscriber or transactional information) and 2) the content of specific communications.

7. Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT Act) Act of 2001, Pub. L. No. 107-56, 115 Stat. 272 (2001) (codified as amended in scattered sections of 18 U.S.C. and 50 U.S.C. (2000 & Supp. 2003)) [hereinafter USA PATRIOT Act].

8. Homeland Security Act of 2002, Pub. L. No. 107-296, 116 Stat. 2135 (2002) (codified as amended in scattered sections of 18 U.S.C. (2000 & Supp. 2003)).

	Historical Information	Real Time Acquisition
Non-Content Records (Subscriber or Transactional Data)	Part 3: Customer Records Requires subpoena (for basic subscriber information only), “section 2703(d)” court order (for other records), or consent.	Part 4: Pen Register and Trap and Trace Devices Requires pen register or trap and trace court order or consent
Content of Communications	Part 5: Stored Electronic Communications Generally requires warrant (for communications in storage less than 180 days), subpoena (more than 180 days), or consent	Part 6: Interception of Electronic Communications Generally requires Title III court order or consent

This Guide briefly discusses each of these categories beginning with non-content customer records (which are generally entitled to the least legal protection) and concluding with the real-time acquisition of the content of communications (which is entitled to the most legal protection). This Guide further discusses the government’s authority to ask ISPs to preserve information in Part 7 and national security investigations in Part 8.

In applying the principles set forth in this Guide, ISPs must take into account their own particular technical structure. Not every ISP uses the same technology; indeed, there can be radical differences in technology that have a substantial impact both on what the law requires and on what the ISP can actually do. Thus, both ISPs and law enforcement agencies should be wary of the notion that the capabilities and obligations of one ISP can be applied freely to other ISPs with different technical structures.

2 WHICH ENTITIES ARE COVERED BY ECPA?

2.1 Provider of “Electronic Communication” or “Remote Computing” Services

ECPA addresses two types of entities: 1) providers of “electronic communication service,” defined as “any service which provides to users

thereof the ability to send or receive wire or electronic communications,” including electronic mail services;⁹ and 2) providers of “remote computing service,” which is defined as “the provision to the public of computer storage or processing services.”¹⁰ An ISP may qualify as either a provider of “electronic communication service” or “remote computing service”—or both. Take, for example, the case of a single e-mail received by an ISP for one of its customers. Before the recipient opens the e-mail, the ISP is providing “electronic communication service” because it is providing the user “the ability to . . . receive . . . electronic communications.”¹¹ Once the e-mail has been opened, however, the electronic communication is complete. If the user does not immediately delete the e-mail, the ISP is now offering a “remote computer service.” It is providing “computer storage” of the opened e-mail for the user.¹²

2.2 What Constitutes an “Electronic Communication”?

For purposes of ECPA, “electronic communication” is defined broadly to encompass a wide range of technologies. Specifically, ECPA defines an “electronic communication” as “any transfer of signs, signals, writing, images, sounds, data, or intelligence of any nature transmitted in whole or in part by a wire, radio, electromagnetic, photoelectronic or photooptical system that affects interstate or foreign commerce[.]”¹³ In the legislative history, Congress specifically identified non-voice communications consisting solely of data, communications transmitted only by radio, electronic mail, digitized transmissions, and video teleconferences as electronic communications.¹⁴ While the statute sought to be comprehensive in the technologies it covered, it can be unclear what constitutes an “electronic communication” and what does not. In light of this ambiguity, ISPs should

9. 18 U.S.C. § 2510(15) (2000 & Supp. 2003).

10. 18 U.S.C. § 2711(2). Providers of electronic communication service are in turn divided into those who provide service to the public and those who do not, with different rights and responsibilities. For example, a company that runs an internal data network is a provider of electronic communication service to itself, but is not a provider “to the public.” By definition, the term “remote computing service” only applies to those who provide such service to the public.

11. 18 U.S.C. § 2510(15).

12. Messages that are sent to group message boards or private fora present a unique problem. It is often difficult, if not impossible to determine if all members of a group have read the message. Thus, for some members the ISP is merely providing storage while for others the ISP is still in the process of delivering the message. How to treat such group message boards thus presents difficult legal and technical questions.

13. 18 U.S.C. § 2510(12).

14. *See* S. REP. NO. 99-541, at 14 (1986), *reprinted in* 1986 U.S.C.C.A.N. 3555, 3568.

contact legal counsel when there is a question about whether a service could potentially qualify as an “electronic communication.”

3 NON-CONTENT CUSTOMER RECORDS

Customer records include historical, non-content records such as basic subscriber information that identifies a customer’s name and address and transactional information about the customer’s use of a service. ECPA sets forth different standards for disclosing such non-content records depending on whether a provider is disclosing the records to a non-governmental entity or the government.

3.1 Voluntary Disclosure to Non-Governmental Entities

As discussed below, a public provider of electronic communication service or remote computing service may generally not disclose the *contents* of customers’ electronic communications¹⁵ to any third party, whether governmental or non-governmental. However, ECPA does not in any manner restrict a public provider from disclosing *records* about a customer (for example, customer user names, other identification information, and transactional records related to the account) to any entity other than a governmental entity.¹⁶

Therefore, a provider is not restricted by ECPA from providing records to non-governmental entities. However, to better protect customer privacy, an ISP may set standards on its own for when records will be disclosed to non-governmental entities. Although disclosure of customer records is generally not restricted by ECPA, other legal limitations may apply. Disclosures that are not in accord with the ISP’s privacy policy, for example, may lead to private or Federal Trade Commission actions.

3.2 Disclosure to Government Entities

Generally, a public provider may only disclose customer records to the government either in response to compulsory process such as a subpoena or court order under § 2703 or pursuant to an exception, such as customer

15. *See infra* Parts 4, 6.

16. 18 U.S.C. §§ 2702(a)(3), (c)(6) (2000 & Supp. 2003). Although disclosure of customer records is generally not restricted by ECPA, it is important to emphasize that other legal limitations may exist. For example, the Federal Trade Commission has promised to pursue civil sanctions against companies that disclose customer records in violation of their stated privacy policies.

consent.¹⁷ Note that in the latter case (exceptions where compulsory process is not required), disclosure is generally permissive: that is, a provider may disclose the pertinent records, but is not required to do so.

The governmental entity obtaining the records or other customer information must reimburse the provider for its “reasonably necessary” costs for assembling and providing such information.¹⁸ Typically, a warrant or “section 2703(d)” court order will instruct the provider not to disclose the existence of the warrant or order.¹⁹

3.2.1 *Government-Compelled Disclosure*

When the government wishes to compel the production of customer non-content records, the type of legal process required depends on the type of records sought. As discussed below, non-content customer records fall into two general categories—basic and transactional.

3.2.1.1 Basic Subscriber Information

Basic information about customers is available in response to an administrative, grand jury, or trial subpoena.²⁰ Of course, a subpoena is not the only way the government may obtain basic information. The government may also opt to use any of the other forms of process (court order, warrant) discussed below.²¹ But subpoenas are easy for investigators to obtain. They require no judicial oversight and no particular showing in terms of evidence. Congress has expanded the category considerably.²² ECPA now mandates disclosure of the following basic information: “the (A) name; (B) address; (C) local and long distance telephone connection records, or records of session times and durations; (D) length of service (including start date) and types of service utilized; (E) telephone or instrument number or other subscriber number or identity, including any temporarily assigned network address; and (F) means and source of payment for such service (including any credit card or bank account number), of a subscriber . . . when the governmental entity uses an administrative

17. A little-used exception also allows the government to access, with a mere written request, the name, address, and place of business of a customer engaged in telemarketing fraud (as defined in 18 U.S.C. § 2325). 18 U.S.C. § 2703(c)(1)(D).

18. 18 U.S.C. § 2706(a).

19. 18 U.S.C. § 2705(b).

20. *See* 18 U.S.C. § 2703(c)(2).

21. *See id.* (noting that basic subscriber information may be compelled by “any means available under paragraph (1)” of 18 U.S.C. § 2703(c)).

22. *See* USA PATRIOT Act § 210 (codified as amended at 18 U.S.C. § 2703(c)(2) (2000 & Supp. 2003)).

subpoena authorized by a Federal or State statute or a Federal or State grand jury or trial subpoena.”²³

Generally the types of subscriber information the government is entitled to receive under a subpoena “relate to the identity of the subscriber, his relationship with his service provider and his basic session connection record.”²⁴ The list of subscriber information in 18 U.S.C. § 2703(c)(2) includes Internet Protocol (“IP”) address information²⁵ and billing information (billing name, address, credit card, and/or bank account number) if the service provider is charging the subscriber for services.²⁶

The last two items on the list are also the two most interesting and occasionally difficult provisions. First, the credit card number used by a subscriber to pay for Internet access is properly treated as basic information subject to subpoena. But when a credit card is used to buy other goods or services, its number is transactional information and may be obtained only with a court order under 18 U.S.C. § 2703(d), known as a “section 2703(d)” court order.²⁷

Second, there is a question whether the government may use a subpoena to obtain a list of all dynamically assigned IP addresses used by a subscriber. The statute calls for production of “any temporarily assigned

23. 18 U.S.C. § 2703(c)(2). Note that some states require ISPs to notify customers when they receive civil subpoenas for identifying information about the customer. *See, e.g.,* VA. CODE ANN. § 8.01-407.1(A)(3) (Michie Cum. Supp. 2003), available at <http://leg1.state.va.us/cgi-bin/legp504.exe?000+cod+8.01-407.1> (last visited on Nov. 15, 2003). California is also considering such a measure, the Internet Communications Protection Act. *See* A.B. 1143, 2003 Leg., 2003-04 Reg. Sess. (Cal. 2003).

24. ORIN KERR, DEP’T OF JUSTICE, SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS § III.C.1, at 90 (2002) [hereinafter KERR, SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS], at <http://www.usdoj.gov/criminal/cybercrime/s&smanual2002.pdf>; *see also* *Jessup-Morgan v. Am. Online, Inc.*, 20 F. Supp. 2d 1105, 1108 (E.D. Mich. 1998). The court in *Jessup-Morgan* stated that

18 U.S.C. § 2510 states that “‘contents’, when used with respect to any wire, oral, or electronic communication, includes any information concerning the substance, purport, or meaning of that communication,” [not information concerning the identity of the author of the communication]. 18 U.S.C. § 2510(8). The “content” of a communication is not at issue in this case. Disclosure of information identifying an AOL electronic communication account customer is at issue. In 18 U.S.C. § 2703(c)(1)(C) this identifying information is specifically acknowledged as separate from the “content” of electronic communications.

Id.

25. 18 U.S.C. §§ 2703(c)(2)(C), 2703(c)(2)(E).

26. *See* 18 U.S.C. § 2703(c)(2)(F).

27. *See infra* Part 3.2.1.2.

network address.”²⁸ The Department of Justice has correctly concluded that this “list does not include other, more extensive transaction-related records, such as logging information revealing the e-mail addresses of persons with whom a customer corresponded during a prior session” and that “these records include the IP address assigned by an Internet service provider to a customer *for a particular session*.”²⁹ A complete list of all dynamic IP addresses that have been assigned to a subscriber can be a large and sensitive body of data. The statute plainly allows disclosure of at least one such address, but it is not clear that it calls for disclosure of multiple addresses. For dial-up ISPs, in particular, one such address is usually sufficient to identify the subscriber, while a complete list may be more sensitive. For portals and free e-mail services, however, the IP address may be the only available method of identifying the subscriber accurately at any particular time, so a more complete list of IP addresses may be needed to provide a user’s identity. In short, this is an area of considerable uncertainty.

3.2.1.2 Transaction and Other Account Records

For all other non-content customer records, the government must obtain a “section 2703(d)” court order³⁰ or a search warrant.³¹ Examples of such customer records include transactional records, such as addresses of web sites visited by the customer and e-mail addresses of other individuals with whom the account holder has corresponded.³² In practice, the re-

28. 18 U.S.C. § 2703(c)(2)(E) (2000 & Supp. 2003).

29. KERR, SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS, *supra* note 24, § III.C.1, at 90 (emphasis added). While this issue has yet to be addressed by a court, a broader reading of the statute, permitting government entities to obtain a log record of session times and durations with a mere administrative subpoena would eviscerate 18 U.S.C. § 2703(d), which requires a court order for information not specifically listed in 18 U.S.C. § 2703(c)(2). Since 18 U.S.C. § 2703(c)(2)(E) specifies “any temporarily assigned network address,” with “address” in the singular, “temporarily” reasonably relating to a specific incident, and making no use of the term “any and all,” there is a strong argument for limiting the information produced in response to administrative subpoenas to a single IP address for identification purposes.

30. In order to obtain a “section 2703(d)” court order (also known as a “specific and articulable facts” order), the government must present to a court “specific and articulable facts showing that there are reasonable grounds to believe that [the specified records] are relevant and material to an ongoing criminal investigation.” 18 U.S.C. § 2703(d). This “specific and articulable facts” standard is a lower standard than the “probable cause” standard the government must show to obtain a warrant.

31. 18 U.S.C. § 2703(c)(1)(B).

32. KERR, SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS, *supra* note 24, § III.C.2, at 90-91.

quirement for a court order or warrant for this data occasionally trips up state and local law enforcement. Because of the sharp distinction drawn by the statute between subpoenas under § 2703(c) and court orders under § 2703(d), ISPs often refuse to provide transactional data in response to court orders that are labeled “subpoenas,” as may happen in some states. Law enforcement should recognize the risk that any document labeled a subpoena will be handled under § 2703(c) and not under § 2703(d), thus limiting the data available.

Similarly, some states issue court orders for discovery that do not meet the requirements of § 2703(d). These requirements include “specific and articulable facts showing that there are reasonable grounds to believe that the . . . records or other information sought, are relevant and material to an ongoing criminal investigation.”³³ A state court order that does not specify these findings will typically be rejected by cautious ISPs, or ISPs will limit disclosure to basic subscriber data.

3.2.1.3 Presence of Officer Not Required

A relatively new provision of the statute makes clear that a law enforcement officer need not be present during the service and execution of a search warrant for the purposes of obtaining customer records.³⁴ This provision overturns the lower court’s decision in *United States v. Bach*, which held that federal law requires a law enforcement officer who obtains a warrant to search a suspect’s e-mail account to be “present” when the ISP’s employees retrieve the information.³⁵ The decision has now been overruled twice—once by Congress as set forth above and again by the Eighth Circuit, which held that allowing ISP employees to conduct a search and seizure without the supervision of an officer was not a violation of the Fourth Amendment.³⁶ The provision allows ISPs to accept service of warrants by fax and to extract records without the distraction and privacy risk of an investigator observing the process.³⁷ Many ISPs do not permit investigators to participate while technicians are examining and extracting data from their systems.

33. 18 U.S.C. § 2703(d).

34. 18 U.S.C. § 2703(g) (2000 & Supp. 2003). This section reflects the amendment introduced by the 21st Century Department of Justice Appropriations Authorization Act, Pub. L. No. 107-273, § 11010, 116 Stat. 1812, 1822 (2002).

35. No. 01-221, 2001 U.S. Dist. LEXIS 21853, at *10 (D. Minn. Dec. 14, 2001).

36. *United States v. Bach*, 310 F.3d 1063, 1066 (8th Cir. 2002), *cert. denied*, 123 S. Ct. 1817 (2003).

37. 18 U.S.C. § 2703(g).

3.2.2 *Voluntary Disclosure of Non-Content Customer Records to the Government*

ECPA allows for the voluntary disclosure of a public provider's non-content subscriber information to the government in three circumstances:

- with the customer's consent;³⁸
- where the disclosure is necessary to protect the provider's rights or property;³⁹ and
- "if the provider reasonably believes that an emergency involving immediate danger of death or serious physical injury to any person justifies disclosure of the information."⁴⁰

Disclosure of the contents of subscriber communications in an emergency is discussed in Part 4.3, which should be read in conjunction with this Part. It is worth reiterating that under these provisions, service providers are under no obligation to disclose customer records to law enforcement entities absent legal process issued pursuant to § 2703. In short, a provider *may* disclose records to the government under these circumstances, but that decision is committed to the sole discretion of the service provider.

4 PEN REGISTERS AND TRAP AND TRACE DEVICES

Subject to certain exceptions discussed below, ECPA prohibits the installation of either pen registers or trap and trace devices without a court order.⁴¹

A pen register records the telephone numbers dialed on outgoing calls, while a trap and trace device records the telephone numbers identifying the origin of incoming calls. Neither mechanism records the *content* of a communication (which is covered by a more rigorous set of restrictions, discussed in Parts 5 and 6).⁴² In the context of e-mail communications, of course, the distinction between content and non-content is not as plain as in the telephone context. In addition to subject lines, which plainly contain

38. 18 U.S.C. § 2702(c)(2).

39. 18 U.S.C. § 2702(c)(3).

40. 18 U.S.C. § 2702(c)(4) (2000 & Supp. 2003). The emergency/voluntary disclosure amendments to ECPA, added by the USA PATRIOT Act, are currently set to expire on December 31, 2005. USA PATRIOT Act § 224 (codified as amended at 18 U.S.C. § 2510 note (2000 & Supp. 2003))

41. 18 U.S.C. § 3121(a).

42. 18 U.S.C. § 3127(3)-(4).

content, even the “To” and “From” lines can contain aliases and “Send As” names that may be considered content.

Law enforcement may also use pen register and trap and trace orders to trace communications on the Internet and other computer networks. Orders for the installation of a pen register or trap and trace device may obtain any prospective non-content information associated with communications—including all “dialing, routing, addressing, [and] signaling information”⁴³—utilized in the processing and transmitting of wire and electronic communications. Such information includes IP addresses and port numbers, as well as the “To” and “From” information contained in an e-mail header. Pen register and trap and trace orders cannot, however, authorize the interception of the content of a communication, such as words in the subject line or the body of an e-mail.⁴⁴

Further, because the pen register or trap and trace “device” often cannot be physically “attached” to the target facility, 18 U.S.C. § 3123 was recently amended by the USA PATRIOT Act to allow law enforcement agencies to use software instead of physical mechanisms to collect relevant pen register or trap and trace information.

4.1 Exceptions to Prohibition Against Installing Pen Registers and Trap and Trace Devices

The prohibition against installing pen registers and trap and trace devices does not apply to devices (or software) used by electronic communication service providers:

- in relation to the operation, maintenance, and testing of their communication service;
- to protect the rights or property of the provider (for example, to bill or to detect hacking);
- to protect users of the service from fraudulent, unlawful, or abusive use of the service; or
- with the consent of the user of the service.⁴⁵

43. 18 U.S.C. § 3127(3).

44. The Justice Department takes the view that the subject line of an e-mail constitutes *content* that may, along with the body of the e-mail, properly be captured only with a valid court order or other legal authority under the wiretap statute. KERR, SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS, *supra* note 24, § III.C.3, at 91.

45. 18 U.S.C. § 3121(b).

4.2 Law Enforcement's General Ability to Use Pen Registers and Trap and Trace Devices

Except in the very limited emergency situations set forth below, law enforcement must obtain a court order before it can lawfully install a pen register or trap and trace device.⁴⁶ An order may only authorize the installation and use of a pen register or trap and trace device for up to 60 days, but law enforcement may obtain an unlimited number of extensions, each of 60-day duration.⁴⁷ Because a court order is available to law enforcement based on its own certification (without independent evaluation by a court of the facts and circumstances supporting the application), the requirements for a pen register or trap and trace order are considerably less stringent than those for interceptions of the contents of electronic communications, discussed below.⁴⁸

4.3 Emergency Law Enforcement Use of Pen Registers and Trap and Trace Devices

Certain specially designated law enforcement officers may authorize the installation and use of a pen register or trap and trace device before obtaining a court order if they 1) determine that an emergency situation exists involving either an immediate danger of death or serious injury or organized crime activities, and 2) subsequently obtain a court order within 48 hours.⁴⁹ Even in this type of emergency situation, the law enforcement agency must stop using the pen register or trap and trace device after 48 hours (at the latest) if it subsequently fails to obtain a court order.⁵⁰

4.4 A Provider's Obligations to Provide Assistance

An order authorizing the installation of a pen register or trap and trace device may direct a provider to furnish the government "forthwith all information, facilities and technical assistance necessary to accomplish the installation."⁵¹ Law enforcement must reimburse a provider for its "rea-

46. 18 U.S.C. §§ 3121(a), 3125 (2000 & Supp. 2003).

47. 18 U.S.C. § 3123(c). Note, national security-related pen registers and trap and trace devices are installed under slightly different standards, pursuant to the Foreign Intelligence Surveillance Act of 1978 ("FISA"), Pub. L. No. 95-511, 92 Stat. 1783 (1978) (codified as amended at 50 U.S.C. §§ 1801-11, 1822-29, 1841-46). FISA-authorized pen registers and trap and trace devices can last up to 90 days, with renewals for up to 90 days at a time. 50 U.S.C. § 1842(e).

48. See *infra* Part 6.

49. 18 U.S.C. § 3125(a).

50. 18 U.S.C. § 3125(b).

51. 18 U.S.C. § 3124(a)-(b).

sonable expenses” in providing such assistance.⁵² As a general rule, the only assistance required in response to such an order is the assistance that can be provided by an ISP’s existing personnel and technology. The principal exception arises in the context of pen registers, where ISPs that cannot carry out the order using their own equipment may be required to install a device (such as Etherpeek or the FBI’s DCS1000) to collect the information. This is because, somewhat anomalously, the pen register provision speaks in terms of authorizing “the installation and use” of a “device.” But extracting data from a complex ISP network is far more complex than attaching one of the old phone line pen registers. Few ISPs are willing to permit the use of foreign equipment on their networks in order to preserve the proper functioning of these complex systems. As a result, ISPs tend to insist on developing their own pen register tools rather than allowing the installation of other equipment for all but the simplest networks.

When a pen register “device” is installed, the law enforcement agency is required to provide to the court under seal within 30 days: 1) the identity of the officers who installed or accessed the device; 2) the date and time the device was installed, accessed, and uninstalled; 3) the configuration of the device at installation and any modifications to that configuration; and 4) the information collected by the device.⁵³

Typically, an order will instruct the provider not to disclose the existence of the pen register or trap and trace device.

As a result of the passage of the USA PATRIOT Act, ECPA was amended to include two new provisions in 18 U.S.C. § 3123 that are worth highlighting. The first gives federal courts the authority to issue pen register and trap and trace orders effective outside the district of the issuing court.⁵⁴ The second provides that law enforcement authorities must file a special report with the court whenever they use a pen register or trap and trace order to install their own monitoring device (such as the FBI’s DCS1000) on computers belonging to a public provider.⁵⁵

As amended, 18 U.S.C. § 3123(a)(1) gives federal courts the authority to compel assistance from any provider of communication services in the United States. This allows a law enforcement agency to serve one order on multiple ISPs. Thus, a prosecutor’s application and the resulting order will not necessarily name all providers in a communications chain. This provi-

52. 18 U.S.C. § 3124(c) (2000 & Supp. 2003).

53. 18 U.S.C. § 3123(a)(3).

54. 18 U.S.C. § 3123(a)(1).

55. 18 U.S.C. § 3123(a)(3).

sion specifies that, when a provider requests it, law enforcement must provide a “written or electronic certification” that the order applies to that provider.

This section also empowers courts to authorize the installation and use of pen registers and trap and trace devices in other districts. Thus, for example, if a federal terrorism or other criminal investigation based in Virginia uncovers a conspirator using an Internet account in New York, the Virginia federal court can compel communications providers in New York to assist investigators in collecting information under its pen register or trap and trace order. Consistent with this, 18 U.S.C. § 3123(b)(1)(C) does not require that federal pen register or trap and trace orders specify their geographic limits. However, since the law gives out-of-district effect to federal pen register or trap and trace orders, 18 U.S.C. § 3127(2)(A) imposes a “nexus” requirement: the issuing court must have jurisdiction over the particular crime under investigation.

5 VOLUNTARY AND COMPELLED DISCLOSURE OF STORED ELECTRONIC COMMUNICATIONS

ECPA generally prohibits unauthorized access to and disclosure of the contents of electronic communications in *electronic storage*.⁵⁶ ECPA defines “electronic storage” as: “[A]ny temporary, intermediate storage of a wire or electronic communication incidental to the electronic transmission thereof; and (B) any storage of such communication by an electronic communication service for purposes of backup protection.”⁵⁷

18 U.S.C. § 2510(17) makes it clear that electronic communications in electronic storage includes unopened e-mail that is less than 180 days old.⁵⁸ E-mail that had been opened and retained on the service was thought not to be in electronic storage because it was not retained “incidental to

56. 18 U.S.C. § 2701(a).

57. 18 U.S.C. § 2510(17). The courts have not found the concept of storage easy to capture. In *United States v. Councilman*, the U.S. District Court of Massachusetts managed to characterize almost all electronic communications as “stored communications” rather than as communications in transit when it concluded that an electronic communication is in electronic storage even when it is in storage for a mere “nano-second juncture.” 245 F. Supp. 2d 319, 320-21 (D. Mass. 2003). If this view stands it would mean that virtually all electronic communications would be treated as stored (and thus subject to seizure with a search warrant) rather than in transit (and thus subject to seizure only with an intercept order). Seizure by intercept order is a far more onerous process.

58. See KERR, SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS, *supra* note 24, § III.B, at 86-87.

... transmission.”⁵⁹ However, in *Theofel v. Farey-Jones*,⁶⁰ the Ninth Circuit held that electronic communications that had been opened and were kept by the Internet service provider merely for the convenience of the customer were still in electronic storage because they were kept for purposes of backup protection and covered by 18 U.S.C. § 2510(17)(B). This has potentially far-reaching implications because governmental entities may obtain communications that are no longer in electronic storage by subpoena with notice, or by court order pursuant to 18 U.S.C. § 2703(d).⁶¹ The holding in *Theofel* means the government must get a search warrant in order to obtain the contents of any electronic communication that has been stored by an ISP for less than 180 days.

Whether the holding in *Theofel* will ultimately be proven correct is open to some question. The opinion was issued just before this Guide went to press and petitions for rehearing have been filed. Moreover, this is the first appellate interpretation of this particular provision and other circuits may come to different conclusions on the meaning of the language in the statute. However, for the present time this opinion is controlling in the Ninth Circuit and represents the only circuit court opinion interpreting 18 U.S.C. § 2510(17).

5.1 Voluntary Disclosure of Stored Electronic Communications

Subject to certain exceptions, providers of electronic communication service or remote computing service *to the public* are prohibited from knowingly divulging the contents of any customer communication (those who provide private communications and storage services are not so regulated).⁶² Disclosure of the contents of a communication is permitted, however, in the following circumstances:

- to an addressee or intended recipient of the communication (or his agent);
- as otherwise authorized by a court order or some other legal authorization;
- “with the lawful consent of the originator or an addressee or intended recipient of such communication or the subscriber in the case of a remote computing service;”

59. 18 U.S.C. § 2510(17)(A) (2000 & Supp. 2003).

60. 341 F.3d 978, 984-85 (9th Cir. 2003).

61. 18 U.S.C. § 2703(a)-(b).

62. 18 U.S.C. § 2702(a).

- for the purpose of forwarding a communication to its destination;
- if necessary to provide the service or to protect the rights or property of the provider;
- to law enforcement, “if the contents (i) were inadvertently obtained by the service provider; and (ii) appear to pertain to the commission of a crime;”
- “to a Federal, State, or local governmental entity, if the provider, in good faith believes that an emergency involving danger of death or serious physical injury to any person requires disclosure without delay of communications relating to the emergency[;]” or
- if the ISP has knowledge of child pornography, required by 42 U.S.C. § 13032 (where violation of child pornography laws is apparent).⁶³

For any other disclosures to the government, the government must obtain a warrant, court order, or subpoena, depending on the age or type of the communication, as described below.

Law enforcement agencies sometimes invoke the “emergency” provision in an effort to avoid the necessity of a subpoena or other process. ISPs often must be firm in pointing out that this provision gives the ISP, *not* law enforcement, authority to decide whether or not to provide information. There is never an “emergency” obligation on an ISP to disclose under § 2702(b)(7). In a true emergency, ISPs are usually ready to respond, but they typically insist that investigators provide a clear basis on which the ISP can conclude that an emergency meeting the statutory criteria exists. The law requires that the ISP reasonably “believe[] that an emergency involving danger of death or serious physical injury to any person requires disclosure [of the information] without delay”⁶⁴

Because of the intense interest of agencies in this exception, it is prudent for an ISP to adopt clear procedures for its use, and to require all government agencies to adhere to the procedures. Some ISPs provide forms to be filled out by investigators, sometimes under penalty of perjury. These forms can help to focus investigators by requiring the information needed to satisfy the statute (for example, “What is the danger of death or serious physical injury?” and “Is the danger immediate?”). Transmitted along with a cover letter on the letterhead of a law enforcement agency, such docu-

63. 18 U.S.C. § 2702(b).

64. 18 U.S.C. § 2702(b).

ment agency, such documentation provides a useful backup in case the ISP's decision to release information voluntarily is ever questioned.

5.2 Law Enforcement's Ability to Compel Production of the Content of Stored Electronic Communications

Although ECPA does not on its face draw a distinction between opened and unopened e-mail, this is a central dividing line in practice. ECPA provides heavy protection for communications while they are in the process of transmission (for example, an unopened e-mail).⁶⁵ As long as the e-mail remains in "electronic storage" (for example, unopened in an inbox), it is protected quite heavily for 180 days. After electronic communications in "electronic storage" have remained undelivered for 180 days, law enforcement may access the contents under less rigorous legal authorizations, including an administrative, grand jury, or trial subpoena.⁶⁶ But all of that depends on the communication remaining undelivered. Once an electronic communication has been opened by its recipient, it is no longer in transmission. The ISP is no longer providing "electronic communication service" but is now providing a "remote computing service." The communication is protected only as a stored communication and is accessible under the less rigorous legal authorizations of 18 U.S.C. § 2703(b).⁶⁷

When a governmental entity obtains the content of electronic communications that have been in electronic storage more than 180 days with process other than a search warrant, the government (not the ISP) is required by law to give prior notice to the customer.⁶⁸ However, the government may delay notification for up to 90 days if there is reason to believe that notification would result in: 1) risk to the life or physical safety of an individual; 2) flight from prosecution; 3) destruction of or tampering with evidence; 4) intimidation of potential witnesses; or 5) otherwise seri-

65. 18 U.S.C. § 2510(17) (2000 & Supp. 2003).

66. 18 U.S.C. § 2703(b). These requirements apply to a remote computing service provider to the extent that

any wire or electronic communication . . . is held or maintained on [its] service—(A) on behalf of, and received by means of an electronic transmission from (or created by means of computer processing of communications received by means of electronic transmission from), a subscriber or customer of such remote computing service.

18 U.S.C. § 2703(b)(2).

67. *See supra* Part 2.1.

68. 18 U.S.C. § 2703(b).

ously jeopardize an investigation.⁶⁹ The government may obtain subsequent extensions of this delay in providing notice.

5.3 A Provider's Obligations to Provide Assistance

A provider must comply with an appropriate legal authorization to provide the government with access to the stored communications identified in the order. The governmental entity obtaining the contents of such stored communications, however, must reimburse the provider for its "reasonably necessary" costs directly incurred for assembling and providing such information.⁷⁰

5.4 Back-up Copies

Under a very rarely used provision, if a governmental entity determines (at its discretion) that there is reason to believe that prior notification to the customer of the existence of a court order or subpoena may result in destruction of or tampering with stored communications, it may require that the provider create a back-up copy of the contents of the electronic communications at issue before such notice is provided to the customer.⁷¹

Without notifying the customer, the provider must create a back-up copy within two days of receipt of the order.⁷² The provider must not destroy this back-up copy until the later of: 1) its delivery to the governmental entity, or 2) the resolution of any proceedings concerning the court order or subpoena.⁷³ Note, unlike the preservation requirement⁷⁴ (which pre-dates a court order and applies to stored communications as well as customer records), the obligation to create back-up copies applies after a court order or subpoena is served and applies only to stored communications (not customer records).

5.5 Civil Requests for E-mail Content

Most ISPs receive discovery requests in civil matters on a routine basis. These requests may not be as numerous as criminal requests, but responding is often more complicated and time consuming. First, the sources of civil requests—federal enforcement agencies such as the Securities and Exchange Commission ("SEC") or the Federal Trade Commission

69. 18 U.S.C. § 2705(a)(1)-(2).

70. 18 U.S.C. § 2706(a).

71. 18 U.S.C. § 2704(a)(5) (2000 & Supp. 2003).

72. 18 U.S.C. § 2704(a)(1).

73. 18 U.S.C. § 2704(a)(3).

74. See discussion *infra* Part 7.

(“FTC”), and private litigants—are often not familiar with the technology or legal constraints related to subscriber data held by ISPs. Second, the legal framework that allows government entities access to subscriber data does not provide helpful guidance on the appropriate mechanisms for private entities to compel production of records. Third, unlike in the criminal context where ISPs are unlikely to provide notice to subscribers, in the civil context many ISPs seek to give their subscribers notice of requests for their account data. Indeed, in certain jurisdictions ISPs may be required to provide notice.⁷⁵

While ISPs are free, subject to their privacy policies, to provide identity and other non-content information to civil litigants (or, indeed, to other private parties), they are more restricted in divulging content. Section 2702(a) prohibits an ISP (either a provider of “electronic communication service” or a “remote computing service”) from divulging “the contents of a communication.” Section 2702(b) offers several exceptions to the prohibition, but none of them expressly permits disclosure pursuant to a civil discovery order unless the order is obtained by a government entity. This issue has not been litigated to our knowledge, but in some cases courts have managed to avoid the issue by ordering the subscriber to give consent to the disclosure of the contents of his or her e-mail, a compromise that only works when the subscriber is subject to the court’s jurisdiction. Thus, the federal prohibition against divulging e-mail contents remains stark, and there is no obvious exemption for a civil discovery order on behalf of a private party. ISPs can, of course, voluntarily preserve the contents of an account pending receipt of a court order, if such an order can be obtained.

6 INTERCEPTION OR DISCLOSURE OF ELECTRONIC COMMUNICATIONS

Subject to certain exceptions discussed below, ECPA prohibits the intentional “interception” of any electronic communication, or the intentional disclosure or use of the contents of an intercepted electronic communication.⁷⁶ This prohibition applies to everyone, not just government officials; however, a violation of this statute occurs only when one “intentionally” intercepts an electronic communication or intentionally uses or divulges such an intercepted communication.

75. VA. CODE ANN. § 8.01-407.1(A)(3) (Michie Cum. Supp. 2003). Other states have also considered notice requirements. *See, e.g.*, H.B. 2203, 84th Gen. Assem., Reg. Sess. (Ark. 2003); A.B. 1143, 2003 Leg., 2003-04 Reg. Sess. (Cal. 2003).

76. 18 U.S.C. § 2511(1).

6.1 What Constitutes an “Interception”?

An “interception” of an electronic communication (such as an e-mail message) occurs when someone other than the intended recipient gains access to the communication during the *transmission* phase.⁷⁷ Technically, ECPA defines “intercept” as “the aural or other acquisition of the contents of any wire, electronic, or oral communication through the use of any electronic, mechanical, or other device.”⁷⁸ As previously discussed, access to communications that are in *electronic storage* is governed by separate rules.⁷⁹

6.2 Exceptions to Prohibition Against Intercepting Electronic Communications

The prohibition against intentionally intercepting electronic communications is subject to the following notable exceptions:

- **“Normal Course of Business”:** A provider whose facilities are used in the transmission of an electronic communication may intercept, disclose, or use that communication in the “normal course of business” if necessary to provide the service or to protect the rights or property of the provider.⁸⁰ For example, caching and temporary buffering of messages during communication fall under this exception. While this exception is fairly broad, it may not apply to malicious acts committed by an employee of the provider.
- **Compliance with Lawful Court Order:** A provider is authorized to provide information, facilities, or technical assistance to law enforcement when the provider has been served with a court order or other appropriate legal authorization.⁸¹
- **Consent of Party to Communications:** Any private citizen may give consent to the interception of an electronic communication where that person is a party to the communication.⁸² Consent may

77. See *United States v. Smith*, 155 F.3d 1051, 1056-58 (9th Cir. 1998); *Steve Jackson Games, Inc. v. U.S. Secret Serv.*, 36 F.3d 457, 460-62 (5th Cir. 1994); *Wesley Coll. v. Pitts*, 974 F. Supp. 375, 385-86 (D. Del. 1997).

78. 18 U.S.C. § 2510(4) (2000 & Supp. 2003).

79. See discussion *supra* Part 5.

80. 18 U.S.C. § 2511(2)(a)(i).

81. 18 U.S.C. § 2511(2)(a)(ii).

82. 18 U.S.C. § 2511(2)(d). This “one-party consent” rule is the Federal rule. The law of some states establishes a more stringent “all party consent” rule, which makes it illegal under state law to intercept a communication without consent of all parties to the communication. Note, however, that it is not an “interception” for the sender or recipient

be explicit or implied, as is the case with “logon banners” displayed when logging onto many governmental networks, which state that by using the system one consents to monitoring by the entity administering the system.

- **Fraudulent, Unlawful, or Abusive Use of Service:** A provider may record the fact that an electronic communication was initiated or completed in order to protect itself, another provider furnishing service to the communicating parties, or a user of the provider’s service from fraudulent, unlawful, or abusive use of the service.⁸³ An example from traditional telephony is when a carrier traces harassing phone calls.
- **Interception of Computer Trespasser Communications:** The owner or operator of a computer who is the victim of an attack by a trespasser may authorize law enforcement to intercept the trespasser’s communications on the protected computer.⁸⁴ Service providers are not *required* to invite law enforcement to monitor such trespasser communications but, if they wish to do so, 18 U.S.C. § 2511(2)(i) provides an exception to the general prohibition on intercepting electronic communications. A “computer trespasser” includes anyone who accesses a protected computer without authorization, but specifically excludes someone who has a contractual relationship with the computer owner/operator for access to all or part of the computer, for example, an authorized user who is merely violating the service provider’s terms of service.⁸⁵ The statute provides immunity from liability for all claims against service providers who invoke this provision.⁸⁶

6.3 Exceptions to Prohibition Against Divulging Contents of Electronic Communications

As noted above, a provider of electronic communication service is generally prohibited from divulging the contents of such communications.⁸⁷ However, that prohibition is subject to the following exceptions:

of an e-mail to disclose the contents to someone else, either to a private individual or to the government.

83. 18 U.S.C. § 2511(2)(h)(ii).

84. 18 U.S.C. § 2511(2)(i) (2000 & Supp. 2003).

85. 18 U.S.C. § 2510(21).

86. H.R. REP. NO. 107-236, pt. 1, at 58 (2001).

87. 18 U.S.C. § 2511(3)(a).

- if either the “normal course of business” or “pursuant to lawful court order” exceptions discussed above in Part 6.2 apply;
- with the lawful consent of the originator or any addressee or intended recipient of the communication (this could include consent provisions that are part of the provider’s standard terms of service);
- to a person employed, or whose facilities are used, to forward the communication to its destination (for example, another ISP);
- to law enforcement, if the communication was inadvertently obtained by the provider and appears to pertain to the commission of a crime,⁸⁸ or
- as may be necessarily incident to the rendition of services or the protection of the rights or property of the service provider.⁸⁹

6.4 Law Enforcement’s Ability to Intercept Electronic Communications

Law enforcement must meet a fairly stringent set of requirements for authorization to intercept electronic communications in *transmission* (although, access to communications in *electronic storage* is governed by a different set of standards discussed in Part 5 above).

Except in the very limited emergency situations set forth below, a governmental entity must obtain a court order before it can lawfully conduct an interception.⁹⁰ ECPA limits the duration of the interception to 30 days (although the order can be renewed in 30-day increments on application to the issuing court).⁹¹

6.4.1 Law Enforcement Interception of Electronic Communications in Emergency Situations

Certain specially-designated law enforcement officers may authorize the interception of electronic communications before obtaining a court order if they 1) determine that an emergency situation exists involving an immediate danger of death or serious injury, activities that threaten national security, or organized crime activities, and 2) apply for a court order within 48 hours.⁹² Even in this type of emergency situation, law enforce-

88. 18 U.S.C. § 2511(3)(b)(i)-(iv); 18 U.S.C. § 2702(b)(1)-(4).

89. 18 U.S.C. § 2702(b)(5).

90. 18 U.S.C. §§ 2516, 2518 (2000 & Supp. 2003).

91. 18 U.S.C. § 2518(4)-(5).

92. 18 U.S.C. § 2518(7).

ment must stop conducting the interception when the communication sought is obtained or when the application for an order is denied, whichever is earlier. If the application is denied, the contents of any communications already intercepted are likely to be viewed as having been obtained in violation of ECPA. Hence it is both prudent and a common practice among ISPs to obtain either a court order or a certification in writing that an emergency situation exists and that no court order is required before conducting an interception of a communication.

6.5 A Provider's Obligation to Provide Assistance

An order authorizing the interception of electronic communications may direct a provider to furnish “forthwith all information, facilities, and technical assistance necessary to accomplish the interception unobtrusively.”⁹³ Although the scope of this assistance is not defined in the statute, such an order would, at least under most circumstances, require a provider to promptly use its equipment or facilities to accomplish the interception unobtrusively. On the other hand, the Supreme Court has held that “the power of federal courts to impose duties upon third parties is not without limits; unreasonable burdens may not be imposed.”⁹⁴

A governmental entity must reimburse a provider for its “reasonable expenses” in providing interception assistance.⁹⁵ Typically, an order will instruct the provider not to disclose the existence of law enforcement’s interception.

7 PRESERVATION REQUESTS

7.1 Preservation

The government may request that a provider *preserve* stored electronic communications, customer records, and other evidence in its possession, without turning them over to the government, pending the issuance of a legal process.⁹⁶

The federal government has informally taken the position that, because the statute does not explicitly require a written request, the obligation to

93. 18 U.S.C. § 2518(4).

94. *United States v. N.Y. Tel. Co.*, 434 U.S. 159, 172 (1977).

95. 18 U.S.C. § 2518(4).

96. 18 U.S.C. § 2703(f) (2000 & Supp. 2003). Note that this provision has a limited effect because “no law regulates how long network service providers must retain account records in the United States.” KERR, *SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS*, *supra* note 24, § III.G.1, at 104.

preserve evidence attaches when the ISP receives a preservation request by telephone. However, an ISP that receives a telephone preservation request should request a confirmation letter for its own protection. Preservation requests only apply to stored communications and records that the provider has in its possession at the time of the request.⁹⁷ Preservation requests potentially can be burdensome, but the government has sometimes been willing to negotiate the scope of the request in order to reduce the burden on companies that receive a large number of such requests.

7.2 Nondisclosure

It is common for law enforcement agencies to instruct an ISP not to disclose to its subscriber the fact that a subpoena or other process targeting the subscriber has been served on the ISP. 18 U.S.C. § 2705(b) provides a basis and a procedure for nondisclosure orders:

A governmental entity acting under section 2703, when it is not required to notify the subscriber or customer under section 2703(b)(1), or to the extent that it may delay such notice pursuant to subsection (a) of this section, may apply to a court for an order commanding a provider of electronic communications service or remote computing service to whom a warrant, subpoena, or court order is directed, for such period as the court deems appropriate, not to notify any other person of the existence of the warrant, subpoena, or court order. The court shall enter such an order if it determines that there is reason to believe that notification of the existence of the warrant, subpoena, or court order will result in—

- (1) endangering the life or physical safety of an individual;
- (2) flight from prosecution;
- (3) destruction of or tampering with evidence;
- (4) intimidation of potential witnesses; or
- (5) otherwise seriously jeopardizing an investigation or unduly delaying a trial.⁹⁸

97. The Justice Department has taken the position that under § 2703(f), a law enforcement agent “can order a provider to preserve records that have already been created, but cannot order providers to preserve records not yet made.” While other provisions in ECPA permit prospective orders, § 2703(f) does not, because law enforcement officers must comply with electronic surveillance statutes to obtain prospective communications. KERR, *SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS*, *supra* note 24, § III.G.1, at 105.

98. 18 U.S.C. § 2705(b).

This is an area in which formalities matter. Law enforcement agencies sometimes serve a subpoena and include nondisclosure language in the cover letter transmitting the subpoena. This procedure does not conform to § 2705(b), which requires a separate nondisclosure order from a court.

This provision raises other difficulties for ISPs. First, what happens when a nondisclosure order from one court is followed by a subpoena from a second court seeking all records in response to the first subpoena? This can occur in criminal cases when the defendant seeks to review the propriety of the prosecutor's searches. As a rule, it is prudent to assume that the nondisclosure order remains in effect and bars the second subpoena unless the order is lifted by the first court, or is at least considered and specifically superseded by the second court. ISPs often solve the problem by insisting that the prosecution give the defendant any materials received in response to the earlier subpoena, rather than requiring the ISP to go through the discovery process twice.

A second problem concerns state nondisclosure orders. State law and practice varies widely, and it is not always easy to know whether a nondisclosure order meets state law requirements. In addition, there is a reasonable, but untested, view that federal law preempts state law on this point, so that states must follow the procedures of § 2705 to impose a nondisclosure obligation on ISPs. Because of this uncertainty, ISPs are reluctant to rely solely on such orders to provide protection from liability for disclosure. ISPs therefore typically adopt privacy policies that notify subscribers that the ISP may respond to criminal investigative subpoenas or other law enforcement discovery orders without providing notice to the subscriber.

Finally, there is a question whether a nondisclosure obligation may be created by something short of a court order. Sections 2703(b)(1) and 2705(a)(1), when read together, seem to suggest that nondisclosure orders may not be available when the government uses a subpoena. Instead, the subpoenaing agency may simply delay its own notice to the subscriber. But the law enforcement agency is likely to argue that the delayed notice provision is binding on the ISP as well. While this view is questionable, no definitive interpretation of the provision is yet available. ISPs asked to follow this view are likely, at a minimum, to demand particularly strict adherence to the procedural requirements for such a delayed notice, including a copy of the "written certification of a supervisory official" required by § 2705(a)(1)(B).

8 NATIONAL SECURITY INVESTIGATIONS

Orders of the Foreign Intelligence Surveillance Court are classified. Government regulations provide that classified information may only be provided to individuals who have the appropriate security clearance and have a need to know the classified information. Although the FBI sometimes presses ISPs to obtain clearances in order to see FISA orders, such clearance is not required. As a rule, the government may not require a private citizen to obtain a clearance and thus become subject to many new legal liabilities and obligations. Nor may the government require a private citizen to obey court orders without letting the citizen see the order. Thus, if the government wants a FISA order carried out, it must show it to the ISP's personnel, whether or not the personnel agree to obtain clearances. In an emergency, the regulations have procedures permitting the government to provide classified information to an uncleared individual. An ISP without cleared personnel may insist on the use of this provision.

The regulations also require that classified information be stored in a government-approved secure facility. The government is on stronger ground when it refuses to allow the ISP to keep copies of classified documents in facilities that are not secure. If a company does not have such a secure facility, the FBI will show the FISA court order to the company and give the company a trust receipt; the FBI will retain the classified court order.

Company personnel who receive, or who are shown, a classified Foreign Intelligence Surveillance Court order may consult with company executives or with legal counsel where necessary. It is the government's position that such company personnel may not provide any classified information (for example, the name of the target of the search or surveillance, any identifying information about the target, or any classified information about the technique the government is using) to uncleared personnel. Certainly it is prudent for ISP personnel to avoid unnecessary disclosure of such information; however, in some cases disclosure of some classified information is necessary to provide a full view of the risks to the CEO or of legal issues to outside counsel. Classified information may not be used as a way to avoid an ISP's normal corporate oversight or legal clearances. The scope of any expected disclosure should be made clear to the government officials who serve the court order, however, because the government may wish to withdraw the order rather than accept the disclosure that the ISP deems necessary to brief management properly.

Whether to obtain security clearances and establish secure facilities for storage of orders is not a simple decision. The government will press

heavily for full security approval, but before agreeing to do so, ISPs should be sure that they know the full cost of security arrangements and whether the government will reimburse those costs.

8.1 Customer Records

As noted before,⁹⁹ a provider is generally not permitted to disclose customer records to the government without appropriate legal authorization. The FBI has three different authorities for obtaining such records in national security investigations.

8.1.1 FISA Order for Business Records

First, the FBI may obtain a court order from the Foreign Intelligence Surveillance court to require “the production of any tangible things (including books, records, papers, documents, and other items) for an investigation to obtain foreign intelligence information not concerning a United States person or to protect against international terrorism or clandestine intelligence activities.”¹⁰⁰ It is not entirely clear whether this authority covers electronic records. For example, are stored customer e-mails “tangible things”? Do they become tangible things if they are burned to a CD? Given the emphasis on documentary evidence in this provision, it seems likely that the government will argue that electronic documents do fall within its scope, notwithstanding their intangibility.

If a provider is served with an order for such business records, it is prohibited from disclosing the fact that the FBI had sought or obtained those items.¹⁰¹ However, this confidentiality requirement is not intended to prevent company personnel from seeking guidance from company executives or inside or outside legal counsel where necessary. A provider that complies in good faith with an order to provide such business records is immunized against any liability.¹⁰²

8.1.2 FISA Order for Physical Search

Second, the government also may obtain a court order to conduct a physical search of premises or property within the United States that is “intended to result in a seizure, reproduction, inspection, or alteration of information, material, or property.”¹⁰³ Under a physical search order, an ISP may be directed to “furnish all information, facilities, or assistance

99. See *supra* Part 3.2.

100. 50 U.S.C. § 1861(a)(1).

101. 50 U.S.C. § 1861(d).

102. 50 U.S.C. § 1861(e).

103. 50 U.S.C. § 1821(5) (2000 & Supp. 2003).

necessary to accomplish the physical search in such a manner as will protect its secrecy and produce a minimum of interference with the services that such [ISP] . . . is providing the target of the physical search.”¹⁰⁴ A FISA order for physical search is analogous to a search warrant in the criminal context and may be used to obtain communications and content stored in the target user’s account. Unlike a search warrant, a FISA order for physical search typically has a duration of 90 days (it may be up to one year under certain circumstances), but this does not mean that a series of searches may be carried out over that period. Instead, it generally authorizes the government to complete a single search within that time.

Service providers are required to maintain records related to a secret search in accordance with security procedures set by the Attorney General and the Director of Central Intelligence.¹⁰⁵ The government must compensate providers “at the prevailing rate” for assistance provided under this section.¹⁰⁶

8.1.3 *Certification for Subscriber Information and Toll Records in Counter-Intelligence Investigations*

The Director of the FBI is authorized to issue a certification, generally described as a “national security letter,” requiring a provider to release certain subscriber information and toll records, as well as “electronic communication transactional records.” Such a letter takes the place of the subpoena, search warrant, or court order that otherwise would be necessary.¹⁰⁷ This provision is an awkward one for ISPs, mainly because it uses terms that do not appear elsewhere in the electronic intercept statutes and are not defined. Viewed in relation to the more precise provisions of § 2703, § 2709 seems to give the government access both to subscriber identity information (subscriber information and toll billing records information) of the sort described in § 2703(c) as well as access to most of the transactional information covered by § 2703(d). But perhaps not all. Because this section is limited to “electronic communication transactional records,” records of other transactions (for example, purchases from an online store) would not be made available in response to a national security letter, but would be available in response to a “section 2703(d)” order.

104. 50 U.S.C. § 1822(a)(4)(A)(i).

105. 50 U.S.C. § 1822(a)(4)(A)(ii).

106. 50 U.S.C. § 1822(a)(4)(B).

107. 18 U.S.C. § 2709. While the certification may require a telecommunications provider to disclose “local and long distance toll billing records” for the subscriber, ISPs do not maintain such records and, therefore, are not required to disclose them.

A provider is prohibited from disclosing that the FBI had served a certification for such basic subscriber information.¹⁰⁸ However, as mentioned above, company personnel may discuss the certification, as necessary, to get guidance from company executives or legal counsel. Providers that comply with a certification to provide such basic subscriber information are immunized against any liability.¹⁰⁹

8.2 Pen Registers and Trap and Trace Devices

As in criminal investigations,¹¹⁰ the government can also obtain a pen register or trap and trace order in national security investigations. Generally, these FISA pen register or trap and trace orders are treated almost identically to criminal pen register or trap and trace orders.

8.2.1 *The Government's Ability to Use FISA Pen Register and Trap and Trace Surveillances*

Except in the very limited emergency situations set forth below, the government must obtain a FISA order before it can lawfully conduct a pen register or trap and trace surveillance in a national security investigation.¹¹¹ An order may only authorize the installation and use of a pen register or trap and trace device for up to 90 days, although the government may obtain an unlimited number of 90-day extensions.¹¹²

Like the modifications to ECPA's pen register and trap and trace provisions,¹¹³ the USA PATRIOT Act also amended FISA to clarify the government's ability to use pen register and trap and trace devices to capture routing information for electronic communications. FISA's original statutory language had referred to "telephone lines," which had raised questions about whether the pen register and trap and trace authority could be extended to Internet communications (such as e-mail or web browsing). The USA PATRIOT Act resolved this ambiguity by expanding the statute's language to include any "other facility to which the pen register or trap and trace device is to be attached or applied."¹¹⁴

108. 18 U.S.C. § 2709(c).

109. 18 U.S.C. §§ 2703(e) & 2707(e) (2000 & Supp. 2003).

110. *See supra* Part 4.

111. 50 U.S.C. § 1842.

112. 50 U.S.C. § 1842(e). In contrast, criminal pen register/trap and trace orders and extensions only last 60 days. 18 U.S.C. § 3123(c).

113. *See supra* Part 4.

114. 50 U.S.C. § 1842(d)(2)(A).

8.2.2 *Emergency Law Enforcement Use of Pen Register and Trap and Trace Surveillances*

The Attorney General may authorize the installation and use of a FISA pen register or trap and trace surveillance before obtaining a court order, but only if the Attorney General: 1) determines that an emergency exists requiring the immediate installation of the surveillance before an order “can with due diligence be obtained;” and 2) subsequently applies to the Foreign Intelligence Surveillance court for an order within 48 hours.¹¹⁵ Even in this type of emergency situation, the government must stop using the pen register or trap and trace surveillance after 48 hours (at the latest) if it subsequently fails to obtain a court order.¹¹⁶

8.2.3 *A Provider’s Obligations to Provide Assistance*

An order authorizing the installation of a pen register or trap and trace surveillance may direct a provider to furnish “any information, facilities, or technical assistance necessary to accomplish the installation and operation.”¹¹⁷ The government must reimburse a provider for its “reasonable expenses” in providing such assistance.¹¹⁸ No cause of action can be brought against a provider for furnishing any assistance or information in accordance with a FISA pen register or trap and trace order.¹¹⁹

As mentioned above, FISA pen register and trap and trace orders are governed by special confidentiality requirements. A FISA order will direct a provider to maintain the order and “any records concerning the pen register and trap and trace device or the aid furnished” in compliance with the security procedures approved by the Attorney General and the Director of Central Intelligence.¹²⁰ Again, however, this confidentiality requirement is not intended to prevent personnel from getting guidance from company executives or inside or outside counsel where necessary.

8.3 **Interception or Disclosure of Communications**

Finally, like criminal investigations,¹²¹ the government can also obtain a wiretap order to intercept a target’s communications in national security

115. 50 U.S.C. § 1843(a)-(b) (2000 & Supp. 2003).

116. 50 U.S.C. § 1843(c).

117. 50 U.S.C. § 1842(d)(2)(B)(i).

118. 50 U.S.C. § 1842(d)(2)(B)(iii).

119. 50 U.S.C. § 1842(f).

120. 50 U.S.C. § 1842(d)(2)(B)(ii)(II).

121. *See supra* Part 6.

investigations.¹²² Generally, these FISA interceptions are treated almost identically to criminal wiretap orders.

8.3.1 *Government's Ability to Intercept Communications*

Except in certain limited situations set forth below, the government must obtain a court order before it can lawfully intercept a target's communications.¹²³ FISA wiretap orders usually last up to 90 days, although an order may last up to one year. The government may obtain an unlimited number of extensions.¹²⁴

8.3.1.1 "Roving" FISA Orders

FISA allows the government to serve a "roving" wiretap order on multiple providers "in circumstances where the Court finds that the actions of the target . . . may have the effect of thwarting the identification" of a single provider.¹²⁵ For example, if a target frequently uses different Internet accounts with various ISPs, the government might serve an order on all of the providers, instructing them to implement surveillance on any account the government believes the target may be using at a particular moment.

As a result, an ISP could receive a Foreign Intelligence Surveillance Court order in which the ISP is not specifically named and the service that is under surveillance is not specifically identified. In order to minimize the risk of miscommunication, the provider should check the description of the target provided in the order to ascertain that the target really is one of its subscribers. ISPs with questions about such an order should discuss the matter with the government official who served them with the court order.

8.3.1.2 Attorney General's Certification to Intercept Communications without a Court Order

In certain circumstances, FISA permits the Attorney General to authorize the interception of communications for up to one year *without* a court order. However, in order to do so, the Attorney General must certify under oath that "there is no substantial likelihood that the surveillance will acquire the contents of any communications to which a United States person is a party" and must file a copy of the certification with the Foreign Intel-

122. 50 U.S.C. § 1805 (2000 & Supp. 2003).

123. 50 U.S.C. § 1805.

124. 50 U.S.C. § 1805(e). In contrast, the duration of criminal wiretap orders and extensions is limited to only 30 days. 18 U.S.C. § 2518(5).

125. 50 U.S.C. § 1805(c)(2)(B).

ligence Surveillance court.¹²⁶ Because of these requirements, such certifications are extremely uncommon.

8.3.2 *Government Interception of Communications in Emergency Situations*

The Attorney General may authorize the interception of communications before obtaining a FISA court order if the Attorney General: 1) determines that an emergency exists requiring the immediate installation of the surveillance before an order “can with due diligence be obtained;” and 2) subsequently applies to the Foreign Intelligence Surveillance Court for an order within 72 hours.¹²⁷ Even in this type of emergency situation, the government must stop intercepting the communications after 72 hours (at the latest) if it subsequently fails to obtain a court order.¹²⁸

8.3.3 *A Provider’s Obligations to Provide Assistance*

An order authorizing a FISA interception may direct a provider to furnish “all information, facilities, or technical assistance necessary to accomplish the electronic surveillance.”¹²⁹ Law enforcement must compensate a provider for its expenses in providing such assistance.¹³⁰ No cause of action can be brought against a provider that furnishes any assistance or information in accordance with a FISA wiretap order.¹³¹ As mentioned above, FISA court orders are governed by special confidentiality requirements. A FISA order will direct a provider to maintain the order and “any records concerning the surveillance or the aid furnished” in compliance with the security procedures approved by the Attorney General and the Director of Central Intelligence.¹³² However, company personnel may discuss these orders, as necessary, to get guidance from company executives or inside or outside legal counsel.

9 REIMBURSEMENT FOR COSTS

Federal law makes it clear that ISPs are not expected to conduct government investigations at their own expense:

126. 50 U.S.C. § 1802.

127. 50 U.S.C. § 1805(f).

128. 50 U.S.C. § 1805(f) (2000 & Supp. 2003).

129. 50 U.S.C. § 1805(c)(2)(B).

130. 50 U.S.C. § 1805(c)(2)(D).

131. 50 U.S.C. § 1805(i).

132. 50 U.S.C. § 1805(c)(2)(C).

(a) Payment—Except as otherwise provided in subsection (c), a governmental entity obtaining the contents of communications, records, or other information under section 2702, 2703, or 2704 of this title shall pay to the person or entity assembling or providing such information a fee for reimbursement for such costs as are reasonably necessary and which have been directly incurred in searching for, assembling, reproducing, or otherwise providing such information. Such reimbursable costs shall include any costs due to necessary disruption of normal operations of any electronic communication service or remote computing service in which such information may be stored.

(b) Amount—The amount of the fee provided by subsection (a) shall be as mutually agreed by the governmental entity and the person or entity providing the information, or, in the absence of agreement, shall be as determined by the court which issued the order for production of such information (or the court before which a criminal prosecution relating to such information would be brought, if no court order was issued for production of the information).¹³³

The exception to this provision is so narrow that it mainly serves to establish that ISPs should always be reimbursed. Paragraph (c) simply excludes searches of “telephone toll records and telephone listings” supplied by a “communications common carrier” from reimbursement and neither of those terms applies to ISPs.

Notwithstanding this clear language, law enforcement agencies are often reluctant to provide reimbursement. This can make it hard to “mutually agree” on an appropriate fee as the statute provides.

ISPs must in any event take the lead in identifying such “costs as are reasonably necessary and . . . directly incurred in searching for, assembling, reproducing, or otherwise providing” the requested records.¹³⁴ At a minimum, those costs should include the hourly cost of any paralegals, engineers, or lawyers involved in the search. In addition to salary, it is reasonable to include benefits and social security, as well as overhead items such as office space and utilities. It is less common for ISPs to include other overhead items, from general administrative allocations to outside counsel costs associated with searches, though the statute would seem to permit their recovery if the costs are properly attributed to searches. When tools have been developed by company engineers to make searches more

133. 18 U.S.C. § 2706.

134. 18 U.S.C. § 2706(a) (2000 & Supp. 2003).

efficient, the cost of the total can be amortized and applied to all searches conducted with the tool until the cost has been recovered.

Some ISPs use an estimate of costs to create a per-search charge; others charge based on the number of hours required. The latter method seems to produce fewer objections from law enforcement.

A particular difficulty arises in the context of preservation requests under § 2703(f). There is no provision for receiving payment from the government when a preservation request is made, even though the preservation itself requires substantial searching and assembling. Of course, these costs can be assessed when the subpoena or other order arrives and the information is released. But close to half of all preservation requests are simply abandoned. No subpoena is ever issued. The costs incurred as a result of these requests cannot be recovered in the subpoena process, except as part of a general office overhead charge.

10 LIABILITY FOR ECPA VIOLATIONS

10.1 Criminal Liability

Violations of ECPA's prohibition of the interception and disclosure of electronic communications¹³⁵ can give rise to criminal liability, including fines and/or imprisonment for up to five years.¹³⁶ Knowing violations of ECPA's provisions concerning pen registers and trap and trace devices¹³⁷ also can give rise to criminal liability, including fines and/or imprisonment for up to one year.¹³⁸

In addition, ECPA imposes criminal liability for intentionally accessing, without authorization, a stored electronic communication.¹³⁹ This provision does not apply, however, to conduct authorized by providers of electronic communication service, and therefore would not apply to ISPs, except in cases of malicious acts by their employees.¹⁴⁰

10.2 Civil Liability

Under ECPA, a person whose electronic communications are intercepted or disclosed may sue a provider for illegally intercepting or disclosing the communications.¹⁴¹ The person must allege, however, that the pro-

135. *See supra* Part 6.

136. 18 U.S.C. § 2511(4)(a).

137. *See supra* Part 4.

138. 18 U.S.C. § 3121(d).

139. 18 U.S.C. § 2701(a); *see supra* Part 5.

140. 18 U.S.C. § 2701(c) (2000 & Supp. 2003).

141. *See supra* Part 6.

vider or one of its employees: a) acted without a facially valid court order or other lawful authorization; b) acted beyond the scope of a court order or lawful authorization; or c) acted in bad faith.¹⁴² A provider, subscriber, or customer may also sue a party for illegally obtaining or divulging his stored communications or records.¹⁴³ Possible relief for violations of either provision includes damages (including, potentially, punitive damages) and reasonable attorneys' fees and litigation costs.¹⁴⁴ ECPA's provisions concerning pen registers and trap and trace devices¹⁴⁵ do not state a cause of action for violations.

10.3 Service Provider Immunity

A lawsuit cannot be brought, under ECPA or otherwise, against a provider for assisting law enforcement in carrying out a court order, warrant, subpoena, or other lawful authorization.¹⁴⁶ In addition, a provider's good faith reliance on court orders or other legal authorizations is a complete defense against any civil or criminal actions brought under ECPA.¹⁴⁷

11 INTERNATIONAL JURISDICTIONAL ISSUES

The global nature of the Internet and many ISPs' networks frequently raises international jurisdictional issues with respect to ECPA. For example, a multinational ISP might be presented with a non-U.S. judicial order for the interception and/or disclosure of electronic communications stored or transmitted within the United States. The ISP would be at risk of violating ECPA, however, if it complied with the non-U.S. order by intercepting or disclosing electronic communications stored or transmitted within the United States, absent an applicable exception to ECPA.¹⁴⁸ Therefore, in order to access such communications, in most cases, the foreign government would have to obtain the appropriate (ECPA-compliant) U.S. legal

142. 18 U.S.C. § 2520(a).

143. 18 U.S.C. § 2707(a); *see supra* Part 5. In one notable ECPA lawsuit, a Virginia ISP paid an undisclosed amount in 1998 to settle a claim by a homosexual Navy sailor that it had improperly disclosed information about him to a Navy investigator. *See Bradley Graham, Gay Sailor Takes Navy Retirement Settlement; AOL Also Will Pay For Privacy Violation*, WASH. POST, June 13, 1998, at A3; *see also* *McVeigh v. Cohen*, 983 F. Supp. 215 (D.D.C. 1998).

144. 18 U.S.C. §§ 2520(b)-(c), 2707(b)-(c).

145. *See supra* Part 4.

146. 18 U.S.C. §§ 2511(2)(a)(ii), 2703(e), 3124(d) (2000 & Supp. 2003).

147. 18 U.S.C. §§ 2520(d), 2707(e), 3124(e).

148. *See supra* Part 5.3.

order through the United States government.¹⁴⁹ Foreign governments have been known to object to such time-consuming processes, and claim that an ISP subject to its jurisdiction must comply with its order without obtaining process through the United States government. ISP staff should coordinate closely with legal counsel in such situations.

The USA PATRIOT Act made two changes to ECPA that are particularly notable in the international context. First, it altered the statutory provisions for search warrants, “section 2703(d)” orders, and pen register and trap and trace orders such that jurisdiction for issuing these orders vests in a court having jurisdiction over the underlying offense. When dealing with international requests for assistance, this change raises a question as to whether or not federal courts have jurisdiction to grant relief. Because the offenses being investigated were committed abroad in circumstances under which a federal district court would not have jurisdiction over the offense, there is an argument that federal courts do not have jurisdiction. A counterargument likely to be advanced by the government is that the court acquires jurisdiction over the offense by virtue of the international request for assistance. As a practical matter, an ISP will only be required to act when presented with an order. Having issued the order, the court will have resolved the issue.

Second, the Act amends the definition of “protected computer” to make clear that this term includes computers outside of the United States so long as they affect “interstate or foreign commerce or communication of the United States.”¹⁵⁰ The United States can now use purely domestic procedures, as opposed to international legal assistance, to join in international hacker investigations when hacking a foreign computer constitutes an offense under U.S. law.

Finally, enactment of the USA PATRIOT Act and the Homeland Security Act of 2002,¹⁵¹ has made international waters a bit more dangerous for ISPs. Section 225 of the Homeland Security Act amends the emergency disclosure provision. Before the amendment, ISPs could make emergency disclosures of the contents of a communication to “a law en-

149. There is a formal process for “domesticating” such foreign orders. The process is often referred to as the MLAT-process—named for the Mutual Legal Assistance Treaties that provide for this form of assistance by the U.S. government—and is operated by the Department of Justice Office of International Affairs. *See* Mutual Legal Assistance and Other Agreements (“MLAT”), at <http://travel.state.gov/mlat.html> (last visited Oct. 17, 2003).

150. 18 U.S.C. § 1030(e)(2)(B).

151. Pub. L. No. 107-296, § 1, 116 Stat. 2135 (2002) (codified as 6 U.S.C. § 101 (Supp. 2003)).

forcement agency.” After the amendment, such disclosures may be made “to a Federal, State, or local governmental entity.” While this may not sound like much of a change, it may exclude emergency disclosures to a foreign law enforcement agency unless that agency can be characterized as a “local government entity.” Of course, whether U.S. law governs such disclosures and whether the subscribers may have consented to such disclosures in an ISP’s acceptable use policies, are separate questions.

12 INTERRELATION WITH STATE LAW

Federal law has “preempted” the states in the field of electronic surveillance and interception of wire and electronic communications. 18 U.S.C. § 2516(2), which authorizes state electronic surveillance laws, lists the offenses for which state statutes may authorize interceptions, and requires that the procedures set out in 18 U.S.C. § 2518 be followed in issuing interception orders. At the same time, since the original wiretap law in 1968, it has been clear that a state may have stricter (but not more lenient) requirements. ECPA has similar provisions regarding state authority¹⁵² and the same logic applies to these other forms of obtaining evidence—that is, state requirements may be stricter than the federal statute, but not more permissive. Pursuant to these authorities, most states have adopted their own wiretap and pen register and trap and trace statutes, as well as “mini-ECPAs.”¹⁵³

After the passage of the USA PATRIOT Act, several states have sought to amend their laws to track the federal amendments. However, proposed state amendments in some states have differed in critical ways from federal law. ISPs should carefully track amendments to state laws, both before and after passage.

Another emerging problem area is the extent to which ECPA, particularly 18 U.S.C. § 2702, may prohibit access to the content of e-mail based on state discovery subpoenas. In addition, several states are considering various privacy laws and criminal statutes relating to illegal Internet content, all of which could have serious consequences for ISPs.

The law in this area is in a state of flux, with many federalism and other constitutional issues yet to be resolved by the courts. ISPs and their

152. *See, e.g.*, 18 U.S.C. § 2703; 18 U.S.C. § 3122 (2000 & Supp. 2003).

153. *See, e.g.*, 720 ILL. COMP. STAT. 5/14-1 (2003) (defining “eavesdropping” to include interception of electronic communications); N.Y. PENAL LAW § 250.05 (McKinney 2000) (defining the offense of eavesdropping as including the “intercepting or accessing of an electronic communication”); VA. CODE ANN. §§ 19.2-61 to 2-70.3 (Michie Cum. Supp. 2003) (Chapter 6—Interception of Wire, Electronic or Oral Communications).

legal counsel should bear in mind that both state and federal law can be the basis for electronic surveillance requests and the legal requirements, provisions for reimbursement, and immunity from liability under state law must be checked carefully with their federal counterparts.

13 CONTACT INFORMATION

If you would like to get in touch with some of the contributors to this Guide, see the contact information below:

Drew C. Arena
Assistant General Counsel for Legal
Compliance
Verizon Communications
1515 North Courthouse Road, Suite 500
Arlington, Virginia 22201-2909
Telephone: 703.351.3007
Fax: 703.351.3667
E-mail: drew.c.arena@verizon.com

Stewart A. Baker
Stephoe & Johnson LLP
1330 Connecticut Avenue, N.W.
Washington, D.C. 20036
Telephone: 202.429.6413
Fax: 202.261.9825
E-mail: sbaker@stephoe.com

Elizabeth Banker
Associate General Counsel
Yahoo! Inc.
701 First Avenue
Sunnyvale, California 94089
Compliance Fax: 408.349.7941

Christopher G. Bubb
Assistant General Counsel
American Online, Inc.
22000 AOL Way
Dulles, Virginia 20166
Fax: 703.265.2305

Kate Dean
Manager
US Internet Service Provider Associa-
tion
1330 Connecticut Avenue, N.W.
Washington, D.C. 20036
Telephone: 202.862.3816
Fax: 202.261.0604
E-mail: kdean@stephoe.com

Susan Kelley Koeppen
Senior Attorney
Microsoft Corporation
One Microsoft Way
Redmond, Washington 98052
Telephone: 425.705.4788
Fax: 425.936.7329
E-mail: susankoe@microsoft.com

14 GLOSSARY

Electronic Communication: For purposes of ECPA, “electronic communication” is defined broadly to encompass a wide range of technologies. Specifically, ECPA defines an “electronic communication” as “any transfer of signs, signals, writing, images, sounds, data, or intelligence of any nature transmitted in

whole or in part by a wire, radio, electromagnetic, photoelectric or photooptical system that affects interstate or foreign commerce[.]”¹⁵⁴

Electronic Communication Service: As defined by ECPA, “any service which provides to users thereof the ability to send or receive wire or electronic communications,” including electronic mail services.¹⁵⁵

Electronic Storage: As defined by ECPA, “any temporary, intermediate storage of a wire or electronic communication incidental to the electronic transmission thereof” and “any storage of such communication by an electronic communication service for purposes of backup protection.”¹⁵⁶

Interception: An “intercept” of an electronic communication (such as an e-mail message) occurs when someone other than the intended recipient gains access to the communication during the transmission phase (as opposed to when it is stored). Technically, ECPA defines “intercept” as “the aural or other acquisition of the contents of any wire, electronic, or oral communication through the use of any electronic, mechanical, or other device.”¹⁵⁷

Pen Register: A device or process that attaches to a telephone line or facility and records outgoing dialing, routing, addressing, or signaling information. It does not record the content of communications.¹⁵⁸

Remote Computing Service: As defined by ECPA, the provision to the public of computer storage and processing services.¹⁵⁹

Trap and Trace Device: A device or process which attaches to a telephone line or facility and records “the incoming electronic or other impulses [for example, telephone numbers] which identify the originating number or other dialing, routing, addressing, and signaling information reasonably likely to identify the source

154. 18 U.S.C. § 2510(12).

155. 18 U.S.C. § 2510(15).

156. 18 U.S.C. § 2510(17).

157. 18 U.S.C. § 2510(4).

158. 18 U.S.C. § 3127(3) (2000 & Supp. 2003).

159. 18 U.S.C. § 2711(2).

of a wire or electronic communication” It does not record the content of communications.¹⁶⁰

160. 18 U.S.C. § 3127(4).

THE BUSINESS METHOD PATENT MYTH

By John R. Allison and Emerson H. Tiller[†]

ABSTRACT

Internet business method patents have been roundly criticized by most observers as being singularly inferior to most other patents. Many have even argued that business methods should not be patentable subject matter. As a result, Congress and the Patent and Trademark Office (“PTO”) singled them out for special treatment. All of these criticisms were, however, voiced without empirical support.

We gathered data on most Internet business method patents issued through the end of 1999 and compared them with a large contemporaneous data set of patents in general. We also compared them with patents in fourteen individual technology areas within the general patent data set. Our comparison focused on several metrics that we believe serve as good proxies for patent quality and value. We found that Internet business method patents appear to have been no worse than the average patent, and possibly even better than most. They also appear to have been no worse, and possibly even better, than patents in most individual technology areas.

These findings lead us to question the conventional wisdom that Internet business method patents were uniquely deficient. We briefly explore some possible explanations for the chasm between the accepted

© 2003 John R. Allison and Emerson H. Tiller

[†] John R. Allison is the Spence Centennial Professor of Business Administration, McCombs School of Business, and Professor of Intellectual Property Law, McCombs School and the College of Engineering, University of Texas at Austin. Emerson H. Tiller is Professor of Law, Northwestern University School of Law.

Partial funding for this study was provided by the National Academy of Sciences (“NAS”), Science, Technology, and Policy Board (“STEP”). We appreciate the ideas and suggestions offered by several members of STEP’s Committee on Intellectual Property Rights in the Knowledge-Based Economy.

The authors thank Thomas Bohman and Xinlei Wang of the Information Technology Services at the University of Texas for statistical consulting. The authors also thank participants at the University of Chicago Law School’s Law and Economics Workshop, University of California at Berkeley, Workshop on Institutional Analysis, Northwestern Law School’s Faculty Workshop, the University of Southern California Law School’s Workshop on Law, Economics and Organization, and the Huber-Hurst Research Seminar, University of Florida, for comments on this Article or portions thereof, as well as several individuals who provided especially helpful comments on earlier drafts, including Connie Bagley and Mark Lemley.

view and what we believe to have been the reality, including the possibility that negative opinions about these patents may have been the result of an information cascade. More importantly, we believe that efforts to single out these patents for special treatment not only lacked sound justification in the particular case but also reveal more fundamental problems associated with ex ante definitions to carve out any particular technology area for different treatment.

TABLE OF CONTENTS

I.	INTRODUCTION	989
II.	CRITICISMS OF BUSINESS METHOD PATENTS	1007
	A. The Diverse Nature of the Outcry	1007
	B. A Bit of Software Patent Déjà Vu	1012
	C. Prior Art-Related Criticisms Migrating from Software to Business Method Patents	1015
III.	RESPONSES BY CONGRESS AND THE PTO	1018
	A. Congressional Responses	1018
	1. <i>The First Inventor Defense Act</i>	1018
	2. <i>H.R. 5364</i>	1021
	3. <i>H.R. 1332</i>	1023
	B. PTO Responses	1024
IV.	DATA ON INTERNET BUSINESS METHOD PATENTS AND GENERAL PATENTS	1026
	A. The General Patent Data Set	1026
	1. <i>Defining the Technology Areas</i>	1027
	2. <i>The Fourteen Technology Areas</i>	1028
	B. The Internet Business Method Patent Data Set	1031
	1. <i>Constructing the Data Set</i>	1032
	2. <i>Categorizing Internet Patents</i>	1034
	3. <i>Excluding the Last Category</i>	1036
V.	COMPARISON OF PATENT CHARACTERISTICS	1036
	A. Number of Prior Art References	1036
	1. <i>Internet Business Method Patents Compared With the Average Patent</i>	1039
	2. <i>Internet Business Method Patents Compared With Patents in Individual Technology Areas</i>	1043
	B. Types of Nonpatent Prior Art References	1045
	1. <i>Internet Business Method Patents Compared With the Average Patent</i>	1048
	2. <i>Internet Business Method Patents Compared With Patents in Individual Technology Areas</i>	1050
	C. Total Patent Claims	1052
	1. <i>Internet Business Method Patents Compared With the Average Patent</i>	1057
	2. <i>Internet Business Method Patents Compared With Patents in Individual Technology Areas</i>	1057
	D. Number of Inventors	1058
	1. <i>Internet Business Method Patents Compared With the Average Patent</i>	1060
	2. <i>Internet Business Method Patents Compared With Patents in Individual Technology Areas</i>	1061
	E. Time Spent in the Patent Office	1063
	1. <i>The Effect of Ancestor Applications on Average Pendency Times</i>	1063

2.	<i>Other Reasons for Variation in Pendency Times</i>	1065
3.	<i>Method of Comparison</i>	1070
4.	<i>Pendency Times from the Original Priority Filing Date: Internet Business Method Patents Compared With General Patents and With Patents in Each Technology Area</i>	1071
5.	<i>Pendency Times from the Most Recent Filing Date: Internet Business Method Patents Compared With General Patents and With Patents in Each Technology Area</i>	1074
VI.	SUGGESTED EXPLANATIONS FOR THE GAP BETWEEN CONVENTIONAL WISDOM AND REALITY	1077
A.	Possible Explanations.....	1077
B.	An Information Cascade of Negative Opinion.....	1079
VII.	CONCLUSION	1081

I. INTRODUCTION

A patent represents a bargain with society. In return for the right to exclude others from making, using, or selling an invention for twenty years, the inventor must disclose the invention's details to the public instead of keeping them secret.¹ To justify such a grant, the inventor must introduce and reveal a technological advance that is both different from what has been done before—the requirement of “novelty”—and different enough so that it represents more than a trivial extension of what is already in the public domain—the requirement of “nonobviousness.”² Although it is the exceedingly rare patent that, by itself, confers significant market power on its owner, such exclusive rights to control knowledge that is otherwise free for the taking may cause a number of economic distortions.³ Thus, there is cause for concern when patents are granted erroneously.⁴ If patents are issued that should not be, the public does not

1. 35 U.S.C. § 112 (2000) (disclosure requirement).

2. *Id.* § 102 (novelty requirement); *id.* § 103 (nonobviousness requirement).

3. There is a vast literature on the economics of patents. One example is the increasingly large theoretical literature on various approaches to the proper scope and allocation of patent rights, thoroughly surveyed in Dan L. Burk & Mark A. Lemley, *Policy Levers in Patent Law*, 89 VA. L. REV. 1575, 1595-1630 (2003).

4. We assume, as do most observers, that patent quality matters. Staking out a distinctly minority position, Mark Lemley has argued that issuance of substandard patents by the PTO does not cause enough economic harm to warrant the expenditure of additional resources to increase patent quality, because only a very small portion are ever licensed or enforced and the federal courts are better equipped to assess patent validity in depth. Mark A. Lemley, *Rational Ignorance at the Patent Office*, 95 NW. U. L. REV. 1495 (2001) [hereinafter Lemley, *Rational Ignorance*]. Although exploring his arguments is beyond the scope of this Article, we observe in passing that he does condition them on significant reforms in patent litigation, such as lowering the burden of proof for one challenging the validity of a patent from “clear and convincing evidence” to “preponderance of the evidence.” *Id.* at 1528-29. Even with such litigation reforms, we remain uncon-

receive the benefit of technological advancement in return for the patentee's exclusive right.

Of late, there has been a great deal of negative commentary about patents on "methods of doing business," as well as targeted action against such patents by both Congress and the PTO. Although the criticism and action have focused on business method patents generally, the majority of these patents entail business processes and techniques implemented on the Internet. Moreover, those business method patents tailored for Internet usage are the most controversial and the most likely to cause economic harm if they are granted when they should not be because of their potential for impeding electronic commerce while it is still maturing.

In this Article, we present the findings of our empirical study comparing two sets of patent data. We compared Internet business method patents issued through December 31, 1999 with a large random sample of general patents issued during a contemporaneous time period. We find that Internet business method patents actually fare quite well statistically, contrary to the conventional wisdom that Internet business method patents issued during the early years of their recognition were much worse than others—that they were somehow exceptional.

In the now famous *State Street Bank & Trust Co. v. Signature Financial Group, Inc.*,⁵ the United States Court of Appeals for the Federal Circuit⁶ held that business methods can be patentable subject matter.⁷

vinced. Overall, however, Lemley's assertions are sufficiently serious that a number of his assumptions and estimates warrant further empirical research, as he himself suggests. *Id.* at 1531.

5. 149 F.3d 1368 (Fed. Cir. 1998) (holding that there is no patentable subject matter exception for methods of doing business).

6. The Federal Circuit is the court of appeals having exclusive appellate jurisdiction over patent cases. 28 U.S.C. §§ 1295(a)(1), 1338(a) (2000).

7. The Federal Circuit actually concluded that there never had been a patentable subject matter exception for business methods, despite much contrary conventional wisdom supported by *dicta* in several older cases. In *State Street*, the court explained that earlier decisions referring to such an exception were decided on other grounds. 149 F.3d at 1375-76. For a discussion of the widespread pre-*State Street* assumption among practicing patent attorneys that there was a patentable subject matter exception for business methods, and the absence of a basis for this assumption, see Rinaldo Del Gallo, III, *Are "Methods of Doing Business" Finally Out of Business as a Statutory Rejection?*, 38 IDEA 403 (1998).

Quickly demonstrating that *State Street* was not a fluke, as some observers believed might be the case, the Federal Circuit followed that decision one year later with *AT&T Corp. v. Excel Communications, Inc.*, 172 F.3d 1352, 1355 (Fed. Cir. 1999), in which the court reemphasized the absence of any subject matter exception to patentability for methods of doing business.

Although the PTO had issued patents on software-embodied business methods⁸ before the *State Street* decision in 1998,⁹ the number exploded thereafter.¹⁰ Along with the increased patenting activity came a chorus of criticism. Some observers expressed alarm at the very idea of patenting software-embodied business methods.¹¹ A larger number of

In both *State Street* and *AT&T*, the Federal Circuit also reiterated its 1994 holding in *In re Alappat*, 33 F.3d 1526, 1544 (Fed. Cir. 1994) (en banc), that mathematical subject matter should be treated no differently in patent law than any other subject matter, and therefore, software is patentable subject matter despite the fact that it consists of mathematical algorithms. Algorithms are only unpatentable in the abstract, but are patentable subject matter if they achieve a “useful, concrete, and tangible result.” *AT&T*, 172 F.3d at 1356-58; *State Street*, 149 F.3d at 1373-75. Although the Federal Circuit spoke in terms of patentable subject matter, the distinction between abstract and useful really implicates the utility requirement of patent law rather than the patentable subject matter requirement. See 35 U.S.C. § 101 (2000).

8. The holding in *State Street* that business methods are patentable subject matter was not limited to those implemented by software, but when people speak of business method patents they ordinarily refer only to software-embodied ones. The ones we studied that envision Internet-related use are necessarily implemented by software. Thus viewed, business method patents are a subset of software patents. To distinguish business method patents from other kinds of software patents, we use the term “software patent” to refer to the broader set minus business method patents.

9. The patent at issue in *State Street* is but one example. The patent covered software described (and “claimed,” using patent law terminology for the structure of the claims that follow the written description) as a data processing system for implementing an investment structure whereby mutual funds pooled their assets in an investment portfolio organized as a partnership in a way that assertedly achieved economies of scale in administration coupled with tax advantages. See U.S. Patent No. 5,193,056 (issued Mar. 9, 1993) (“Data Processing System for Hub and Spoke Financial Services Configuration”).

10. Although business method patents may be found in several classifications within the PTO’s classification scheme, Class 705, “Data processing: financial, business practice, management, or cost/price determination,” consists almost exclusively of them. The PTO placed 469 patents in classification 705 in 1998, 833 in 1999, and 1,006 in 2000.

Many have commented on the large number of business method patents issuing in a relatively short time. See, e.g., Michael J. Meurer, *Business Method Patents and Patent Floods*, 8 WASH. U. J.L. & POL’Y 309 (2002) (pointing to the existence of a patent flood in the area of business methods, predicting that the flood will worsen, and describing the economic harm caused by patent floods); Arti Rai, *Addressing the Patent Gold Rush: The Role of Deference to PTO Patent Denials*, 2 WASH. U. J.L. & POL’Y 199, 211 (2000) (documenting and discussing dramatic increase in filings of business method patents, and calling for more deference by the Federal Circuit to PTO application rejections).

11. See, e.g., John R. Thomas, *The Patenting of the Liberal Professions*, 40 B.C. L. REV. 1139, 1185 (1999) [hereinafter Thomas, *Liberal Professions*] (observing that recognizing the patentability of business processes opens the door for patenting new developments in all of human experience, and that patents should remain grounded in science and engineering); see also Alan L. Durham, “Useful Arts” in the Information Age, 1999 BYU

commentators, however, maligned business method patents for their purported lack of quality, devoting much of their attention to the perceived failure of patent applicants and the PTO to adequately take into account the “prior art.”¹² The argument is that patents granted without proper consideration of the prior art are not as likely to meet the statutory requirements of novelty and nonobviousness.¹³

L. REV. 1419, 1526-28 (similarly arguing that software-embodied business method patents should not be patentable subject matter).

12. For present purposes, prior art consists of prior patents and other printed publications relevant to the invention in question, for example, “patent prior art” and “nonpatent prior art.” There are other types of prior art, but these are the only ones that can be analyzed in a study like ours.

13. Most of the prior art-based criticisms represent extensions of similar complaints about software patents generally. It has been observed, for example, that much fundamental programming knowledge was developed long before software was recognized as a candidate for patenting and that, consequently, this knowledge was kept as trade secrets and is not available to those searching for prior art. See, e.g., Julie E. Cohen, *Reverse Engineering and the Rise of Electronic Vigilantism: Intellectual Property Implications of “Lock-Out” Programs*, 68 S. CAL. L. REV. 1091, 1177-79 (1995) [hereinafter Cohen, *Electronic Vigilantism*]; Simson L. Garfinkel, *Patently Absurd*, WIRED, July 1994, available at http://www.wired.com/wired/archive/2.07/patents_pr.html (last visited Nov. 9, 2003). We explain later that previous inventions in many fields are secret, and thus do not create prior art because companies of all kinds often prefer trade secret protection over patent protection, especially for methods (techniques and processes).

Generally speaking, prior art refers to past revelations about the technology being patented. The types of prior art cited in patent applications and issued patents are prior U.S. and foreign patents and prior printed publications from anywhere in the world. Printed publications include those from a vast array of hard copy and electronic sources, including academic and trade journals, company and industry-sponsored publications, university and government reports, software and its documentation, the popular press, published patent applications and search reports, various web sites, and so on. In almost all cases, prior art must have been accessible to interested members of the public, although the patent statute provides for one instance in which “secret prior art” may be used to prevent a patent from issuing or to subsequently invalidate it in court. If X filed a patent application before Y invented, and X later received a patent, X’s patent can be prior art against Y’s application for a patent even though X’s application was held in secret by the PTO at the time Y invented. 35 U.S.C. § 102(e) (2000 & Supp. 2003).

Other types of prior art, which are typically discovered by an accused infringer seeking to invalidate the patent rather than by the PTO, include evidence of a prior public (nonconfidential) use of the invention or evidence that the invention had been placed on sale. Regarding the “critical date,” if a piece of prior art reveals all of the elements of the invention now in question, and if the prior art became accessible either (1) before this inventor invented or (2) more than one year before this inventor filed her patent application, the inventor cannot receive a valid patent. If the prior art became accessible before this invention, we say that this invention lacks novelty. *Id.* § 102(a). If the prior art became accessible more than one year before this inventor filed a patent application, we say

Two of the business method patents often cited as probably failing the requirements of novelty or nonobviousness are Amazon.com's patent on its "one-click" technique for more efficiently ordering merchandise online,¹⁴ and Priceline.com's patent on the reverse auction technique for buying airline tickets on the Internet.¹⁵ Although patents in other areas of technology have brought forth complaints from various quarters,¹⁶ the

that there is a statutory bar to patentability. *Id.* § 102(b). Obviously, §§ 102(a) and (b) can sometimes overlap.

Even if no single piece of prior art revealed all of the elements of the invention in question, two or more pieces of prior art in the relevant field of technology, or in a closely related field, may be combined by the PTO or a court to reach a conclusion that this invention is "obvious." The requirement that an invention have been nonobvious at the time the invention was fully conceived is found in 35 U.S.C. § 103(a) (2000). From the perspective of a hypothetical "ordinarily skilled practitioner in the art," the invention must not have been merely a trivial, or obvious, extension of the cumulative prior art. According to § 102, this invention must be *different* from any single piece of prior art, and according to § 103, this invention must be *different enough* from the cumulative prior art.

14. U.S. Patent No. 5,960,411 (issued Sept. 28, 1999) ("[M]ethod and system for placing a purchase order via a communications network"). Amazon sued Barnesandnoble.com for patent infringement and won a preliminary injunction. *Amazon.com, Inc. v. Barnesandnoble.com, Inc.*, 73 F. Supp. 2d 1228, 1249 (W.D. Wash. 1999). On appeal, however, the Federal Circuit reversed, noting that there were sufficient obviousness-based doubts about the patent's validity to preclude a finding of substantial likelihood of success on the merits. *Amazon.com, Inc. v. Barnesandnoble.com, Inc.* 239 F.3d 1343, 1366 (Fed. Cir. 2001).

15. U.S. Patent No. 5,797,127 (issued Aug. 18, 1998) ("Method, apparatus, and program for pricing, selling, and exercising options to purchase airline tickets"). The reverse auction is sometimes referred to as a Dutch auction. *See, e.g.*, Eugene R. Quinn, Jr., *The Proliferation of Electronic Commerce Patents: Don't Blame the PTO*, 28 RUTGERS COMPUTER & TECH. L.J. 121, 122 n.5 (2002).

16. Gene patents appear to be the only type that has garnered more publicity than business method patents. A survey of articles in the Lexis "News Group File, All" using the simple "patent w/10 gene" search request reveals that comparable or even greater publicity has been given to patents on isolated, purified genes. This publicity in the popular press has been mixed, and most negative comments reveal the perception that such patents reward mere discovery rather than invention. This is not the case, however, because in their natural state, genes cannot exist in isolation. It is only through human intervention that genes can be isolated, purified, and put to use in medical and pharmaceutical research. *See, e.g.*, John M. Golden, *Biotechnology, Technology Policy, and Patentability: Natural Products and Invention in the American System*, 50 EMORY L.J. 101, 122-28 (2001).

Regardless of the reasons stated in objections to gene patentability, the magnitude of reported comment, criticism, and opinion on the subject likely results not so much from the expressed concerns or even from the relative newness of the technology as from underlying fears about privacy, the morality of private ownership of the basic building blocks of life, the uncertain consequences of genetic alterations that may pass to progeny,

magnitude of adverse commentary and reportage on business method patents has been unprecedented.¹⁷

In 1999, Congress responded to the outcry with the First Inventor Defense Act,¹⁸ which creates a special patent-infringement defense for a prior business-method inventor who used it in secret and was later sued for infringement by a second comer who obtained a patent. U.S. patent law provides no such defense for similarly situated "secret" first inventors in other areas of technology, and Congress usually does not provide special treatment for any category of technology.¹⁹

Late in the 2000 congressional session, Representatives Howard Berman and Rick Boucher introduced H.R. 5364, "The Business Method Patent Improvement Act of 2000."²⁰ H.R. 5364 sought to significantly raise the patentability bar for business method patent applications. Among other features, the bill proposed to increase the difficulty of satisfying the requirement that an invention be nonobvious, and called for post-grant

and other similarly unsettling ideas. *See generally* Leon R. Kass, *Triumph or Tragedy? The Moral Meaning of Genetic Technology*, 45 AM. J. JURIS. 1 (2000). These are quite different concerns than those regarding business method patents, but certainly help explain the vast amount of public discourse on biotechnology patents.

17. We thoroughly explore the nature and magnitude of this criticism *infra* at Part II.A.

18. First Inventor Defense Act of 1999, Pub. L. No. 106-113, 113 Stat. 1536 (codified as amended at 35 U.S.C. § 273 (2000)).

19. Though unusual, congressional action to single out a particular technology is not completely without precedent. For example, since 1996, physicians, hospitals, and other health care providers have been immunized from infringement liability for using patented medical procedures. 35 U.S.C. § 287(c) (2000). Also in 1996, Congress relaxed the standard for nonobviousness of biotechnology processes such that, if either the starting material or the product of the process is novel, the process is presumed to have been nonobvious in light of the prior art. *Id.* § 103(b). In the mid-1980's, Congress created *sui generis* protection for semiconductor mask works including concepts from both copyright and patent law. It was obsolete upon or shortly after taking effect because it had become impractical for copyists to use the proscribed methods. Semiconductor Chip Protection Act of 1984, 17 U.S.C. § 901 (2000); *see also* Steven P. Kasch, *The Semiconductor Chip Protection Act: Past, Present, and Future*, 7 HIGH TECH. L.J. 71 (1992); John G. Rauch, *The Realities of Our Times: The Semiconductor Chip Protection Act of 1984 and the Evolution of the Semiconductor Industry*, 75 J. PAT. & TRADEMARK OFF. SOC'Y 93 (1993); Pamela Samuelson, *Intellectual Property and Contract Law for the Information Age: Foreword to a Symposium*, 87 CALIF. L. REV. 1, 16 n.16 (1999). Congress has also provided for limited extensions of the patent term for pharmaceuticals to allow drug companies to recoup some of the time lost during the lengthy process of securing marketing approval from the Food and Drug Administration. 35 U.S.C. § 156.

20. Business Method Patent Improvement Act of 2000, H.R. 5364, 106th Cong. (2000).

opposition proceedings.²¹ Early in the 2001 session, Berman and Boucher introduced H.R. 1332, which was very similar to the previous year's bill.²² One day after the introduction of the second bill targeting business method patents for heightened PTO scrutiny, Congress conducted oversight hearings on business method patents that focused on whether such legislation was needed.²³

In March of 2000, the PTO—perhaps moved by the public outcry, the dread of legislation that might be difficult to implement, or the crush of the increasingly large number of business method patent applications—undertook new procedures for examining business method patents. The PTO initiative sought to hire more examiners with qualifications suited for examining business method patent applications, provide more training for such examiners, require examiners to search specifically identified prior art sources for particular subclasses of business method patents, subject business method patents having 705 as their main classification to a second-level review after allowance, and, once issued, select a sample of these patents for review by the Office of Patent Quality Review that is larger than the samples selected for periodic quality review in other technical areas.²⁴ The initiative represented a rare commitment of resources by the PTO to a single subset of the patenting enterprise.²⁵

21. Post-grant opposition provisions exist for all patents in many countries other than the United States. Such proceedings probably make much more sense in these other nations, because their courts, unlike ours, cannot rule on the validity of an issued patent. See John R. Thomas, *Collusion and Collective Action in the Patent System: A Proposal for Patent Bounties*, 2001 U. ILL. L. REV. 305, 326-42 (2001) [hereinafter Thomas, *Collusion and Collective Action*] (critiquing proposed reforms of patent system, especially relating to business method patents, and rejecting European-style post-grant oppositions in favor of a system of bounties to third parties who submit relevant prior art to the PTO).

22. Business Method Patent Improvement Act of 2001, H.R.1332, 107th Cong. (2001). No such legislation was introduced in 2002.

23. *Business Method Patents: Hearings on H.R. 1332 Before the House Comm. on the Judiciary, Subcomm. on Courts, the Internet, & Intell. Prop.*, 107th Cong. (2001) [hereinafter *Hearings on H.R. 1332*], available at http://commdocs.house.gov/committees/judiciary/hju72299.000/hju72299_of.htm (April 4, 2001) (last visited Nov. 19, 2003).

24. U.S. PAT. & TRADEMARK OFF. [hereinafter USPTO], A USPTO WHITE PAPER: AUTOMATED FINANCIAL OR MANAGEMENT DATA PROCESSING METHODS (BUSINESS METHODS) 8-22 [hereinafter BUSINESS METHODS WHITE PAPER], available at <http://www.uspto.gov/web/menu/busmethp/whitepaper.pdf>. (last visited Nov. 9, 2003). The action plan, discussed in detail at these sites, is summarized at <http://www.uspto.gov/web/offices/com/sol/actionplan.html>. USPTO, *Business Methods Patent Initiative: An Action Plan*, <http://www.uspto.gov/web/offices/com/sol/actionplan.html> (last visited Nov. 9, 2003). The sets of specifically identified prior art sources for particular subclasses of business method patents that examiners are now required to search are necessarily very

Patent quality is an elusive concept, but it essentially consists of the likelihood that a patented invention meets the requirements of novelty and nonobviousness, and thus will be found valid if challenged in litigation. Although most of the complaints about Internet business method patents have been couched in terms of perceived low quality, we believe that it is also important to examine indicators of value.²⁶

Knowing more about patent value is important for several reasons. Given the limitations of any inferences one might draw from the evidence on patent quality, additional information about patent value may buttress those inferences about quality because patent owners (patentees) are more likely to care about quality when they perceive high value. Patentees' perceptions about value should be relatively accurate. Patentees should have more relevant information about their patents' value than anyone else and, on the average, can be expected to act rationally to increase patent quality when they perceive that value is high.²⁷ Patent value is also important independent of the effect it may have on quality. For example, perceptions of patent value affect license negotiations, the decision to

incomplete, but do represent a good faith effort by the PTO to meet criticisms aimed at perceived deficiencies in the amount of nonpatent prior art cited in business method patents. See USPTO, NON-PATENT LITERATURE DATABASES FIGURES FOR WHITE PAPER ON BUSINESS METHODS PATENTS, available at <http://www.uspto.gov/web/menu/busmethp/figurenpl.htm> (last visited Nov. 9, 2003). For a general summary of post-*State Street* developments, see Peter R. Lando, *Business Method Patents: Update Post State Street*, 9 TEX. INTELL. PROP. L.J. 403 (2001).

25. The PTO has previously made special efforts to clarify patenting standards in specific technology areas. It has, for example, issued guidelines for examiners of software patents. USPTO, *Examination Guidelines for Computer-Related Inventions: Training Materials Directed to Business, Artificial Intelligence, and Mathematical Processing Applications*, <http://www.uspto.gov/web/offices/pac/compexam/examcomp.htm> (last visited Nov. 9, 2003). It has also issued examination guidelines on the utility requirement for biotechnology patents. Utility Examination Guidelines, 66 Fed. Reg. 1092 (USPTO Jan. 5, 2001), available at <http://www.uspto.gov/web/offices/com/sol/notices/utilexmguide.pdf> (last visited Nov. 9, 2003). Guidelines such as these are intended to serve an explanatory purpose for examiners and applicants. The PTO measures targeting business method patents for additional scrutiny seek to codify disparate treatment rather than to merely clarify an existing patentability requirement in a technology area.

26. We speak only of private value and not of social value. The former is the value of patents to their owners and the latter is the value of patents to society. The two are obviously correlated because the characteristics of invention novelty and nonobviousness that contribute so much to private patent value are the sine qua non of innovations that benefit the public in the long run.

27. We know that not everyone will behave this way. There may be a few patent applicants who knowingly seek and receive low quality patents with the idea of using them in an extortionate way, but we believe that our assumption holds true for most patent applicants in most circumstances.

settle or litigate an infringement claim, and the assessment of whether private or public resources expended in the patenting enterprise are being properly allocated.

Furthermore, because patent quality and value are interwoven in inextricable ways, one seeking a definition of patent value would in most cases be wise to start with patent quality. On average, value can probably be characterized as quality plus other factors. We recognize that situations exist in which a patent may be valuable to its owner despite its low quality. For example, it is possible for a single patent to have little, if any, value on its own, and even be of highly questionable validity, but to form part of a patent portfolio with a value far greater than the sum of its parts. Portfolio value can manifest itself in licensing negotiations, especially cross-licensing, or merely in the greater *in terrorem* effects it creates for competitors.²⁸

Many other variables can affect the degree to which a patent has worth to its owner. Ascertaining those factors, much less measuring them, is challenging. For instance, if other things are equal, a patent that helps launch a new technology should be more valuable than one that adds to an already mature technology. There is neither a generally accepted definition²⁹ nor a unified theory of patent value. Instead, researchers have

28. One of several manifestations of portfolio value is in a “patent thicket,” what Carl Shapiro has called “a dense web of overlapping intellectual property rights that a company must hack its way through in order to actually commercialize new technology.” Carl S. Shapiro, *Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard-Setting*, in 1 INNOVATION POLICY AND THE ECONOMY 1-2 (Adam Jaffe et al. eds., 2001), available at <http://faculty.haas.berkeley.edu/shapiro/thicket.pdf>. (last visited Nov. 19, 2003). In other words, many closely related patents may cover a single product, making it much more difficult for competitors to invent noninfringing substitutes. Patent thickets increase the probability of “hold-up” licensing, that is, exercising the ability to charge a premium for patent licenses in the case of technologies in which competitors have already invested heavily. *Id.* A patent thicket is just one instance of portfolio value, because a group of patents on related technologies can have a value greater than the sum of its parts even if the patents do not create overlapping rights in the same product. Regardless of the particular manifestation of portfolio value, previous research has not captured this aspect of patent value, and we have not ascertained a way to estimate the effect of a patent’s contribution to a portfolio apart from whatever stand-alone value it may or may not have.

29. In the course of showing how the number of times a patent is cited as a reference in later patents may be an indicator of patent value or importance, Manuel Trajtenberg stated: “By value I mean the social benefits generated by the innovation in the form of the additional consumer surplus and the profits stemming from the innovation.” Manuel Trajtenberg, *A Penny for Your Quotes: Patent Citations and the Value of Innovations*, 21 RAND J. ECON. 172, 173 n.1 (1990). Here, Trajtenberg seems to refer to both private value and social value in the same sentence.

simply proposed and tested value indicators.³⁰ We can only do the same by seeking to measure several indicators suggesting patent quality and value.

We propose and test five value indicators in comparing Internet business method patents with other types of patents: the number of prior art references, type of prior art references, number of claims within the patents, number of inventors, and time spent in the PTO before issuance. We examine the number and type of prior art references in both patent data sets for two reasons. First, the number and type of prior art references appear to relate directly to patent quality (and thus value).³¹ Second, the alleged failure of business method patents to properly cite prior art served as a centerpiece for much of the criticism of business method patents. In addition to analyzing the number and type of prior art references, we also examine the total number of claims within these patents, because number of claims is another variable that may relate to patent value.³² On average,

Economists sometimes even use the term patent “quality” as being synonymous with the private value of patents to their owners, without providing useful definitions of either quality or value. *See, e.g.*, JEAN O. LANJOUW & MARK SCHANKERMAN, THE QUALITY OF IDEAS: MEASURING INNOVATION WITH MULTIPLE INDICATORS 1 (Nat’l Bureau of Econ. Research, Working Paper No. 7345, 1999) [hereinafter LANJOUW & SCHANKERMAN, QUALITY OF IDEAS]; Mark Schankerman & Ariel Pakes, *Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period*, 96 ECON. J. 1052, 1052 (1986).

30. Josh Lerner has, for example, used the valuation of patent-dependent biotechnology start-ups by venture capital firms during multiple rounds of financing as such an indicator. Joshua Lerner, *The Importance of Patent Scope: An Empirical Analysis*, 25 RAND J. ECON. 319 (1994) [hereinafter Lerner, *The Importance of Patent Scope*]. For further discussion of patent value indicators, see John R. Allison et al., *Valuable Patents*, 89 GEO. L.J. (forthcoming Jan. 2004) [hereinafter Allison et al., *Valuable Patents*], available at http://papers.ssrn.com/sol3/delivery.cfm/SSRN_ID426020_code030722630.pdf?abstractid=426020.

31. Patent lawyers speak of prior art references, whereas economists typically speak of “backward citations” to refer to the same thing. The amount and type of prior art in a patent is associated with the extent of the applicant’s effort to differentiate its invention from what has gone before. Thorough prior art searches also are costly. Moreover, previous research has shown that prior art found by the accused infringer that was not cited in the patent and, thus, not considered by the PTO in its examination prior to issuance, is one of the most common reasons for judicial declarations of patent invalidity. *See* John R. Allison & Mark A. Lemley, *Empirical Evidence on the Validity of Litigated Patents*, 26 AM. INTELL. PROP. ASS’N Q.J. 185, 231-34, 251 (1998) [hereinafter Allison & Lemley, *Validity of Litigated Patents*] (examining litigated patents leading to final written decisions on validity or invalidity during 1989-96). The amount and type of prior art should thus be associated with patent quality and value.

32. Section 112 of the Patent Act requires that a patent application have a thorough and concise description of the invention coupled with drawings (engineering and/or con-

patents that become the subject of infringement litigation appear to be more valuable than those that do not,³³ and prior research shows that litigated patents have more claims than unlitigated ones.³⁴ There also are strong intuitive reasons to believe that the average number of claims in a group of patents is associated with patent value. Finally, we examine two other variables that may impart information about patent value: the number of inventors, and the amount of time that applications spend in the PTO before issuance as patents. Although previous research has not tested the possible relationship between the number of inventors and patent value, we believe that there is sufficient reason to suspect an association at least warrant investigation.³⁵ The amount of time an application spends in

ceptual). The description and drawings must be “enabling,” that is, they must reveal information about the invention that is sufficient to enable a hypothetical ordinarily skilled practitioner in the art to make the invention without undue experimentation. 35 U.S.C. § 112 (2000). Following the description, the application must contain “claims” that precisely delineate the intangible property interest being sought by the applicant. *Id.* The claims define the invention. When a patentee sues for patent infringement, it actually sues for infringement of one or more of the claims within the patent. The total number of claims is thus likely to be correlated to the likelihood of proving infringement.

33. Patent infringement litigation is frightfully expensive, and one would generally expect patent owners to resort to it only when the stakes are relatively high. The American Intellectual Property Law Association recently estimated the median expense of patent infringement litigation for cases with between \$1 million and \$25 million at stake as \$1.001 million through pretrial discovery and \$2 million through trial and appeal. AM. INTEL. PROP. LAW ASS’N (AIPLA), 2003 REPORT OF THE ECONOMIC SURVEY 21-22 (2003) [hereinafter AIPLA, 2003 ECONOMIC SURVEY]. Factors other than the size of the stakes also affect the propensity to litigate patents, however. *See, e.g.*, JEAN O. LANJOUW & MARK SCHANKERMAN, ENFORCING INTELLECTUAL PROPERTY RIGHTS (Nat’l Bureau of Econ. Research, Working Paper No. 8656, 2001) [hereinafter LANJOUW & SCHANKERMAN, ENFORCING INTELLECTUAL PROPERTY RIGHTS] (finding several variables to be correlated with the fact of patent litigation, and finding that the fact of litigation, but not the outcome, is significantly correlated with patent value); Jean O. Lanjouw & Josh Lerner, *The Enforcement of Intellectual Property Rights: A Survey of the Empirical Literature*, 49/50 ANNALES D’ECONOMIE ET DE STATISTIQUE 223-46 (1998) [hereinafter Lanjouw & Lerner, *Literature Survey on I.P. Rights Enforcement*] (surveying empirical studies in economics on intellectual property litigation, including the relationship between patent litigation and patent value).

34. *See, e.g.*, Jean O. Lanjouw & Mark Schankerman, *Characteristics of Patent Litigation: A Window on Competition*, 32 RAND J. ECON. 129 (2001) [hereinafter Lanjouw & Schankerman, *Characteristics of Patent Litigation*]. Another far more comprehensive study of litigated and unlitigated patents by one of the authors, John Allison, along with Mark Lemley, Kimberly Moore, and Derek Trunkey, finds that litigated patents have significantly more claims than unlitigated ones. *See generally* Allison et al., *Valuable Patents*, *supra* note 30.

35. The number of inventors may be an indicator of resources committed to the inventive endeavor, complexity of the technology, and costs associated with patenting, thus

the PTO before issuance is often referred to as “pendency time” or “time-in-prosecution.” Another previously untested variable, average pendency time is likely to reveal several types of information, including some that may relate to patent quality and perhaps value. Although the time a patent application spends in the PTO prior to issuance affects the term of patent protection, we believe that average pendency times are far more important because they reveal information about the seriousness of the applicant and the thoroughness of the PTO’s examination.³⁶

Others have proposed several additional indicators of patent value that we do not measure. For example, some have suggested that the number of International Patent Classifications (“IPCs”) into which the PTO places a patent correlates with its value.³⁷ The argument is that because each IPC defines a technology area, the more IPCs a patent is assigned to, the more technology areas it touches upon and the larger the number of potential infringers. If the average number of IPCs is a useful measure of patent scope, one might expect it to be correlated with private patent value and litigation propensity. The relevance of this variable as a measure of patent value is highly questionable, however, and we were, in any event, unable to employ it.³⁸ Others have also found that the number of different nations

suggesting a perception by the patent applicant that the invention has substantial value, such perception often affecting the reality. *See infra* Part V.D.

36. Prior to June 8, 1995, the term of patent protection was seventeen years from the date of issue. For applications originally filed after that date, the term of a patent issued from such application is twenty years from the original filing date. For patent applications pending on June 8, 1995, or for patents currently in force on that date, the patentee was given a choice of the two patent terms. 35 U.S.C. §§ 154(a)(2), (c). When a patent’s term is calculated from the original filing date, pendency time obviously affects the length of the term.

37. *See* Lerner, *The Importance of Patent Scope*, *supra* note 30, at 320-21. Professor Lerner found a positive association between the values assigned to patent-dependent biotechnology start-ups by venture capital firms in multiple rounds of financing and the average number of four-digit IPCs into which the biotechnology firms’ patents had been placed by the PTO. Lerner thus suggested an association between the number of four-digit IPCs and patent value. The PTO assigns patents to both its own classifications and to IPCs, and Lerner offers explanations for why he believes the IPC system is superior to the PTO’s own system for this purpose. *Id.*

38. *See* JEAN O. LANJOUW & MARK SCHANKERMAN, *STYLIZED FACTS OF PATENT LITIGATION: VALUE, SCOPE AND OWNERSHIP* (Nat’l Bureau of Econ. Research, Working Paper No. 6297, 1997) [hereinafter LANJOUW & SCHANKERMAN, *STYLIZED FACTS OF PATENT LITIGATION*] (finding that the average number of IPCs is *not* correlated with propensity to litigate); Dominique Guellec & Bruno Van Pottelsberghe de la Potterie, *Applications, Grants, and the Value of Patent*, 69 *ECON. LETTERS* 109 (2000), available at http://www.ulb.ac.be/cours/solvay/vanpottelsberghe/resources/Pap_Econ_letters_1.pdf (last visited Nov. 9, 2003); Dominique Guellec & Bruno Van Pottelsberghe de la Potterie,

in which patent protection has been obtained for the same invention is associated with value.³⁹ Because of the large expense required to obtain patent protection in multiple countries, an association with patent value is to be expected, but we were not able to make use of the statistic because the business method patent phenomenon is purely an American one.

We used two data sets in our research. The first was a set of 1,000 randomly selected patents of all kinds that were issued between mid-1996 and mid-1998. These data were developed by John Allison and Mark

The Value of Patents and Patenting Strategies: Countries and Technology Areas Patterns, 11 ECON. INNOVATION & NEW TECH. 133 (2002) (reasonably assuming that issued patents in Europe reflect more valuable inventions than published patent applications that were then either abandoned or finally rejected, finding that issued patents actually were characterized by a *smaller* average number of four-digit IPCs than abandoned or rejected applications). Although Professor Lerner was undoubtedly correct in finding the association he did, we believe it to have been idiosyncratic to his set of data and not generalizable. Allison et al., *Valuable Patents*, *supra* note 30 (manuscript at 12 n.36) (finding that the number of IPCs is not a significant predictor of litigation, and is actually negatively correlated with litigation propensity).

The average number of technology areas per patent probably is an indicator of patent value, but IPCs simply do not do a good job of identifying technology areas at a conceptual level for the same reasons PTO classifications do not do a good job of identifying technology areas at a conceptual level. Each classification system is designed to aid in prior art searching and thus operates at a functional rather than conceptual level, and at a very low level of abstraction. Using the same fourteen subjectively identified technology areas that we employ in this Article when we compare Internet business method patents to other patents, another study has found that the average number of technology areas per patent is highly related to the propensity to litigate, litigated patents clearly being more valuable on average than unlitigated ones. Allison et al., *Valuable Patents*, *supra* note 30. It was found in that study that litigated patents averaged 1.99 technology areas per patent, and unlitigated patents 1.59, a difference that is statistically significant far beyond 99% ($p < 0.0001$).

Even if the average number of IPCs were a value indicator, in our study it is not meaningful to compare the number of IPCs into which Internet business method patents have been placed with the average patent in all technology areas because business method patents all lie in a very narrow technical field whereas the “average” patent often will involve multiple areas of technology. It likewise is not useful to use this indicator for comparing Internet business method patents with patents in each of our defined technology areas because the former also are innately narrower than any of the latter.

39. See, e.g., LANJOUW & SCHANKERMAN, *QUALITY OF IDEAS*, *supra* note 29, at 12. Economists have referred to the existence of protection in multiple nations as a “patent family.” Patent Information Users Group, Inc., *Patent Families*, <http://www.piug.org/patfam.html> (last visited Dec. 10, 2003). What they measure, however, is geographic scope, and the term “patent family” probably would be better used to describe multiple patents on closely related inventions rather than patents in multiple countries on the same invention.

Lemley for use in two previous studies,⁴⁰ although more data were added for the instant study. The second data set of Internet business method patents was developed especially for this study and mirrors the data categories from the set of general patents. The data set of business method patents consists of those that were clearly aimed at use with the Internet.⁴¹ The characteristics of these two data sets, as well as our data mining techniques, are more fully described below.

Using bivariate and multivariate statistical methods,⁴² we obtained statistically significant findings that both defied conventional wisdom and surprised us. Our findings are listed below:

40. John R. Allison & Mark A. Lemley, *The Growing Complexity of the United States Patent System*, 82 B.U. L. REV. 77 (2002) [hereinafter Allison & Lemley, *Growing Complexity*] (comparing many variables in a randomly selected set of one thousand patents issued during a two-year period in 1996-98 with counterpart variables in an identical sample drawn from patents issued during a two-year period twenty years earlier, revealing that the patenting enterprise had grown more complex over twenty years in all dimensions); John R. Allison & Mark A. Lemley, *Who's Patenting What? An Empirical Exploration of Patent Prosecution*, 53 VAND. L. REV. 2099 (2000) [hereinafter Allison & Lemley, *Who's Patenting What?*] (exploring empirically, among many other things, the types of entities receiving patents, the geographic origin of patents, and technology areas in which patents are granted).

41. Although not all business method patents are so targeted, we chose to collect those that are Internet-related because (1) the great bulk of patents one may characterize as covering business methods are Internet-related; (2) the business method patents receiving the most attention and criticism in recent years are Internet-related; and (3) any relevant policy concerns about the quality of business method patents are allegedly most acute in the case of those aimed at Internet usage because of their potential for unduly propertizing and thus creating stenoses in the Internet. See LAWRENCE LESSIG, *THE FUTURE OF IDEAS: THE FATE OF THE COMMONS IN A CONNECTED WORLD* 204-17, 316-24 (2001) [hereinafter LESSIG, *FUTURE OF IDEAS*] (arguing that the granting of patents on Internet-related software threatens to fence off large areas of the Internet); see also JAMES BOYLE, *SHAMANS, SPLEENS, AND SOFTWARE* (1996) (offering a wide-ranging objection to the growing propertization of information).

42. Bivariate methods, such as the independent groups *t*-test or *z*-test for comparing means from two samples, seek to ascertain whether differences in a variable in two data sets are statistically significant. See, e.g., WILLIAM E. BECKER, *STATISTICS FOR BUSINESS & ECONOMICS* 400-46 (1995). Multivariate methods are used to simultaneously analyze data on several variables. See, e.g., AMIR D. ACZEL, *COMPLETE BUSINESS STATISTICS* 895-963 (1989). Thus, when we compare several variables in one sample or population with counterpart variables in another sample or population, we use bivariate methods to determine whether there are statistically significant differences between a given variable in one set and its counterpart variable in the other set. When we speak of a statistically significant difference, we use the standard "95% confidence" level; in other words, a difference is viewed as statistically significant if we are at least 95% confident that the difference is not due to chance. Stated somewhat differently, we report statistical significance when the *p*-value is 0.05 or less. In all instances in which we report statistical sig-

- First, with respect to prior art, Internet business method patents had significantly more patent references, nonpatent references, and total references than patents in general. They also had significantly more of all types of references than patents in most of the fourteen individual technology areas within the general patent data set that we identified.⁴³
- Second, although it was impossible to directly measure the relevance or informational value of prior art references, we developed a taxonomy of *nonpatent* prior art that allowed us to make rough comparisons of the likely quality of these nonpatent references in the two data sets. Internet business method patents appear to have cited nonpatent prior art of a similar quality to that in the average patent.⁴⁴
- Third, Internet business method patents had significantly more claims than patents in general, and significantly more than in all but one individual technology area.⁴⁵
- Fourth, Internet business method patents had significantly more inventors than patents in general, although this finding is weaker than the findings pertaining to references and claims.
- Fifth, applications for Internet business method patents spent much more time in the PTO's examination process than we expected.⁴⁶ There is thus no evidence that Internet business method patent

nificance, the *p*-value is actually far below 0.05. Multivariate methods then provided general confirmation of the bivariate results by revealing that significant differences between variables in the two data sets were not rendered insignificant by interactions among variables within each data set.

43. We emphasize that the Internet business method patents in our data set were all issued before the PTO initiated its program of greater scrutiny for certain business method patent applications.

44. Only chemistry, semiconductor, and pharmaceutical patents out of our fourteen technology areas appear to have cited higher quality prior art than Internet business method patents, and only pharmaceutical patents cited prior art that is better when taking into account both quantity and quality.

45. Internet business method patents actually had substantially more claims than patents in this technology area, acoustics, but the fact that this area included a relatively small number of patents and that acoustics patents did have more claims than the other technology areas caused the difference to be statistically insignificant.

46. For several reasons, we anticipated that other kinds of patents would have had much longer pendency times than Internet business method patents, especially when measuring from the original priority filing date. These reasons are explained *infra* in Part V.E

applications received less examination than other patents, or that their owners committed fewer resources to obtain them.⁴⁷

Overall, our data demonstrate that Internet business method patents were no worse than patents in general in the late 1990s. Indeed, our empirical evidence suggests that they may have been better than average.⁴⁸

Why, then, have these patents been so widely disparaged and singled out for disparate treatment? Stated differently, why is there a wide gap between the conventional wisdom and the apparent reality shown by hard data? There are several possible explanations.

One possibility is that all of the alleged deficiencies in Internet business method patents resulted from the patent system trying to cope with a new subject matter that had been recognized only recently as patentable. Although superficially appealing, such an argument quickly collapses upon further consideration. We studied Internet business method patents when they were still new—patents issued through the end of 1999—and found that they compared favorably with patents generally.

One might also argue that the approach used by the PTO and the courts for addressing the requirement of “nonobviousness” provides an explanation.⁴⁹ The mere fact that an invention appears trivial after a quick look does not mean that it is obvious for patentability purposes.⁵⁰ The PTO or a challenger in litigation must produce prior art showing that the invention represents only an obvious advance over the prior art. This legal attitude toward the nonobviousness requirement may explain why we find patents in all areas that appear questionable, but it does not support an

47. The average patent did spend slightly more time in the PTO than did Internet business method patents, but the significance of the difference disappeared when we accounted for the influence that the number of references and claims have on pendency time. Moreover, patents in only four of the fourteen technology areas had longer pendency times. When we removed the effect of multiple applications in a chain and considered only those applications leading most directly to the patents in question, Internet business method patents surprisingly spent a lot more time in the PTO than the average patent and than those in most technology areas. These differences remained statistically significant even after taking account of the effect that the number of references and claims each have on pendency times.

48. If there are serious problems with the quality of issued U.S. patents, they appear to be systemic rather than specific to a certain technology. We realize that this statement has negative implications for the patent system as a whole, but examining the entire system or the work quality of the PTO is a task left for others.

49. 35 U.S.C. § 103 (2000).

50. *See, e.g.,* *Panduit Corp. v. Dennison Mfg. Co.*, 774 F.2d 1082, 1093-96 (Fed. Cir. 1985), *vacated by* 475 U.S. 809, 810 (1986), *remanded to* 810 F.2d 1561 (Fed. Cir. 1987).

inference that patents in one area are inferior to those in other areas. Many have contended that more patents were issued on obvious Internet business methods than in other areas because many of these methods had been previously practiced but kept secret so that little accessible prior art was created. Not only was this intuitively attractive argument not borne out by the data, but it is far from clear that prior secret innovations are substantially more common in the area of business methods than in other areas.⁵¹

The possibility also exists that our analysis of prior art cited in patents overemphasizes quantitative measures and fails to adequately account for variations in the relevance and content quality of those references. We will show, however, that the quantitative measures we use say a great deal about patent quality and value. We also attempt a reasonable assessment of the caliber of the prior art cited in the patents we study. A complete qualitative study of the cited prior art takes money and time, and would require the employment of experts in every technology area and sub-area.

A final possibility, which we believe to be the most likely, is that negative opinions about business method patents and consequent actions to further spread this impression are the subject of a classic information cascade. Such phenomena are pervasive in all parts of society and can be found in various fads and fashions, in riots and mobs, and in the herd behavior of stock market investors.⁵²

51. Despite rapidly increasing rates of patenting in practically all industries, most research and development executives view trade secrets and other means as superior to patents in appropriating returns on R&D investment. See WESLEY M. COHEN ET AL., PROTECTING THEIR INTELLECTUAL ASSETS: APPROPRIABILITY CONDITIONS AND WHY U.S. MANUFACTURING FIRMS PATENT (OR NOT) 9-11 (Nat'l Bureau of Econ. Research, Working Paper No. 7552, 2000) [hereinafter COHEN ET AL., WHY U.S. MANUFACTURING FIRMS PATENT] (showing that the importance of patents as a means for appropriating returns from research and development investments varies among industries and that patents are among the least favorite means in almost all of them despite the fact that patenting activity has increased dramatically); Richard C. Levin et al., *Appropriating the Returns from Industrial Research and Development*, in 1987 BROOKINGS PAPERS ON ECON. ACTIVITY 783, 794-97 (1987) (showing, in a seminal study, that the importance of patents as a means for appropriating returns from R&D investments varies among industries, and that they are among the least favorite means in most of them).

52. They go by many other names, as well, such as "bandwagon effects," and occur when a critical mass of information is reached and leads to rapid acceptance of a norm or belief and similar action by large numbers of people. People's own actions can come to depend to an increasing degree on the information conveyed by the actions of others. The phenomenon can occur regardless of whether the information on which it is based is correct or incorrect. See, e.g., Sushil Bikhchandani et al., *A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades*, 100 J. POL. ECON. 992 (1992)

In a closely related vein, the “confirmation” bias, which has been found by researchers in behavioral decision theory to be a pervasive psychological phenomenon, may have played a part in perpetuating the view that Internet business method patents were uniquely bad.⁵³ Further supporting the idea that surface perceptions may have overwhelmed the public debate is the fact that Internet business methods probably require much less capital investment and start-up time to reach the marketplace than many other inventions. As a result, by the time a business method patent issues, the product may already be quite familiar to the public. Such early familiarity may contribute to widespread perceptions that the method is obvious.

Part II of this Article outlines the most frequently voiced criticisms of Internet business method patents and notes the unusual volume and intensity of these complaints, especially in the popular, business, and legal press. Part III outlines steps taken by Congress and the PTO in response to the widespread panning of business method patents. Part IV describes our data mining techniques and the two data sets we developed for comparison. Part V describes the patent attributes we measured, the results of our statistical analyses, and the import of these results. Part VI offers some possible reasons for the differences between our findings and the general consensus that business method patents were worse than other kinds, including the likelihood that excoriating them simply became a fad. Part VII offers some concluding observations. We conclude that actions to single out Internet business method patents for special treatment were not only unjustified, but also were probably futile and counterproductive.

[hereinafter Bikhchandani et al., *Theory of Fads*] (showing how information cascades can often be mistaken even though individuals are acting rationally, how a few early individuals can have a disproportionate effect on the development of a cascade, and how people’s beliefs and actions can depend to an increasing degree on information learned from the decisions of others).

53. Decision-making “heuristics,” or short cuts, may be the most efficient means for processing information in many circumstances, but forming beliefs and making decisions based on beliefs so derived can also create systematic decision-making biases. Regarding the “confirmation bias,” once relatively firm beliefs or theories are formed, most people tend to pay more attention to and indeed seek out further information that confirms preexisting beliefs. Such an approach is not the logical way to test a belief (or a hypothesis in statistical testing) and is diametrically opposed to the scientific method. Sometimes this tendency is referred to as a logical fallacy that produces “belief perseverance.” See, e.g., ROBIN M. HOGARTH, EDUCATING INTUITION 119-21 (2001) (discussing the confirmation bias). Once firm beliefs were formed that business method patents were either inferior or should never have been recognized as protectible, many may have clung to those beliefs and paid far more attention to subsequent evidence confirming them, even if further evidence consisted merely of the unsubstantiated opinions of others.

II. CRITICISMS OF BUSINESS METHOD PATENTS

A. The Diverse Nature of the Outcry

Business method patents antedated their official 1998 recognition by the Federal Circuit.⁵⁴ However, the sprinkle became a deluge after *State Street*.⁵⁵ Along with the accelerated patenting of business methods came a great deal of attention in the popular, legal, and business press, as well as in academic circles. In general, these patents received poor reviews. The substance of the criticism has ranged from the plausible to the puzzling, and the rhetorical means from the cautiously alarmed to the polemical. Some observers simply viewed business processes as an inappropriate subject matter for patent protection. A notable theme in this strain of criticism was that allowing such patents opened the door for patenting new developments in all of human experience, and that patents should remain grounded in science and engineering, a theme expressed especially well by John R. Thomas:

Each issue of the Patent Office Gazette seems to include proprietary processes from an unlikely collection of disciplines. Although we once might have relegated these claims to some popular compilation of unusual patents, the Federal Circuit opinion in *State Street* has imbued them with newfound vitality. With the Patent Office open for patents on business methods, the frontiers of the patent system appear virtually without limit. The patent system now seems poised to impact callings ranging from the arts, to the social sciences, to the law itself.

There is much to commend the adoption of the standard of industrial application in the United States patent law. Our patent law should comport with our perception of what technology is, not defy it. Restoring a patentability standard firmly grounded in industrial applicability, rather than equating technology with anything artificial, would enable us to maintain the integrity of our current patent system. Moreover, it would enable us to respect the boundary between the whole expression of our

54. See BUSINESS METHOD WHITE PAPER, *supra* note 24, at 8, 10; see, e.g., U.S. Patent No. 5,664,115 (issued Sept. 2, 1997) ("Interactive Computer System to Match Buyers and Sellers of Real Estate, Businesses and Other Property Using the Internet"); U.S. Patent No. 5,664,110 (issued Sept. 2, 1997) ("Remote Ordering System") (an Internet-based system for building and using order-lists for remotely ordering goods and services).

55. See, e.g., Julia Angwin, 'Business Method' Patents, *Key to Priceline, Draw Growing Protest*, WALL ST. J., Oct. 3, 2000, at B1; see generally Meurer, *supra* note 10; Rai, *supra* note 10, at 211; Maeir, *supra* note 10.

humanity and that small part of it that is properly called technological. However central to contemporary life and worthy of nurturing through the patent system, technology is but one manifestation of the human experience.⁵⁶

Others have made similar arguments. After an exhaustive search for a definition of “technology,” Alan Durham concluded that even if the art of computer programming is within the definition and should be treated as patentable subject matter, a software-embodied business method is not and should not.⁵⁷ Similarly, the well-known science writer James Gleick observed that business method patents (as well as software patents generally) had crossed a crucial boundary between “a substantial, tangible, nuts-and-bolts world . . . into the realm of thought and abstraction.”⁵⁸ Rochelle Cooper Dreyfuss opined against the patentability of business methods for slightly different reasons. Dreyfuss argued that, contrary to the situation in many other fields, patents are unnecessary to create incentives to innovate in business methods, and thus there is no social benefit to offset the social cost of exclusive rights.⁵⁹ Commentators have also expressed particular concerns about lock-in effects and network

56. Thomas, *Liberal Professions*, *supra* note 11, at 1185.

57. Durham, *supra* note 11, at 1525-28; *see also*, Joseph H. Sommer, *Against Cyberlaw*, 15 BERKELEY TECH. L.J. 1145, 1220 (2000) (stating in passing that business methods have nothing to do with technology, and that the patent system is going beyond technological innovation to protect social innovation). For allusions to the nontechnological nature of business method patents, rather than arguments, *see* Julie E. Cohen & Mark A. Lemley, *Patent Scope and Innovation in the Software Industry*, 89 CALIF. L. REV. 1, 27 n.99 (2001) [hereinafter Cohen & Lemley, *Patent Scope and Innovation*] (claiming “[t]he floodgates for non-technological patents were opened by *State Street* . . . which allowed the patenting of pure business methods. A number of patents had already issued for such non-technological concepts as methods of holding a golf putter, however.”).

58. James Gleick, *Patently Absurd*, N.Y. TIMES, Mar. 12, 2000, § 6 (Magazine), at 44; *see also* Angwin, *supra* note 55, at B1 (quoting Tim O’Reilly stating that business method patents give exclusive rights to pure ideas).

59. Rochelle Cooper Dreyfuss, *Are Business Method Patents Bad for Business?*, 16 COMPUTER & HIGH TECH. L.J. 263, 274-77 (2000). For similar arguments, *see*, for example, Jared Earl Grusd, *Internet Business Methods: What Role Does and Should Patent Law Play?* 4 VA. J.L. & TECH. 9 (1999); Robert A. Kreiss, *Patent Protection for Computer Programs and Mathematical Algorithms: The Constitutional Limitations on Patentable Subject Matter*, 29 N.M. L. REV. 31, 84 n.358 (1999); Leo J. Raskind, *The State Street Bank Decision: The Bad Business of Unlimited Patent Protection for Methods of Doing Business*, 10 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 61, 90-93, 101-02 (1999).

externalities stemming from Internet business method patents.⁶⁰ Arti Rai has said that, because of the relatively short Internet business cycle, first-

60. A lock-in effect may exist if a product or service user faces substantial switching costs, that is, costs of changing from one product or service to another. This may deter switching to a competitor, or even deter entry by a potential competitor, even if the latter offers better quality or a lower price. It has been posited that a user may be reluctant to switch to another Internet merchant because of the tedium of entering one's credit card and other information again at a competitor's web site. See Dreyfuss, *supra* note 59, at 271; Thomas, *Collusion and Collective Action*, *supra* note 21, at 305, 320-21. We fail to see much in the way of sunk costs in using an Internet merchant that would significantly dissuade a user from switching.

Network externalities exist as a positive effect on the incremental value of a network—such as a telephone system, real estate multiple listing service, or operating system software—as additional users participate in the network. Such networks become more valuable to each user as the number of users increases. Although network economic effects are generally positive, they can tend toward monopolistic results when the network's basic technical standards are closed, as by software copyright protection, patent protection, or other standards-closing mechanisms. See, e.g., Dreyfuss, *supra* note 59, at 271; Thomas, *Collusion and Collective Action*, *supra* note 21, at 305, 318-21. There is little question that a proprietary network gives greater control than a proprietary non-network technology, because competing alternatives to proprietary technical standards are much less likely to develop when large positive externalities draw users into the network and ultimately make them highly dependent on it. One only has to consider Microsoft's operating systems, the code for which is vigorously protected by copyright. On network externalities more generally, see Mark A. Lemley & David McGowan, *Legal Implications of Network Economic Effects*, 86 CALIF. L. REV. 479 (1998); Howard A. Shelanski & J. Gregory Sidak, *Antitrust Divestiture in Network Industries*, 68 U. CHI. L. REV. 1 (2001).

Opponents of patenting software in general, or software-embodied business methods in particular, would argue that the Internet would not have developed as it did had its basic standards been patented. We first note that these standards were technical software communications protocols, and not business methods. It is true, however, that if software generally had not reached patentable status, neither would software-embodied business methods. We agree that the Internet would have developed differently if these early protocols had been someone's personal property, and probably would not possess the marvelous end-to-end architecture that makes it what it is. Had today's standards of software patentability existed during the years leading up to development of the Internet, however, the basic protocols would not necessarily have been patented. Mark Lemley observes that some of the Internet's original standards were developed by government, and some were collaborative ventures among many different educational institutions such that sorting out patent rights would have been very difficult if not impossible. E-mail from Mark A. Lemley, Professor of Law, University of California at Berkeley, Boalt Hall School of Law, to John R. Allison (July 1, 2003) (on file with author). He also notes that many standards in telecommunications today are open because standards-setting groups agree to keep them nonproprietary. *Id.* Patents could be a cost to the Internet, Lemley believes, but it is far from certain. *Id.* He also notes that many standards in telecommunications today are open because standards-setting groups agree to keep them nonproprietary. *Id.*

mover advantages may cause more economic harm than such advantages cause in other markets.⁶¹ Robert Merges stated that, although it was too early to tell whether business method patents as a class are worthwhile, “the increased volume of patent applications stemming from this newly patentable subject matter has pushed the patent system into crisis,”⁶² and Professor Thomas argued that low quality business method patents were further damaging the public perception of the patent system.⁶³ One *Forbes Magazine* writer assumed that business method patents are bad and circularly attributed their badness to a PTO overtaxed by the huge number of business method patent applications.⁶⁴ Others asserted that business methods are bad without providing any rationale,⁶⁵ and one even argued that granting business method patents is unconstitutional.⁶⁶

Recently, Michael Meurer argued that patent flooding in the area of business methods has already occurred and will get worse. This particular flood will produce more serious economic harm than similar phenomena in other fields, such as biotechnology, because previous patent floods were triggered by technological breakthroughs not present in business methods.⁶⁷ Even more recently, Richard Posner observed that business

61. Rai, *supra* note 10, at 212. The reasoning here, which is related to concerns about lock-in effects, is that the relatively short duration of Internet business cycles prevents the market from having enough time to correct the anticompetitive consequences of ill-considered business method patents even if they are later deemed to be invalid.

62. Robert P. Merges, *One Hundred Years of Solicitude: Intellectual Property Law, 1900-2000*, 88 CALIF. L. REV. 2187, 2232 (2000) [hereinafter Merges, *One Hundred Years of Solicitude*].

63. See Thomas, *Collusion and Collective Action*, *supra* note 21, at 322 (“Another worrisome trend is that public perception of the patent system is in a rapid tailspin. Citing proprietary rights in electronic commerce concepts such as the one-click patent, major newspapers and magazines have once more found the patent system a convenient target of scathing criticism.”).

64. Philip E. Ross, *Patently Absurd*, FORBES, May 29, 2000, at 180.

65. See, e.g., Larry A. DiMatteo, *The New “Problem” of Business Method Patents: The Convergence of National Patent Laws and International Internet Transactions*, 28 RUTGERS COMPUTER & TECH. L.J. 1, 22-25 (2002).

66. Malla Pollack, *The Multiple Unconstitutionality of Business Method Patents: Common Sense, Congressional Consideration, and Constitutional History*, 28 RUTGERS COMPUTER & TECH. L.J. 61 (2002) (attacking business method patents as unconstitutional because (1) common sense shows that they do not promote progress, (2) Congress has not said that business method patents promote progress, (3) the term “useful arts” as used in the Constitution’s intellectual property clause does not contemplate business method patents, and (4) the generation that ratified the Constitution would have considered business method patents to be “an abuse of the rights of Englishmen”). Although we believe these to be really strange arguments, dissecting or challenging these or other arguments is not our purpose despite any faulty fact or logic they may display.

67. See Meurer, *supra* note 10, at 321.

method patents create the potential to obtain enormous monopoly power and generate a reward greatly in excess of the cost of the invention.⁶⁸ Lawrence Lessig asserted that software development projects are being rendered more costly because of business method and other software patents,⁶⁹ and that business method patents work to the disadvantage of small businesses by imposing additional legal expenses that they can ill afford.⁷⁰ Both Lessig and James Boyle cited business method patents as one of the prime threats to the Internet and the public domain,⁷¹ and Lessig predicted that business method patents, along with expanded copyright protection, will devastate the Internet, turning it into a big business-controlled institution “that will disable once again the independent, unaffiliated, critical, questioning creativity that the Internet of the last ten years produced.”⁷² Comparing those attempting to enforce them to Mafiosi,⁷³ and calling for a moratorium on their issuance,⁷⁴ Lessig also

68. Richard A. Posner, *The Law & Economics of Intellectual Property*, DAEDALUS, Spring 2002, at 5. Posner apparently was serious about the totally implausible assertion pertaining to monopoly power. See also Declan McCullagh, *Left Gets Nod From Right on Copyright Law*, CNET NEWS.COM, Nov. 20, 2002 (reporting on lecture by Posner to the American Enterprise Institute in which he criticized recent expansions of copyright law and attacked the granting of “very questionable” business method patents), available at <http://news.com.com/2100-1023-966595.html>.

69. *Controlling the 'Net: How Vested Interests Are Enclosing the Cybercommons and Undermining Internet Freedom*, MULTINATIONAL MONITOR, Mar. 1, 2002, at 23 (interview with Lawrence Lessig).

70. Daniel S. Levine, *One on One with Lawrence Lessig, Author*, SAN FRAN. BUS. TIMES, Nov. 30, 2001, available at <http://sanfrancisco.bizjournals.com/sanfrancisco/stories/2001/12/03/newscolumn10.html>.

71. LESSIG, *FUTURE OF IDEAS*, *supra* note 41; James Boyle, *Fencing Off Ideas: Enclosure & the Disappearance of the Public Domain*, DAEDALUS, Mar. 22, 2002, at 13; see also Anjana Ahuja, *Patently It Doesn't Work*, LONDON TIMES, Features, Mar. 22, 2000 (quoting Lawrence Lessig that business methods are a disaster and the biggest threat to innovation in cyberspace).

72. Lawrence Lessig, *The Death of Cyberspace*, 57 WASH. & LEE L. REV. 337, 345-47 (2000). Lessig has expressed the same sentiments elsewhere many times. See, e.g., Lawrence Lessig, *Foreword*, 52 STAN. L. REV. 987 (2000).

73. Lawrence Lessig et al., *Criticism of the US Patent & Trademark Office*, Transcript: *Weekend All Things Considered* (National Public Radio, Apr. 20, 2002). Lessig has, however, also made the far more sensible statement that the major problem today is the federal government's role in over-protecting intellectual property generally. See Lawrence Lessig, *The Problem With Patents*, *Industry Standard*, Apr. 23, 1999, previously available at <http://www.thestandard.com/article/display/0,1151,4296,00.html> (printed copy on file with authors). From what we have discovered in our data, he is on less firm ground in the same opinion piece when he asserts that business method patents provide the clearest example of bad patents. *Id.*

74. Lawrence Lessig, *Online Patents: Leave Them Pending*, WALL ST. J., Mar. 23, 2000, at A22. Lessig took swipes at business method patents at many other times and

referred to business method patents as “new monster[s] called forth from an old statute.”⁷⁵

B. A Bit of Software Patent Déjà Vu

Software patents as a genus had been targeted for patentable subject matter criticism several years before the business method species became news. Some of these jeremiads, such as the assertion that software is not technology⁷⁶ and that patents were unnecessary to create productive incentives in the software industry,⁷⁷ foreshadowed similar censures of business method patents. Pamela Samuelson stated that the most profound problem with using patent law to protect software is that innovation in the field is usually accomplished in increments too small to be viewed as inventive steps.⁷⁸ On the whole, however, complaints about software as patentable subject matter were less one-sided, less varied, and less strident than later rebukes of business method patents.⁷⁹

Detractors of software patents expressed similar wariness about the adequacy of prior art. The idea that software patents were insufficiently supported by prior art references was proffered early on by some members

places, as well, sometimes alongside criticisms of software patents generally. *See, e.g.*, Jesse Walker, *Cyberspace's Legal Visionary: Lawrence Lessig on the fate of copyrights and computer networks in the digital future*, REASONONLINE, June 2002, available at <http://reason.com/0206/fe.jw.cyberspaces.shtml> (last visited Nov. 21, 2003).

75. Lawrence Lessig, *Patent Problems*, THE INDUSTRY STANDARD, Jan. 21, 2000, available at <http://www.lessig.org/content/standard/0,1902,8999,00.html> (last visited Nov. 21, 2003).

76. *See, e.g.*, Gleick, *supra* note 58 (observing that patents on both software generally and software-embodied business methods showed that patents were now being issued on thought and abstraction). *But cf.* Durham, *supra* note 11, at 1525-28 (stating that software programming is technology, but business methods are not).

77. *See, e.g.*, Pamela Samuelson, *Benson Revisited: The Case Against Patent Protection for Algorithms and Other Computer Program-Related Inventions*, 39 EMORY L.J. 1025, 1142-43 (1990) (contending that the most powerful argument against software patentability was that “the industry has grown from being a nonexistent industry to a major, flourishing, and highly innovative industry without patent protection”).

78. Pamela Samuelson et al., *A Manifesto Concerning the Legal Protection of Computer Programs*, 94 COLUM. L. REV. 2308, 2344-46 (1994) [hereinafter Samuelson et al., *Manifesto*].

79. Although opinion on the wisdom of patenting software was mixed, opponents were clearly in the majority. *See, e.g.*, Pamela Samuelson, *Once Again, Patents*, 35 COMMUN. OF THE ACM 15 (1992) (noting correctly that her negative view of software patents was in the ideological mainstream, and then arguing that software is sufficiently different from previous technologies, and that Congress should act after having heard from economists, computer professionals, and others with varied perspectives). For an early minority view that software should be patentable subject matter, see Donald S. Chisum, *The Patentability of Algorithms*, 47 U. PITT. L. REV. 959 (1986).

of the industry and the technical academic community,⁸⁰ and later by many others. Julie Cohen's observations in 1995 are illustrative:

[I]n the field of computers and computer programs, much that qualifies as prior art lies outside the areas in which the PTO has traditionally looked—previously issued patents and previous scholarly publications. Many new developments in computer programming are not documented in scholarly publications at all. Some are simply incorporated into products and placed on the market; others are discussed only in textbooks or user manuals that are not available to examiners on line. In an area that relies so heavily on published, “official” prior art, a rejection based on “common industry knowledge” that does not appear in the scholarly literature is unlikely.⁸¹

Contemporaneously with Professor Cohen, Kenneth Dam noted the prior art problem with software patents in the course of a theoretical economic analysis of intellectual property rights in software.⁸² Although numerous others have concurred,⁸³ they never provided empirical support

80. Randall Davis, professor of computer science at the Massachusetts Institute of Technology, chaired a broadcast panel discussion of software patents on Oct. 30, 1990, at MIT in which both the inherent patentability of algorithms and prior art concerns were significant topics of discussion. Michael D. Ernst, *Intellectual Property in Computing: (How) Should Software Be Protected? An Industry Perspective* (May 1992), available at <http://pag.lcs.mit.edu/~mernst/pubs/> (last visited Nov. 28, 2003). One of the panelists, Mitchell Kapor, founder of Lotus Development Corporation, which was at the time embroiled in software copyright litigation regarding its Lotus 1-2-3 spreadsheet software, observed: “The real problem is not in copyright; it’s in patents So what we’ve got is a kind of Bhopal of software patents. There are all of these terrible patents that should never have been issued because there’s prior art.” *Id.*

81. Cohen, *Electronic Vigilantism*, *supra* note 13, at 1178.

82. Kenneth W. Dam, *Some Economic Considerations In The Intellectual Property Protection of Software*, 24 J. LEGAL STUD. 321, 369 (1995).

83. See, e.g., Thomas P. Burke, *Software Patent Protection: Debugging the Current System*, 69 NOTRE DAME L. REV. 1115, 1164-65 (1994); Cohen & Lemley, *Patent Scope and Innovation*, *supra* note 57, at 12; Lemley, *Rational Ignorance*, *supra* note 4, at 1495-96 & nn.1-2 (2001); Mark A. Lemley & David W. O’Brien, *Encouraging Software Reuse*, 49 STAN. L. REV. 255, 300-04 (1997); Peter S. Menell, *The Challenges of Reforming Intellectual Property Protections for Computer Software*, 94 COLUM. L. REV. 2644, 2653 (1994); Maureen A. O’Rourke, *Toward a Doctrine of Fair Use in Patent Law*, 100 COLUM. L. REV. 1177, 1240-41 (2000); Samuelson et al., *Manifesto*, *supra* note 78, at 2363 & nn.221-24; Chad King, Note, *Abort, Retry, Fail: Protection for Software-Related Inventions in the Wake of State Street Bank & Trust Co. v. Signature Financial Group, Inc.*, 85 CORNELL L. REV. 1118, 1172 & n.336 (2000); Garfinkel, *supra* note 13; Brian Kahin, *The Software Patent Crisis*, TECH. REV., Apr. 1990, at 52; Joseph Milizzo, *Computer Industry Fears Patent Wave*, COMP. & SOFTWARE NEWS, May 22, 1989, at 4; Rachel Parker, *Refac’s Unworthy Patent May Rally the Rest of the Industry*, INFOWORLD,

for the assertion that software patents cite less prior art than other kinds of patents.⁸⁴

Critics have pointed to problems with the PTO classification system as adding to the prior art problems allegedly plaguing software patents.⁸⁵ It is true that patents on software have been strewn among many PTO classifications. There appear to be two reasons. First, PTO classifications and subclassifications focus on detailed function, not conceptual definition. They focus at a very low level of abstraction on what inventions achieve rather than on how they achieve it, giving PTO classifiers no reason to conceptually identify an invention as covering software.⁸⁶ Second, and relatedly, a great many inventions touch upon multiple technology areas. These difficulties with using PTO classifications clearly are not limited to software.

Aug. 7, 1989, at 42; Richard Stallman & Simson Garfinkle, *Against Software Patents; Software Should Be Copyrighted, Not Patented*, COMM. OF THE ACM, Jan. 1992, at 17; Jeff Ubois, *The Case Against Software Patents*, MIDRANGE SYSTEMS, Jan. 7, 1992, at 44.

84. Only commentator Gregory Aharonian, a well-known critic of the PTO generally and software patents particularly, has amassed data on the issue. On his website and through his e-mail patent newsletter, one can see that for a number of years he has consistently excoriated software patents for the inadequacy of their prior art, especially nonpatent prior art. See Gregory Aharonian, *Bad Patents Home Page*, <http://www.bustpatents.com> (last visited Nov. 30, 2003). Instructions for subscribing to his newsletter are at <http://www.bustpatents.com/ipns.htm> (last visited Nov. 30, 2003). In February 2003, he began sending his newsletter from patnews@ns1.patenting-art.com. In 1998, he reported in his newsletter that software patents averaged fewer than five references to the prior art. *17,500 Software Patents to Issue in 1998*, INTERNET PATENT NEWS SERVICE (Gregory Aharonian). Oct. 18, 1998, available at <http://lpf.ai.mit.edu/Patents/ipns/ipns-19981018.txt> (printed copy on file with authors) (last visited Nov. 30, 2003); see also Robert P. Merges, *As Many as Six Impossible Patents Before Breakfast: Property Rights for Business Concepts and Patent System Reform*, 14 Berkeley Tech. L.J. 577, 589 & n.31 (1999) [hereinafter Merges, *Six Impossible Patents*] (referring to this statistic from Aharonian's Oct. 18, 1999 newsletter). Ultimately, however, Aharonian's data appeared to combine disparate categories into larger ones, introducing data comparability problems too serious for reliable statistical analysis. Our present study does not examine software patents overall, so we cannot comment on this finding, but we certainly did not find similarly small numbers of prior art citations in Internet business method patents.

One study of the characteristics of patents in general that also included a category for software patents, found that software patents cited somewhat less *patent* prior art than the average patent, more *nonpatent* prior art, and about the same amount of *total* prior art when compared with the average patent and with patents in most other areas of technology. Allison & Lemley, *Who's Patenting What?*, *supra* note 40, at 2131-32, 2149, 2158-60.

85. See, e.g., Garfinkel, *supra* note 13.

86. The International Patent Classifications ("IPCs") operate in the same manner, and, as a consequence, neither system works well as a means for identifying areas of technology.

C. Prior Art-Related Criticisms Migrating from Software to Business Method Patents

Business method patents likewise have been repeatedly indicted for failing to cite sufficient prior art. These patents may have been more susceptible than software patents to prior art-based criticisms because of the belief, true or not, that much evidence of prior business practices is unlikely to have found its way into written sources. By and large, it is said, these things have just been “done,” and if a company perceived that a particular business practice gave it a competitive advantage, it would have protected it as a trade secret. As with software patents, only a few writers have initially offered reasoned arguments along these lines,⁸⁷ and many others have simply agreed.⁸⁸ Charles Cella, a former patent attorney, Jeff

87. See Dreyfuss, *supra* note 59, at 268-69; Rebecca S. Eisenberg, *Analyze This: A Law and Economics Agenda for the Patent System*, 53 VAND. L. REV. 2081, 2090 (2000); Merges, *Six Impossible Patents*, *supra* note 84, at 577, 589-90; Thomas, *Collusion and Collective Action*, *supra* note 21, at 318-19.

88. See, e.g., Kevin M. Baird, *Business Method Patents: Chaos at the USPTO or Business as Usual?*, 2001 U. ILL. J.L. TECH. & POL'Y 347, 347-48, 354-56, 359-64 (2001) (repeating arguments about prior art inadequacies and offering several suggestions for improvement, including mandatory prior art searching by applicants); Sandra Szczerbicki, Comment, *The Shakedown on State Street*, 79 OR. L. REV. 253, 273-82 (2000) (stating, incorrectly, that many commentators believe that interest in business method patents will be short-lived, criticizing business method patents, and observing that prior art inadequacies are likely); Brett Dorny, *Intellectual Piracy*, CIO MAG., Feb. 15, 2001, available at http://www.cio.com/archive/021501/fine_content.html (repeating prior art criticism); Kris Frieswick, *Are Business Method Patents a License to Steal?*, CFO MAG., Sept. 1, 2001, available at <http://www.cfo.com/printarticle/0,5317,4803|A,00.html> (last visited Nov. 25, 2003) (repeating prior art criticism); David Ignatius, *Firestorm in Cyberspace*, WASH. POST, Mar. 19, 2000, at B07 (reporting that Internet patent “rebels” at PC Forum conference had agreed to create a prior art database to prevent future issuance of business method patents that inadequately refer to prior art); Michael Mahoney, *Patents: The Weapon of Choice for E-Biz Giants*, E-COMMERCE TIMES, Apr. 6, 2001, available at <http://www.ecommercetimes.com/perl/story/8757.html> (repeating prior art criticism) (last visited Nov. 25, 2003); Seth Shulman, *Software Patents Tangle the Web*, TECH. REV., Mar./Apr. 2000, at 68 (observing prior art inadequacies in both software and business method patents); see also Raymond Van Dyke, *E War—Episode One: The Patent Menace*, 6 COMPUTER L. REV. & TECH. J. 1, 6-7 (2001) (noting inadequacy of prior art, but opining that time should rectify the problem); Gleick, *supra* note 58, § 6, at 44 (noting serious prior art problems in the examination of business method patent applications, placing these patents side-by-side with patents on such inventions as a method for exercising a cat and a method for determining bra size by measuring the breasts).

We make this statement only as an observation and do not intend it to be judgmental. Of course, the relatively few individuals expressing sanguinity about business method patents have also relied on assumptions and impressionistic judgments rather

Bezos, CEO of Amazon.com and co-inventor of the one-click patent,⁸⁹ and Tim O'Reilly, a well-known Internet publisher, went so far as to found the now-defunct BountyQuest.com in 2001. BountyQuest.com allowed patent challengers to post patents for which someone was seeking invalidating prior art. By providing relevant prior art, BountyQuest.com users could receive a bounty.⁹⁰ Although the venture subsequently posted patents in several technology fields, the publicity surrounding business method patents clearly provided the impetus for its creation.⁹¹

As with software patents, prior art-related criticisms of business method patents were almost exclusively based on impressionistic judgment without the benefit of empirical analysis. There have been references in the literature to numbers of software and business method patent applications and grants, but no systematic data collection and analysis.⁹²

As of this writing, Josh Lerner is the only one to our knowledge who has done in-depth research on business method patents. He found 445 financial patents in five subclasses of classification 705 issued from the

than data. See Greg S. Fine, Note, *To Issue or Not to Issue: Analysis of the Business Method Patent Controversy on the Internet*, 42 B.C. L. REV. 1195, 1206-09 (2001) (relating the arguments of the few supporters of these patents who have spoken up).

89. Yes, the irony is palpable. Bezos applied for another Internet business method patent on March 1, 2001, the application for which was published on February 14, 2002. See U.S. Patent App. No. 20020019856, Ser. No. 797503 (published Feb. 14, 2002) ("Method and System for Information Exchange Between Users of Different Web Pages"). Earlier, however, Bezos had posted an open letter on Amazon.com stating that, although it was his duty to protect his company's intellectual property to the fullest extent possible, he believed that Internet business method patents should be singled out for different treatment than other patents, including a shorter term of protection. See Posting of Jeff Bezos, *An Open Letter from Jeff Bezos on the Subject of Patents*, originally posted to <http://www.amazon.com> on Mar. 9, 2000, available at http://www.oreilly.com/news/amazon_patents.html (last visited Nov. 28, 2003); see also *Bezos Calls for Patent Reform*, ZDNET U.K. NEWS, Mar. 10, 2000 (providing further discussion regarding BountyQuest.com), available at <http://news.zdnet.co.uk/story/0,,t269-s2077612,00.html> (last visited Nov. 28, 2003).

90. See *Every Patent is a Double-Edged Sword*, BUS. WK. ONLINE, Mar. 5, 2001 (report compiled from e-mail correspondence between Business Week correspondent and Charles Cella), available at http://www.businessweek.com/bwdaily/dnflash/mar2001/nf2001035_724.htm (last visited Nov. 25, 2003); *Q&A: Junk-Patent Perps, Beware of BountyQuest*, BUS. WK., Apr. 30, 2001, at 30, available at http://www.businessweek.com/magazine/content/01_18/b3730038.htm.

91. See *BountyQuest: Our Vision and Our History*, at <http://www.bountyquest.com/bqcorp/bqcorp.htm>. (The BountyQuest website is no longer available.)

92. As with software patents, Greg Aharonian has been the main exception. His effort has been laudable, but as previously noted, his data are not such as to allow for reliable statistical testing.

early 1970s through February 2000, and analyzed the prior art in a sample of 100.⁹³ Among other findings, he determined that (a) patents had increasingly issued in fields of finance in which there is highly relevant academic finance literature, particularly in asset allocation, risk management, and valuation;⁹⁴ (b) some of the work published by finance academics was suitable subject matter for patenting;⁹⁵ (c) financial patents are characterized by a lower ratio of examiner-added prior art to applicant-submitted prior art than is the case in other areas of patenting involving academic research;⁹⁶ and (d) an apparent lack of sufficient examiner familiarity with academic research in finance has probably contributed to an overall inadequacy of citation to relevant academic research in financial patents.⁹⁷

Although Professor Lerner's research is valuable and interesting, it does not address our concern as to the largest and most controversial segment of business method patents, those targeting Internet-related uses. His comparison of finance patents with patents in other technologies in which academic research is important, such as biotechnology, was quite appropriate because finance is the one clearly identifiable area in business that appears distinctly in issued patents *and* has been the subject of highly focused academic research for a substantial period of time. Patents on financial tools and techniques are also a special case among business method patents in that one is not likely to know how to find and understand the relevant research without a Ph.D. in finance or economics. A person with a doctorate in a field such as math or physics could

93. Josh Lerner, *Where Does State Street Lead? A First Look at Finance Patents, 1971-2000*, 57 J. FIN. 901, 904-05 (2002) [hereinafter Lerner, *Where Does State Street Lead?*].

94. *Id.* at 913-14.

95. *See id.* at 918.

96. *Id.* at 927-28. To determine which references were cited by the applicant and which were added by the examiner, one must go through the "file wrapper," or prosecution history of each patent. This is, as Lerner admits, a very slow and laborious task, which is the reason why he limited his study to 100 of the 445 financial patents. *Id.* at 905. Such an undertaking was simply not feasible for our comparison of 1,093 Internet business method patents with 1,000 patents-in-general, and we primarily relied on circumstantial evidence to draw an inference that most prior art references are submitted by the applicant rather than being added by the examiner. However, in 2001, the PTO began identifying on the face of the patent which prior art references had been added by the examiner, and we are now able to cite statistics firmly supporting the non-empirical evidence. *See infra* at Part V.A.

97. Lerner, *Where Does State Street Lead?*, *supra* note 93, at 927-28.

understand the math if she were taught how to find the relevant literature.⁹⁸

III. RESPONSES BY CONGRESS AND THE PTO

A. Congressional Responses

1. *The First Inventor Defense Act*

Faced with mounting complaints characterized by a “how-could-the-PTO-do-such-a-thing?” tone, Congress and the PTO answered with measures targeting business method patents for special treatment. Congress passed the First Inventor Defense Act in 1999 to create a new defense to an accusation of patent infringement for a prior inventor of a business method that is later patented by another.⁹⁹

Unlike patent systems in many other countries, there is no “prior user right” (“PUR”) in U.S. patent law that protects a first inventor who has kept her invention secret (“prior secret user”) from being sued for infringement by a second inventor who patents the same invention. If the first inventor makes a *nonsecret* use of the invention (for example, a public use, demonstration or disclosure), or if she sells the invention (whether secretly or not), she creates prior art that should prevent a second inventor from obtaining a patent in the first place or invalidate the patent if issued and later challenged by an accused infringer. On the other hand, a first inventor keeping the invention secret does not create prior art that prevents the second inventor from patenting, and remains vulnerable to an infringement suit by the patenting second inventor.¹⁰⁰

98. Professor Lerner at Harvard Business School and the colleagues in finance whom we contacted at the University of Texas expressed their belief that a person having an MBA degree with a finance concentration could not adequately understand the academic finance literature.

99. First Inventor Defense Act of 1999, Pub. L. No. 106-113, 113 Stat. 1536 (codified as amended at 35 U.S.C. § 273 (2000)).

100. Despite the fact that state trade secret law and federal patent law may constitutionally coexist, the knowledge disclosure and dissemination policies underlying the latter disfavor the former. *Kewanee Oil Co. v. Bicron Corp.*, 416 U.S. 470, 491 (1974). One example is illustrative. Suppose that X invents a novel and nonobvious machine or process, and keeps it secret. X then uses the invention to make a product and markets the product commercially. Assuming that the product does not reveal the secret machine or process (which it is very unlikely to do), X’s invention does not lose its status as a protectable trade secret. If X decides to patent her invention, however, she must do so within one year after first marketing the product commercially, despite the fact that X never actually disclosed the invention itself. The patent law policy discouraging secrecy and favoring disclosure leads to the result that X’s sale of a product made by her secret inven-

In contrast, patent laws in many other nations have long provided for PURs that provide an infringement defense for a first secret user—the first inventor who creates no prior art and does not patent—in the event she is later sued by a second inventor who obtained a patent.¹⁰¹ PUR creates a personal defense that merely allows the prior inventor to continue practicing the invention within her business to the extent she had been doing so before the patent issued. Even though the U.S. patent regime has never included PUR, in 1999 Congress created such a defensive right for business methods out of the fear that a substantial number of business method patents might have reflected techniques previously developed and practiced in secrecy.¹⁰² However, Congress made no attempt to define the term “business method.” Nor did Congress make any findings that business methods are practiced in secrecy to any greater extent than processes or techniques in other fields.¹⁰³ To date, there has been no reported case in which the First Inventor Defense Act was at issue, and we thus have not yet had a chance to sympathize with the plight of a court seeking to ascertain what a business method actually is.¹⁰⁴

tion is a “public use” of that invention. *See Metallizing Eng'g, Co. v. Kenyon Bearing & Auto Parts, Co.*, 153 F.2d 516, 518 (1946).

101. All nations except the United States employ a first-to-file system for determining which inventor has priority for patenting purposes. Thus, in other countries the presence of PURs can add some of the assumed fairness of a first-to-invent system to the greater efficiency of a first-to-file system. This does not mean, however, that PURs will not work under the U.S. first-to-file rule for determining priority between contemporaneous inventors seeking patents on the same invention. Indeed, if PURs are a good idea in general, they should work equally well in the United States for all types of inventions.

102. For more detail on the First Inventor Defense Act and an argument that congressional failure to define business methods may open the door to a more generally applicable PUR, see Jeffrey P. Duke, Comment, *The First Inventors Defense Act (35 U.S.C. § 273) Have Prior User Rights in Patent Law Been Resurrected?*, 20 ST. LOUIS U. PUB. L. REV. 223, 242 (2001). *See also* Thomas, *Collusion and Collective Action*, *supra* note 21, at 337-40 (discussing the First Inventor Defense Act and arguing that it increases the probability of collusion between the first secret inventor and the subsequent patentee to suppress invalidating prior art). *See generally* James R. Barney, Comment, *The Prior User Defense: A Reprieve for Trade Secret Owners or a Disaster for the Patent Law?*, 82 J. PAT. & TRADEMARK OFF. SOC'Y 261 (2000) (criticizing the First Inventor Defense Act of 1999).

103. In all areas of technology, trade secrets are preferred over patents by research and development executives as a means of generating returns from R&D investments. *See generally* COHEN ET AL., WHY U.S. MANUFACTURING FIRMS PATENT, *supra* note 51; Levin et al., *supra* note 51. Thus, if creating PUR for business method patents is a good idea, it should be an equally good idea for patents in all areas.

104. In common parlance, a business method is a method, technique, or process that has something to do with the practice of business. We wonder whether, for example, a chemical refining process would be a business method for a firm in the business of refin-

Although it is not entirely unheard of for Congress to single out a particular field for different protective scope under patent law, it is quite rare. In 1996, Congress granted immunity to medical practitioners and health-care entities so that they are no longer liable for infringing medical and surgical procedure patents.¹⁰⁵ That action by Congress has been criticized, though not widely, because of the difficulties likely to be encountered in defining “medical or surgical procedures.”¹⁰⁶ Even if one believes that it was socially optimal to remove the threat of an infringement action so as to give physicians the freedom to use any procedure they choose, it is unlikely that a reasonable estimate of costs and benefits would lead to the same conclusion for business methods.

ing petroleum products. Although probably not within the contemplation of most people who have given thought to business methods, why would the refining process not be included within such a definition? Other examples are innumerable.

105. Omnibus Consolidated Appropriations Act, 1997, Limitation On Patent Infringements Relating To A Medical Practitioner's Performance Of A Medical Activity, Pub. L. No. 104-208, 110 Stat. 3009, § 616 (codified as amended at 35 U.S.C. § 287(c) (2000)). The act immunizes “medical activity” performed by a “medical practitioner,” *id.* § 287(c)(1), defines “medical activity” as a “medical or surgical activity performed on a body,” *id.* § 287(c)(2)(A), and does not include as an immunized activity “(i) the use of a patented machine, manufacture, or composition of matter in violation of such patent, (ii) the practice of a patented use of a composition of matter in violation of such patent, or (iii) the practice of a process in violation of a biotechnology patent,” *id.* § 287(c)(A)(i)-(iii). “Medical practitioner” is unhelpfully defined as “a natural person who is licensed by a state to perform the medical activity.” *Id.* § 287(c)(2)(B). Related health care entities would include hospitals, nursing homes, and similar places where patients are treated and with which the medical practitioner has a professional affiliation. *Id.* § 287(c)(2)(C). Several bills sought the outright exclusion of such procedures from the definition of patentable subject matter, and the ultimately adopted legislation immunizing physicians from liability represented a compromise. See Brett G. Alten, Note, *Left To One's Devices: Congress Limits Patents on Medical Procedures*, 8 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 837, 859-97 (1998).

106. See Alten, *supra* note 105, at 359-97.; see also Steve Dirksen, *A Reconsideration of the Physicians' Immunity Statute*, 2001 DUKE L. & TECH. REV. 27 (2001) (arguing that, to adequately recompense physician research and create incentives for medical procedure innovation, a licensing scheme should replace the physician exemption); Cynthia M. Ho, *Patents, Patients, and Public Policy: An Incomplete Intersection at 35 U.S.C. § 287(c)*, 33 U.C. DAVIS L. REV. 601 (2000) (generally critiquing the exemption); Scott D. Anderson, Comment, *A Right Without a Remedy: The Unenforceable Medical Procedure Patent*, 3 MARQ. INTELL. PROP. L. REV. 117, 117 (1999) (referring to Congress's action as a “radical amendment” that was “buried within an appropriations measure” and “passed as an impetuous response to a single lawsuit”); Joel J. Garris, Comment, *The Case for Patenting Medical Procedures*, 22 AM. J.L. & MED. 85, 104 n.186 (1996) (critiquing congressional proposal to completely exclude medical and surgical procedures from patentability, and observing the problematic nature of defining a term like “surgical procedure”).

Business practices lack the social imperative of medical treatments.¹⁰⁷ Moreover, treating different technologies differently places too great a premium on *ex ante* definitions, such that the definitional scheme will be at least partially defeated because of the significant transaction costs associated with attorney efforts to opt into or out of a definition by carefully tailoring invention descriptions and patent claims.¹⁰⁸

2. *H.R. 5364*

On October 3, 2000, Representatives Howard Berman and Rick Boucher introduced H.R. 5364, “The Business Method Patent Improvement Act of 2000.”¹⁰⁹ Berman and Boucher admittedly did not expect the measure to pass so late in the session, but they intended to promote discussion of the issue. Congressman Boucher observed, in what we now consider to be classic political hyperbole, that “few issues in the 107th Congress will be more important than deciding whether, and under what conditions, the government should be issuing ‘business method’ patents.”¹¹⁰

H.R. 5364 proposed to increase the difficulty of moving business methods from application to issuance, and to make them easier to challenge once issued. The 2000 Berman-Boucher bill proposed means for restraining the issuance of business method patents, which included (1) requiring publication of business method patent applications eighteen months after filing, regardless of whether the applicant also contemplated

107. There have been a few other instances in which Congress has treated particular technology areas differently. These are briefly discussed *supra* at note 19.

108. Although this will not always be successful, it can be done, and undoubtedly will be done if a significant premium is placed on whether an invention is defined as a business method. *See, e.g.,* Jeffrey R. Kuester & Lawrence E. Thompson, *Risks Associated with Restricting Business Method and E-Commerce Patents*, 17 GA. ST. U. L. REV. 657, 678-79 (2001) (practitioners are adept at drafting patent claims so as to avoid penalties based on categorization). Further, the concept of a “business method” is likely to prove exceptionally difficult to define. In its 1996 revised guidelines to assist examiners with software inventions, the PTO observed that its personnel had experienced “difficulty in properly treating claims directed to methods of doing business,” and further stated that patent “[c]laims should not be categorized as methods of doing business,” but “should be treated like any other process claims.” Examination Guidelines for Computer-Related Inventions, 61 Fed. Reg. 7478-79 (USPTO Feb. 28, 1996).

109. H.R. 5364, 106th Cong. (2000).

110. 146 CONG. REC. E1,651-52 (daily ed. Oct. 3, 2000) (statement of Rep. Boucher). Congressman Boucher also stated that “[s]omething is fundamentally wrong with a system that allows individuals to get patents for doing the seemingly obvious We’re introducing this legislation in an effort to repair the system before the PTO awards more monopoly power to people doing the patently obvious.” *Id.* at 1651.

filing in another country;¹¹¹ (2) allowing the public to submit prior art after application publication and protest the issuance of the patent;¹¹² (3) instituting a post-grant administrative opposition procedure for challenging the patent's validity that closely resembles opposition procedures employed in countries that do not allow courts to invalidate issued patents;¹¹³ (4) requiring applicants for these patents to disclose whether and to what extent they have conducted a search of the prior art;¹¹⁴ (5) presuming that the invention is obvious if the business method is previously known and the only new feature is computer implementation;¹¹⁵ and (6) imposing a preponderance of the evidence burden of proof for the challenger to a business method patent's validity rather than the clear and convincing evidence standard employed for other patent validity challenges.¹¹⁶ In the first attempt to legally define a business method, H.R. 5364 described it as: "(1) a method of—(A) administering, managing, or otherwise operating an enterprise or organization, including a technique used in doing or conducting business;

111. H.R. 5364 § 3. As part of the American Inventors Protection Act of 1999 (of which the First Inventors Defense Act was also a part), Congress amended the Patent Act to provide for publication of patent applications eighteen months after filing, unless the application declared that it would not also file in a foreign country. 35 U.S.C. § 122(b) (2000). Previously, the PTO maintained U.S. patent applications in confidence until issuance of the patent. In other nations, however, patent applications traditionally were published at the eighteen-month point after filing, although usually not in English. Allowing an exception to eighteen-month publication in the United States for applicants not filing outside the United States was seen primarily as a bow to individual inventors and small businesses who are less likely to file overseas and thus less likely to have faced eighteen-month publication in another nation. The PTO began implementation of the new publication measure in March 2001.

Requiring eighteen-month publication of *all* applications for business method patents possibly could have had noticeable effects because applications for such patents rarely would be subject to the more general publication requirement since such applications are unlikely to be filed in other countries, given the fact that most other countries do not recognize patents on business methods. If publication of all applications is a good idea, however, there is no good reason to limit the practice to a single ill-defined technology area.

112. H.R. 5364 § 3.

113. *Id.*

114. *Id.* at § 5. There is no requirement in patent law that an applicant conduct a prior art search. Instead, there is only a requirement that the applicant disclose to the PTO any relevant material prior art of which the applicant is aware. 37 C.F.R. § 1.56 (2002); *see also* Upjohn Co. v. Mova Pharm. Corp., 225 F.3d 1306, 1315 (Fed. Cir. 2000). The combination of no required prior art search and a severe penalty if one is conducted but results not disclosed can have perverse effects. *See infra* text at Part V.A.1.

115. H.R. 5364 § 4.

116. *Id.* at § 3.

or (B) processing financial data; (2) any technique used in athletics, instruction, or personal skills; and (3) any computer-assisted implementation of a method described in paragraph (1) or a technique described in paragraph (2).”¹¹⁷ This definition has been criticized on several grounds, including the likelihood that it also could apply to patents on computer hardware,¹¹⁸ and more importantly, that its fluidity could logically encompass process inventions in many fields. Greg Aharonian observes, for instance, that if “part of the definition is ‘a technique used in doing or conducting business’ . . . isn’t using toluene instead of benzene in a chemical reaction a technique used in doing or conducting chemical engineering business?”¹¹⁹ He also provides an example of how the Amazon.com one-click patent can be redrafted so that it would be placed into one of the PTO’s software engineering subclasses and escape special treatment as a business method patent.¹²⁰

3. *H.R. 1332*

After Congress had expectedly taken no action on H.R. 5364, Congressmen Berman and Boucher introduced H.R. 1332, “The Business Method Patent Improvement Act of 2001” on April 3, 2001. The bill, which also received no action by Congress, defined a business method similarly to the previous year’s bill. The definition in the second bill, however, emphasized the inclusion of only software-implemented business methods.¹²¹ Although almost everyone thinks only of software-embodied business methods whenever the subject comes up, the previous definition could have been interpreted more expansively. Two observers

117. *Id.* at § 2.

118. Kuester & Thompson, *supra* note 108, at 678.

119. Gregory Aharonian, *Business Method Law Proposed*, INTERNET PAT. NEWSLETTER, Oct. 6, 2000 (on file with author). Aharonian’s publicly available archive of newsletters has an end point shortly before the date of this one, but will be updated in the future. See www.bustpatents.com (last visited Nov. 25, 2003) (including link to list of articles in archive, the text of which are available on a CD-ROM). In his comments about H.R. 5364, Aharonian uses his characteristically subtle style to state: “In short, IT IS UTTERLY IDIOTIC TO TRY TO DEFINE ‘BUSINESS METHODS’ AS A CLASS OF PATENT APPLICATIONS TO BE TREATED DIFFERENTLY.” *Id.* (capitalization original).

120. *Id.*

121. H.R. 1332 defined a business method as: “(1) a method—(A) of—(i) processing data; or (ii) performing calculation operations; and (B) which is uniquely designed for or utilized in the practice, administration, or management of an enterprise; (2) any technique used in athletics, instruction, or personal skills; and (3) any computer-assisted implementation of a method described in paragraph (1) or a technique described in paragraph (2).” H.R. 1332, 107th Cong. § 2 (2001).

criticized H.R. 1332's definition as "similarly inexact."¹²² Hearings on H.R. 1332, although naturally displaying some of the unease with business method patents that prompted the bill's introduction in the first place,¹²³ also revealed concerns that the question whether an invention can be characterized as a business method would impose unnecessary burdens on the federal judiciary.¹²⁴ Additionally, the hearings revealed observations that an initial onslaught of questionable biotechnology patents in the 1980s had ameliorated with time and experience;¹²⁵ that the application of traditional novelty and nonobviousness concepts would increase the quality of business method patents over time;¹²⁶ and that more fundamental problems such as improving general PTO funding and prior art databases needed to be addressed.¹²⁷

No business method patent bill was introduced in the 2002 session, perhaps because previous sponsors concluded that most members of Congress believed the issue to be a tempest in a teapot; because their attention was diverted by terrorism and a deepening economic recession; or because interested members like Berman and Boucher were able to tell those pressuring them to tighten the screws on business method patents that the PTO had taken up the cause.

B. PTO Responses

In March 2000, the PTO unveiled a new program making it more difficult to obtain a business method patent. The PTO's March 2000 initiative included a program to establish a formal Customer Partnership with the software, Internet, and electronic commerce industries similar to an earlier biotechnology industry outreach.¹²⁸ Part of the program consisted of a "Roundtable Forum" to discuss issues and possible solutions surrounding business method patents. The PTO envisioned that

122. Kuester & Thompson, *supra* note 108, at 679 n.125.

123. See *Hearings on H.R. 1332*, *supra* note 23, at 13, 19-20, 62 (statements of Rep. Howard L. Berman, Rep. John Conyers, and Andrew B. Steinberg, Executive Vice President, Travelocity.com, a web-based travel agency).

124. *Id.* at 51-53 (statement of Ronald E. Myrick, President, Intell. Prop. Owners Ass'n).

125. *Id.* at 51-53.

126. In the congressional hearings on H.R. 1332, Nicholas P. Godici, Acting Director of the USPTO at the time, testified that the PTO was flexible and could adapt, whereas changes in the law itself would prove much less flexible. *Id.* at 27 (statement of Nicholas P. Godici, Acting Under Sec'y of Com. for Intell. Prop. and Acting Dir. of the U.S. Pat. and Trademark Off.).

127. *Id.* at 34-50 (statement of Michael K. Kirk, Executive Director of the Am. Intell. Property Law Ass'n (AIPLA)).

128. *Id.* at 48.

the Forum would also facilitate its effort to obtain industry feedback on the PTO's existing prior art databases and other information sources, with a view towards ultimately expanding its prior art collection.¹²⁹

Another effort was designed to make examination of patents in Class 705 more robust. The PTO created Class 705¹³⁰ ("Data Processing: Financial, Business Practice, Management, or Cost/Price Determination") in 1997 from the business and cost/price subclasses of Classes 395 and 364.¹³¹ The predecessors to Class 705 were originally dominated by computerized postage metering and cash register systems, but by 1994, financial transaction systems commanded the largest share of patents in the predecessor classes. Since the creation of Class 705, patents on electronic shopping, advertising management systems, records management, and many other software-embodied methods relating to business enterprises have gradually become the majority in this category.¹³²

As part of its stated goal of enhancing the quality of issued business method patents, the PTO greatly increased the number of examiners for applications placed in Class 705,¹³³ emphasized the hiring of examiners with both software expertise and business experience, and vowed to provide additional business method training.¹³⁴ The PTO identified online databases and periodicals that Class 705 examiners are required to

129. *Id.*

130. PTO classifications do not suggest different legal treatment for different technologies, but instead serve to aid the search for prior art.

131. *See* BUSINESS METHOD WHITE PAPER, *supra* note 24, at 6. Class 705 is defined as follows:

This is the generic class for apparatus and corresponding methods for performing data processing operations, in which there is a significant change in the data or for performing calculation operations wherein the apparatus or method is uniquely designed for or utilized in the practice, administration, or management of an enterprise, or in the processing of financial data. This class also provides for apparatus and corresponding methods for performing data processing or calculating operations in which a charge for goods or services is determined.

USPTO, *U.S. Patent Classification System—Classification Definitions as of June 30, 2000: Class 705*, available at <http://www.uspto.gov/web/offices/ac/ido/oeip/taf/def/705.htm> (last visited Nov. 30, 2003). Class 705 includes approximately one hundred subclasses.

132. *See* BUSINESS METHOD WHITE PAPER, *supra* note 24, at 7.

133. The number of examiners in Class 705 increased from seventeen in late 1997 to seventy-seven in April 2001. *See Hearings on H.R. 1332*, *supra* note 23, at 32 (statement of Nicholas P. Godici).

134. *See* BUSINESS METHOD WHITE PAPER, *supra* note 24, at 12.

search,¹³⁵ established “Electronic Information Centers” that provide examiners with access to more than 900 databases—many of which include business and financial information,¹³⁶ and instituted a second level of examiner review for applications in Class 705.¹³⁷ From March 2000 to March 2001, the grant rate for business method patent applications declined from 56% to 36%.¹³⁸ According to PTO Director James Rogan, this allowance rate had stabilized at 35-40% by October 2002.¹³⁹

IV. DATA ON INTERNET BUSINESS METHOD PATENTS AND GENERAL PATENTS

A. The General Patent Data Set

Our investigation employs two data sets. The first, developed by John Allison and Mark Lemley in connection with previous studies,¹⁴⁰ measures a large number of patent attributes for 1,000 randomly selected patents of

135. *Id.* at 15.

136. *Hearings on H.R. 1332, supra* note 23, at 31 (statement of Nicholas P. Godici).

137. *See* BUSINESS METHOD WHITE PAPER, *supra* note 24, at 21.

138. *Hearings on H.R. 1332, supra* note 23, at 52 (statement of Ronald E. Myrick).

139. Declan McCullagh, *Patent Office Swamped—“Hurting Technology”*, ZDNET NEWS, Oct. 16, 2002, available at <http://zdnet.com.com/2100-1106-962182.html> (last visited Nov. 25, 2003). Mr. Rogan, a former congressman appointed by President Bush in late 2001, stated, “We were granting 65 or 70 percent of these things. Now the rejection rate is around 65 or 70 percent.” *Id.*

Although Director Rogan was comparing apples to apples and there apparently has been a real decline in the allowance rate for business method patents, one generally must be cautious in relying on the PTO’s reported allowance rates because they do not properly take into account previous continuing applications in a chain leading to the patent in question. A significant number of patents relate back to one or more “parent” (or “ancestor”) applications, and many of these ancestor applications result in their own patents on very closely related inventions. *See* Cecil D. Quillen, Jr. & Ogden H. Webster, *Continuing Patent Applications and Performance of the U.S. Patent and Trademark Office*, 11 FED. CIR. B.J. 1, 3 (2001) (estimating the actual allowance rate as being far higher than the 66% reported by the PTO for 1993-98, perhaps as high as 80-97%) [hereinafter Quillen & Webster, *Continuing Patent Applications*]; Cecil D. Quillen, Jr. et al., *Continuing Patent Applications and Performance of the U.S. Patent and Trademark Office—Extended*, 12 FED. CIR. B.J. 35 (2002) (adjusting assumptions from previous study to correct a probable flaw, but still producing an estimated allowance rate much higher than reported, and higher than in Japan and Germany). The effects of ancestor applications on the time patent applications spend in the PTO prior to issuance are discussed *infra* at Part V.E.1.

140. Allison & Lemley, *Growing Complexity, supra* note 40 (statistically comparing variables in the 1996-98 data with counterpart variables in a set of one thousand randomly selected patents issued twenty years earlier, 1976-78); Allison & Lemley, *Who’s Patenting What?*, *supra* note 40 (statistically analyzing variables within the data set).

all kinds issued during a two-year period between mid-1996 and mid-1998.¹⁴¹ Some of the many patent attributes measured in that data set are not relevant to our current research on whether business method patents were of lower quality and value than others, and thus we report only the characteristics relevant to our study.¹⁴² For the current study we collected additional data for the set of general patents and made slight modifications to it for comparison with Internet business method patents. The parameters we compare are the number and type of prior art references, number of claims within the patents, number of inventors, and time spent in the PTO before issuance.¹⁴³

1. *Defining the Technology Areas*

In developing the general patent data set, we placed each of the 1,000 patents in one or more of fourteen technology areas. We did not use the PTO's classification system for our definitions because, as anyone who is familiar with that system knows, it identifies most classes and subclasses at a very low level of functional abstraction and is not especially helpful in conceptually defining technologies from a science or engineering perspective. PTO classes and subclasses also are frequently entwined in rather cryptic ways.¹⁴⁴ The system is designed to assist patent examiners in

141. A complete description of that data set is found elsewhere. Allison & Lemley, *Who's Patenting What?*, *supra* note 40, at 2108-13.

142. For example, we developed extensive data on the geographic origin of patents in both data sets, by state and by country, but these data are not germane to the current study's objectives.

143. All of these data were extracted from the patents themselves. Additional data on the entity status of patent owners at issuance—individual, small business, nonprofit, or large business—were obtained from the PTO, but we decided not to use it in this Article because it strays a bit from our main focus. These data did show, however, that the portion of patents owned by large businesses and the combined small-entity types was essentially the same in both data sets. The only difference was that, in the case of Internet business method patents, there was a significant ownership shift from individuals to small corporations.

144. For example, Class 345 is defined as "Computer Graphics Processing, Operator Interface Processing, and Selective Visual Display Systems." The class definition states:

This class provides for processes and apparatus for selective electrical control of two or more light-generating or light-controlling display elements in accordance with a received or stored image data signal. The image data includes character, graphical information or display attribute data. The image data may include, for example, information data from a peripheral input device, from the reception of a television signal, from the recognition of image data, or from the generation or creation of image data by a computer. This class also provides for digital data processing systems or methods for data processing for visual presentation, wherein the processing of data includes the creation

searching for prior art by invention function, not by technological concept. Nor did we find that the World Intellectual Property Organization's ("WIPO") system of International Patent Classifications ("IPCs") well suited for identifying technology areas at a conceptual level, because it also was designed for purposes quite different than ours, and similarly operates at too low a functionally oriented level of abstraction to serve our definitional objectives. We have traded the objectivity of a third-party classification scheme for the subjective but far more conceptually accurate taxonomy we have chosen.¹⁴⁵ Although almost all of the fourteen areas are susceptible to more than one definitional approach, we believe ours to be at least as reasonable as other possible alternatives.

While it is possible to devise an almost endless list of categories and subcategories, these fourteen categories were chosen because they reflect the areas of technology into which inventions in the sample generally fell after a trial run involving the careful study of several hundred patents. The size of the sample further suggests that this is a reasonable approach to categorizing technology areas in patents generally.

2. *The Fourteen Technology Areas*

Following, in alphabetical order, are the technology areas we identified and the definitions we developed.

- 1) Acoustics: A process or product that consists solely or in critical part of an invention using sound waves. Such an invention may also be included in another classification, such as medical device or computer-related.

or manipulation of graphic objects (*e.g.*, artificial images), text or use of an operator interface by a digital data processing system prior to use by or within a specific display system.

USPTO, *U.S. Patent Classification System—Classification Definitions: Class 345*, available at <http://www.uspto.gov/go/classification/uspc345/defs345.htm> (last visited Nov. 30, 2003).

Not only does the class combine different technology areas, such as computer hardware, software, and optics, but it also does so in a very function-oriented way rather than a concept-oriented way. To further illustrate, Subclass 74 within Class 345 is "Machine Element or Mechanism, particularly Subclass 471 for control elements which move in two planes," showing both the system's often bewildering structure and its focus on low-abstraction functionality. *Id.*

145. One must understand that the act of defining areas of technology in today's world is as much art as it is science, and that many inventions involve more than one technology, such as optics and electronics or acoustics and software. Some might reasonably disagree with these definitions.

- 2) Automotive-related: An invention that is intended for use with automobiles or trucks. An invention in this classification necessarily will also be included within another classification, such as mechanics, electronics, or computer-related.
- 3) Biotechnology: A process involving advanced genetic techniques intended to construct new microbial, plant, or animal strains, a product created from such a process, or the way such a process or product is used in biotechnology research.
- 4) Chemistry: A process that consists solely of chemical reactions, a product resulting from such a process, or an invention of which a chemical process or product is a critical part. Closely related inventions such as those on novel metal alloys are also included. An invention in the field of chemistry may be included in one or more other classifications, such as electronics or optics.
- 5) Communications-related: An invention intended to improve the state of the art in communications. An invention placed in this classification necessarily will also be included within another classification, such as optics, electronics, or computer-related.
- 6) Computer-related: (a) A process or product for improving computer hardware (except for advances in semiconductor technology, which are in a separate, mutually exclusive classification). (b) An invention solely embodied in software. (c) An invention in which a microprocessor or other integrated logic circuit is expressed in the patent as being a critical part of the invention (again excluding advances in semiconductor technology itself). Any invention in part (c), and some inventions in parts (a) and (b), of the computer-related classification will necessarily also be classified in one or more other categories.
- 7) Electronics: A process or product in which the sole or a critical part of the invention makes use of traditional electronic circuitry or involves electric energy storage. An invention in this classification may also be included in other classifications, including chemistry, mechanics, or optics.
- 8) Energy-related: Any invention that intends to advance the state of the art in the production, processing, or transmission of energy. An invention is also included in this classification if its intended use is research into some aspect of the production, processing, or transmission of energy. The definition of “energy” includes that

- produced by any means from any source, including fossil fuels, nuclear power, electricity, and the many forms of radiation. An invention in this classification necessarily will also be included within another classification, such as mechanics, electronics, acoustics (for example, seismological inventions for detection of oil and gas), optics, chemistry, or computer-related.
- 9) **Mechanics:** A process or product that consists solely of the use of mechanical parts, sometimes combined with heat, hydraulics, pneumatics, or other power sources; or an invention in which the above is a critical part. Some inventions classified as mechanical also will be in one or more other classifications, such as electronics. While many different types of inventions fit into this category, it is not a catchall “other” category.
 - 10) **Medical devices:** An apparatus to be used for the diagnosis or treatment of diseases in humans or animals including devices used in medical research. An invention classified as a medical device will also fall within at least one other classification, such as computer-related, electronics, mechanics, or optics.
 - 11) **Optics:** A process or product intended to advance the state of the art in the use of light waves or imaging. This may be its sole purpose or it may be a critical part of an invention also having other purposes. Optics technology often will also be classified in one or more other areas, such as medical devices, semiconductors, electronics, or chemistry.
 - 12) **Pharmaceutical:** A process or substance to be used in the diagnosis or treatment of diseases or other abnormal conditions in humans or animals, including processes or substances used in medical research. In this data set, a technology classified as pharmaceutical will also be within either the chemistry or biotechnology areas.
 - 13) **Semiconductors:** A process or product intended to advance the state of the art in researching, designing, or fabricating semiconductor computer chips. Sometimes a semiconductor patent will also be placed in another classification, such as a semiconductor manufacturing process employing a purportedly new chemical process.
 - 14) **Software:** An invention that is completely embodied in software, even if the claims of the patent refer to a system or article of manufacture. A pure software invention is also placed in the

computer-related classification. The instructions embodied in software code can often be embodied in semiconductor chips in a device; this is done in the obvious instances of modern consumer electronic devices, automobiles, and other devices in which the instructions are very specific to a particular function of the device and the use of software for logic instructions simply is not practically feasible. Another researcher might include within the software classification those inventions in which the algorithms are embodied in chips, but we have chosen to include within our definition of software only those inventions that consist purely of software that is not embodied in hardware.

B. The Internet Business Method Patent Data Set

We set out to construct a database of all patents issued through the end of 1999 that were clearly targeted for use with the Internet.¹⁴⁶ Our goal then was to extract from those patents the ones that could be characterized as “methods of doing business.” Using Lexis-Nexis, we experimented with many search terms before settling on “Internet or World Wide Web.”¹⁴⁷ Although patents covering technological advancements ultimately contributing to the operability of the Internet undoubtedly issued before the term Internet became part of the American vernacular, and others having an impact on the Internet surely continued to issue after use of the term became widespread, we did not attempt to gather all patents that may have *affected* the Internet because the effort would have taken us into the realm of speculation.¹⁴⁸ Instead, we sought only patents

146. We only sought patents issued through the end of 1999 because we began the effort in 2000 and, once the data set was in place, it was quite unwieldy to add to it piecemeal.

147. Search terms in LEXIS, of course, are not case-sensitive. Although we found it very unusual for a patent to use the term “World Wide Web” without also using the term “Internet,” we also included the former because it might pick up a small number of Internet-related patents with little chance of including irrelevant ones.

148. After we had developed our operational search request and had done much of our data collection, a member of the review committee at NAS suggested that we add a search term such as “computer network.” Although we were far enough into systematic study of the patents and data coding that doing a reconfigured search request would have caused substantial delay, we saw the suggestion as a reasonable one and experimented with it by searching for patents using the request “computer network! and not (Internet or World Wide Web)”. The search did return a substantial number of patents, but after closely reading over 200 of them we concluded that even though a relatively small number of additional patents probably could be included in the data set, attempting to do so would force us into too much speculation about whether inventions were truly targeted at the Internet. Without seeing a substantial advantage to be gained from a reconfigured search, we decided that further delaying a long and arduous project was unjustified.

on software inventions that clearly were aimed at use with the Internet. We then categorized the patents that made the cut.

1. Constructing the Data Set

Our initial search for all patents issued from the beginning of 1990 to the end of 1999 that included the terms “Internet or World Wide Web” yielded over 9,000, most of which used these terms in an incidental fashion (such as a cite to a piece of prior art on the Internet).¹⁴⁹ Although a survey of these patents found a few initially promising candidates in older computer industry classifications such as 345, 365, 370, and 375, we found by far the heaviest concentration in the newer 700-series “data processing” or “electrical computer” classifications.¹⁵⁰ In particular, most contenders for inclusion appeared in Class 705, 707, and 709.¹⁵¹ Although

149. Most of the patents we found had been issued in 1998 and 1999, with a far smaller number in 1997 and only a few before 1997.

150. When it created the 700-series classes, the PTO did go back several years and add the newer class numbers to selected patents in the older classes. Patents often are placed in more than one PTO classification and it was not uncommon to find patents that had been placed in one or more of the newer classes as well as in an older class.

151. These class titles and definitions are:

Class 705: Financial, business, practice, management, or cost/price determination: This is the generic class for apparatus and corresponding methods for performing data processing operations, in which there is a significant change in the data or for performing calculation operations wherein the apparatus or method is uniquely designed for or utilized in the practice, administration, or management of an enterprise, or in the processing of financial data. This class also provides for apparatus and corresponding methods for performing data processing or calculating operations in which a charge for goods or services is determined.

This class additionally provides for subject matter described in the two paragraphs above in combination with cryptographic apparatus or method.

USPTO, *U.S. Patent Classification System—Classification Definitions: Class 705*, available at <http://www.uspto.gov/go/classification/uspc705/defs705.htm> (last visited Nov. 30, 2003).

Class 707: Data Processing: database and file management, data structures, or document processing: This is the generic class for data processing apparatus and corresponding methods for the retrieval of data stored in a database or as computer files. This class provides for data processing means or steps for organizing and inter-relating data or files (e.g., relational, network, hierarchical, and entity-relationship models). This class also provides for data processing means or steps for generic data, file and directory upkeep, file naming, and file and database maintenance including integrity consideration, recovery, and versioning.

some possibly relevant patents were also found in other 700-series classes, a closer study of these showed that those potentially relevant to our study were also placed in 705, 707, or 709.¹⁵² Using our search terms, we retrieved over 2,800 patents in these three classes.¹⁵³

USPTO, *U.S. Patent Classification System—Classification Definitions: Class 707, available at* <http://www.uspto.gov/go/classification/uspc707/defs707.htm> (last visited Nov. 30, 2003).

Class 709: Electrical computers and digital processing systems: multiple computer or process coordinating: This class provides for an electrical computer or digital data processing system or corresponding data processing method including method or apparatus for transferring data or instruction information between a plurality of computers or processes wherein the computers or processes employ the data or instructions before or after transferring and the employing affects said transfer of data or instruction information. The class includes the following subject matter:

A. Process or apparatus for administrating process or job execution over a digital data processing system.

B. Process or apparatus for transferring data among a plurality of spatially distributed (i.e., situated at plural locations) computers or digital data processing systems via one or more communications media (e.g., computer networks).

C. Process or apparatus for exchanging data or messages between two executing programs or processes, generally independent of the hardware used in the communication.

D. Process or apparatus for synchronizing control or regulation of clocking or timing operations of two or more processors.

USPTO, *U.S. Patent Classification System—Classification Definitions: Class 709, available at* <http://www.uspto.gov/go/classification/uspc709/defs709.htm> (last visited Nov. 30, 2003).

152. For example, we found some patents that appeared not only in 705, 707, or 709, but also in classes such as: “Class 700: Data processing: generic control systems or specific applications,” USPTO, *U.S. Patent Classification System—Classification Definitions: Class 700, available at* <http://www.uspto.gov/go/classification/uspc700/defs700.htm> (last visited Nov. 30, 2003); “Class 703: Data processing: structural design, modeling, simulation, and emulation,” USPTO, *U.S. Patent Classification System—Classification Definitions: Class 703, available at* <http://www.uspto.gov/go/classification/uspc703/defs703.htm> (last visited Nov. 30, 2003); and “Class 713: Electrical computers and digital processing systems: support,” USPTO, *U.S. Patent Classification System—Classification Definitions: Class 713, available at* <http://www.uspto.gov/go/classification/uspc713/defs713.htm> (last visited Nov. 30, 2003).

153. The large difference between the nine thousand patents we first uncovered and the twenty-eight hundred we found after re-running the search only in Classes 705, 707, and 709 resulted from the far greater number of duplicate patents in the first exploratory search because many more classifications had been included and so many patents were in multiple classes.

In our multiple readings to ascertain whether a given patent should be characterized as Internet-related, we focused mainly on the written description of the invention (not just the abstract), although we also examined the claims.¹⁵⁴ We approached the study of these patents with the attitude that we would take the inventors at their word about whether the inventions would be used in connection with the Internet. In the case of employed inventors, of course, these words presumably reflected not only the inventors' own visions for the inventions but also those of corporate managers and patent attorneys familiar with their clients' patenting strategies. If the written description of the invention clearly demonstrated that it was intended to be used with the Internet, it was included; otherwise, it was discarded. We eventually excluded approximately one-half of these patents for a variety of reasons, yielding a data set of 1,423 patents.¹⁵⁵

2. *Categorizing Internet Patents*

Our study of the more than 2,800 patents also revealed that Internet-related patents fall into several discernible categories. Many described relatively broad models for doing business on the Internet and clearly were of the sort that comes to mind when most observers think about business method patents. The Priceline.com reverse auction model for purchasing airline tickets was prototypical of this type. Among the many others, there is U.S. Patent No. 6,006,265, "Hyperlinks resolution at and by a special network server in order to enable diverse sophisticated hyperlinking upon a digital network," a method for distributing hypervideo, i.e., digital video broadcasting incorporating hyperlinks.¹⁵⁶ Other examples include U.S. Patent No. 6,009,412, "Fully integrated on-line interactive frequency and award redemption program," describing a system for Internet shopping

154. Had our inquiry dealt with questions of validity or infringement, we obviously would have given most of our attention to the claims.

155. Three examples of reasons for exclusion from the data set follow. First, the search produced a number of duplicate patents because some had been placed into two or sometimes even all three of these classifications (705, 707, 709), although, as noted, the number of duplicates was far less than in our initial search. Second, and as in our first search, many patents in these classes that included the terms "Internet" or "World Wide Web" used the terms in only an incidental way and did not clearly envision use of the software invention with the Internet. For instance, there were substantial numbers of patents in which these terms were found only in a prior art reference to an item that had been found on the Internet, for example, a web page. Third, several patents described only software application products with some discussion of how they could be delivered over the Internet. Through this process we discarded approximately fifty percent, yielding a set of 1,423.

156. U.S. Patent No. 6,006,265 (issued Dec. 21, 1999).

incorporating an awards program,¹⁵⁷ and the self-explanatory U.S. Patent No. 5,978,780, “Integrated bill consolidation, payment aggregation, and settlement system.”¹⁵⁸ We labeled these as “Internet business model patents.”

We also encountered inventions narrower in scope, intended not as business models, but only as means for solving specific business problems. The Amazon.com one-click patent is illustrative. Other examples include U.S. Patent No. 6,006,332, “Rights management system for digital media,” which provides a method for controlling unauthorized access to copyrighted material distributed by content owners over the Internet,¹⁵⁹ and U.S. Patent No. 6,009,382, “Word storage table for natural language determination,” a method by which a company can distribute a document in multiple human languages and have a particular recipient’s computer automatically select the recipient’s language.¹⁶⁰ We labeled these patents as “Internet business technique patents.”¹⁶¹

Our study of Internet-related patents revealed a third type involving techniques that purport to make the Internet more efficient and effective for conducting electronic commerce by solving a technical software problem. These are easily distinguishable from broadly-described business methods, but are sometimes difficult to differentiate from narrowly drawn ones. One example of the third variety is U.S. Patent No. 6,003,077, “Computer network system and method using domain name system to locate MIB module specification and web browser for managing SNMP agents,”¹⁶² which was fairly easy to distinguish from a patent on a broad business model because of its very narrow, technical nature. Another, U.S.

157. U.S. Patent No. 6,009,412 (issued Dec 28, 1999).

158. U.S. Patent No. 5,978,780 (issued Nov. 2, 1999).

159. U.S. Patent No. 6,006,332 (issued Dec. 21, 1999).

160. U.S. Patent No. 6,009,382 (issued Dec. 28, 1999).

161. The PTO and most commentators have associated patents on business models and business techniques solely with Class 705. Although Internet business model and business technique patents are more highly concentrated in Class 705 (especially models), they can also be found in 707 and 709. As noted earlier, the PTO frequently places patents in more than one classification, the first one listed being the “main classification.” We point this out because the PTO’s March 2000 initiative, which includes a second-level review before issuance, applies only to patents with a main classification of 705. Therefore, a substantial number of business method patent applications are not subject to this program of greater scrutiny because they either are not in 705 at all or because they are in 705 as a secondary classification. Although the end point of our Internet business method data set, December 31, 1999, is prior to the PTO’s institution of this initiative, we make these observations to further substantiate our position that placing a premium on subject matter definition or classification creates its own set of problems.

162. U.S. Patent No. 6,003,077 (issued Dec. 14, 1999).

Patent No. 6,005,939, "Method and apparatus for storing an Internet user's identity and access rights to world wide web resources," presented us with a definitional issue about which reasonable people could differ. We ultimately placed it with others of a similar nature in a category we called "Internet software technique patents,"¹⁶³ because the invention description emphasized the solution to a software problem much more than it emphasized the solution to a business problem.

3. *Excluding the Last Category*

After finishing development of our taxonomy of three categories of Internet patents and placing all 1,423 into one of the three, we concluded that only the first two (business models and business techniques) should remain in our data set for comparison with the general patent data set. The third category more logically belongs in the larger universe of "software patents," of which software-implemented business method patents are a subset. Although the third category consists of Internet-related software patents, few knowledgeable observers would characterize them as business method patents. As a result, our final Internet business method data set includes 1,093 patents issued through the end of 1999. As one would expect, most of these patents issued in 1998 or 1999, with a much smaller number in 1997 and only a handful in 1996. Because the general patent data set includes patents of all types issued between mid-1996 and mid-1998, the two data sets are sufficiently contemporaneous for statistical comparison.¹⁶⁴

V. COMPARISON OF PATENT CHARACTERISTICS

A. Number of Prior Art References

Aside from objections that business methods should not even be patentable subject matter, the most common reason for castigating them was the perception that they are granted with little attention to the prior art. In the examination of patent applications, prior patents and other printed publications ("nonpatent references") describing related

163. U.S. Patent No. 6,005,939 (issued Dec. 21, 1999).

164. Contemporaneity is very important for data comparability in statistical comparison of patent groups because the values of almost all patent variables such as references, claims, and inventors have increased significantly over time. See Allison & Lemley, *Growing Complexity*, *supra* note 40, 97-103 (demonstrating this for the twenty years between the late 1970s and late 1990s).

technological advances are the two most important types of “prior art.”¹⁶⁵ There is a strong intuitive appeal to the idea that the number of prior art references in a patent may relate to patent quality, and thus to patent value. One reason for this appeal is the fact that uncited prior art—art that was not before the patent examiner—is the most common basis for court decisions invalidating U.S. patents.¹⁶⁶ Therefore, a larger number of references in a patent should, on average, increase the probability that it would be held valid if challenged. Naturally, we are speaking only of statistical tendencies over a large group of patents.

When patent applicants spend more time, effort, and money conducting more thorough prior art searches to better differentiate their inventions from what others have done, they probably do so mainly out of a conviction that their innovations have greater value. More vigorous differentiation from the prior art leads, in turn, to more valuable patents because they have a better chance of withstanding validity challenges. We thus witness a likely relationship between the perception and the reality of greater patent value. This notion of self-selected value, which we revisit in other contexts, seems perfectly acceptable because applicants should have more relevant information than anyone else about the probable private value of their inventions. Moreover, applicants’ inclusion of more prior art may result in more thorough examinations in the PTO, further contributing to stronger patents.¹⁶⁷ Relating the number of prior art references to average patent value is thus an intuitively attractive concept.¹⁶⁸

165. The other types of prior art consist of evidence that an invention had been in public use or had been placed on sale. 35 U.S.C. §§ 102(a), (b) (2000). Although evidence of a prior public use or of prior activity that placed an invention on sale often forms the grounds for a judicial invalidity finding, these types of prior art are far less likely to be a factor in the PTO examination process.

166. Allison & Lemley, *Validity of Litigated Patents*, *supra* note 31, at 231-34.

167. Prior art references are listed in the patent by both the applicant and the examiner, but it is not feasible to determine for a large number of patents which references were cited by the applicant and which by the examiner because one must study the prosecution history in the PTO to make this determination. However, there are reasons to believe that the great majority of prior art referenced in patents have been cited by applicants rather than by the PTO examiner. See Allison & Lemley, *Who’s Patenting What?*, *supra* note 40, at 2137. Allison and Lemley found that U.S. patents on foreign-origin inventions cite much more foreign-origin prior art and much less U.S.-origin prior art than do U.S. patents on U.S.-origin inventions. There is reason to believe that these foreign applicants for U.S. patents have better access to foreign-origin prior art in their language. However, if very much additional prior art were cited by the U.S. patent examiner, one would expect it to be English-language prior art. The fact that this did not appear to occur supports the inference that most of the prior art is cited by the applicant and not the examiner.

One study found empirical support for the notion that the number of prior art references is positively correlated to independent measures of patent value,¹⁶⁹ but two others were unable to find a statistically significant relationship.¹⁷⁰ Those studies counted only references to prior

Another observation clearly provides strong support for this conclusion. One finds wild variations in the number of patent and nonpatent prior art references among U.S. patents in the same area of technology. We certainly found this to be true in the case of Internet business method patents. Unless all or most of this prior art is cited by applicants, then patent examiners in the same technology area do not have access to the same resources, they do not communicate with each other, there is little supervision by primary examiners, or all of the above. There is absolutely no reason to believe any of these possibilities and, therefore, most prior art is probably cited by applicants.

Moreover, the fact that the distribution of prior art references in a group of patents tends to be bimodal, as discussed *infra* at Part V.A.1, also provides strong support for an argument that most prior art is cited by applicants rather than examiners. The reasons for such a distribution relate only to applicants, not examiners.

Our theory that most prior art is cited by applicants rather than examiners was recently confirmed empirically. Beginning on January 1, 2001, the PTO began placing an asterisk beside each prior art reference that was cited by the examiner. Bhaven Sampat found that, between January 1, 2001 and June 25, 2002, 40% of the citations to prior U.S. patents, 13% of foreign prior art (foreign patents, foreign search reports and published foreign patent applications), and 3% of nonpatent prior art was cited by examiners. Bhaven N. Sampat, *Analysis of Examiner Inserted Citations, Patent Information User Group (PIUG) Discussion List Archives*, Mar. 7, 2003, available at <http://www.questel.orbit.com/piug/current/0173.html> (last visited Nov. 11, 2003).

168. As with other variables we measure, the relationship may or may not hold in the case of any particular patent. We speak of statistical relationships in large groups of patents. To ascertain whether references in a given patent strengthen it, one must be knowledgeable in the technical area and study the difference between the patent's claims and the prior art.

169. Dietmar Harhoff et al., *Citations, Family Size, Opposition and the Value of Patent Rights* (Working Paper 1999), available at <http://emlab.berkeley.edu/users/bhhall/harhoffetal99.pdf> (last visited Nov. 11, 2003).

170. LANJOUW & SCHANKERMAN, *STYLIZED FACTS OF PATENT LITIGATION*, *supra* note 38, 15; Lanjouw & Schankerman, *Characteristics of Patent Litigation*, *supra* note 34, 138.

There is stronger empirical evidence that the number of forward citations is a predictor of patent value. The term "forward citations" has been used by some economists to describe what most people would call "citations received," that is, the number of citations received by a patent in later issued patents. The proposition that patents receiving much more attention as prior art in later patents have more value has both instinctive attraction and strong empirical support. See generally BRONWYN HALL ET AL., *MARKET VALUE AND PATENT CITATIONS: A FIRST LOOK*, (Nat'l Bureau of Econ. Research, Working Paper No. 7741, 2000), available at <http://www.nber.org/papers/w7741> (last visited Nov. 11, 2003); Dietmar Harhoff et al., *Citation Frequency and the Value of Patented Inventions*, 81 *REV. OF ECON. & STAT.* 511 (1999); Trajtenberg, *supra* note 29. We did not measure forward citations because the patents in our data sets were so recent, those in one set issuing from mid-1996 to mid-1998 and those in the other set almost all issuing

patents, however, and not references to nonpatent prior art, a somewhat glaring omission given the importance of nonpatent prior art in many fields.¹⁷¹ Removing some of the ambiguity in previous research, a recent study in which one of the authors participated found that patents involved in patent infringement litigation have significantly more prior art references of all kinds than the general population of patents.¹⁷² Litigated patents, in turn, have been shown to be more valuable than patents in general.¹⁷³ In addition to the existence of empirical support for the existence of a relationship between the number of prior art references and patent value, as well as the intuitive appeal of an argument that such a relationship exists, we heavily emphasize prior art in the present study because prior art has served as a centerpiece of business method patent criticism.

1. *Internet Business Method Patents Compared With the Average Patent*

The following table shows the results of our comparison of patent, nonpatent, and total references in Internet business method patents and general patents, including the untransformed means.¹⁷⁴

during 1998-99. Our study design and data collection began early in 2000 and ended before the middle of 2001. Collection of data on forward citations should be done when sufficient time has passed for these data to be meaningful for the entire data set. The NBER database compiled by Bronwyn Hall and her colleagues, *supra*, reports citations received only through December 31, 1999 for patents issued between 1963 and 1999.

171. The probable reason that prior studies may have failed to take account of nonpatent prior art is that extracting these data is extremely tedious and time-consuming.

172. Allison et al., *Valuable Patents*, *supra* note 30 (manuscript at 25-26 & tbl.1). Indeed, using the term "significant" is an understatement. The study found that litigated patents had more prior art references by a margin that far exceeded the typical .05 level for statistical significance. *Id.*

173. See generally LANJOUW & SCHANKERMAN, ENFORCING INTELLECTUAL PROPERTY RIGHTS, *supra* note 33; Lanjouw & Lerner, *Literature Survey on I.P. Rights Enforcement*, *supra* note 33.

174. Although we used log-transformed means to smooth out extreme variations in the distributions for most statistical tests in this study, we report untransformed means in the summary tables for ease of understanding.

**Table 1: Prior Art References: Internet Business Method Patents
Compared with General Patents**

	N	Variable	Mean	Median	Std. Dev.	Min.	Max.
Internet Business Method Patents	1093	Patent Refs	14.90	10	23.76	0	314
		NonPat Refs	10	2	38.56	0	391
		Total Refs	24.90	13	53.15	0	457
General Patents	1000	Patent Refs	12.79	9	14.13	0	154
		NonPat Refs	2.37	0	6.56	0	68
		Total Refs	15.16	10	16.29	0	163

Statistical comparison of these means showed that Internet business method patents had significantly more patent references, nonpatent references, and total references than general patents.¹⁷⁵ The null hypothesis of “no difference” was rejected with a confidence level of over 99.99% in the case of total references and nonpatent references ($p < 0.0001$), and over 97% in the case of patent references ($p < 0.0235$).

175. Based on the descriptive analysis of the distributions of the three types of prior art references, log transformations were useful for normalizing the number of patent references and total references, but not for nonpatent references because it was not necessary. Although our standard for statistical significance is 95% or greater, i.e., a p -value of 0.05 or less, p -values were often far lower than 0.05, thus signifying very high levels of significance. Reported p -values are always two-tailed. These and almost all other comparisons in this study used independent-groups t -tests with means that had been adjusted by performing logarithmic transformations to adjust for extreme values in the data distributions. The data distributions for all of our variables, except the number of inventors, were substantially influenced by extreme upper outliers, i.e., very long right-hand tails in the distributions. A simple comparison of untransformed means would thus have not provided reliable results because those means were not good indicators of central tendencies. As an additional check on our results, we employed Wilcoxon nonparametric tests in almost all instances.

The distribution of prior art references in a large group of patents is bimodal, with many applicants citing either very few or a large number.¹⁷⁶ Bimodality is even more pronounced in a distribution of nonpatent references, where one often finds none at all, because it is more time-consuming and thus more expensive to search for nonpatent prior art than it is to search for relevant prior patents. Overall, this phenomenon is easily explainable by the fact that patent law does not require applicants to conduct a prior art search at all, but PTO Rule 56 does require applicants to cite all material prior art of which they are aware.¹⁷⁷ A violation of this duty of candor regarding known prior art can result in a judicial finding that the patentee committed “inequitable conduct,”¹⁷⁸ which may render an otherwise valid and infringed patent unenforceable. Indeed, such a finding can produce consequences substantially more severe than an invalidity finding. Both infringement and validity are determined on a claim-by-claim basis, leading to situations in which certain claims in a patent may be invalid while others remain valid. A patentee’s inequitable conduct, however, renders the entire patent unenforceable,¹⁷⁹ as well as entitling the accused infringer to successfully raise inequitable conduct as a defense to attorney fees.¹⁸⁰ Furthermore, related patents springing from the same original application may also become unenforceable.¹⁸¹ The absence of a duty to search the prior art and the cost of doing so, coupled with potentially severe penalties for not citing what has been found, tends to cause many applicants to take extreme approaches, conducting either minimal searches or painstakingly thorough ones.¹⁸²

176. As explained at *supra* note 167, most prior art references are almost certainly cited by applicants rather than examiners.

177. 37 C.F.R. § 1.56 (2002).

178. *See, e.g.*, *Kingsdown Med. Consultants, Ltd. v. Hollister Inc.*, 863 F.2d 867, 872 (Fed. Cir. 1988) (en banc) (discussing inequitable conduct defense).

179. *See, e.g.*, *J.P. Stevens & Co. v. Lex Tex, Ltd.*, 747 F.2d 1553, 1561 (Fed. Cir. 1984).

180. *See, e.g.*, *Gentry Gallery, Inc. v. Berkline Corp.*, 134 F.3d 1473, 1480 (Fed. Cir. 1998).

181. *Consol. Aluminum Corp. v. Foseco Int’l, Ltd.*, 910 F.2d 804, 809-10 (Fed. Cir. 1990) (finding six related patents unenforceable because of inequitable conduct in intentionally failing to adequately disclose the best mode for one of them). *See generally* Scott D. Anderson, Comment, *Inequitable Conduct: Persistent Problems and Recommended Solutions*, 82 MARQ. L. REV. 845 (1999) (discussing the various consequences of inequitable conduct).

182. We confirmed the existence of bimodality by examining the distributions of each outcome including references (patent, nonpatent, and total) using histograms and Quantile-Quantile (“QQ”) plots.

In addition to bimodality, the distribution of the numbers of prior art references is also characterized by extreme upper outliers.¹⁸³ Some patent applicants cite a large amount of prior art. These distribution characteristics existed in both data sets. In addition to performing logarithmic transformations of the means to adjust for the skewed distributions before comparing them,¹⁸⁴ we compared the median number of prior art references in the two data sets to further ensure that patents with extremely large numbers of references did not over-influence our finding that Internet business method patents cited significantly more prior art than patents in general. Nonparametric tests of medians confirmed our finding that Internet business method patents contain significantly more patent, nonpatent and total prior art references than patents overall.

One may gain further insight into the data by comparing the proportion of patents in the Internet and general patent data sets that cited no prior art at all. We were unable to conduct statistical comparisons between the number of patents in each data set containing zero references to prior patents, or between the number of patents in each data set containing zero total prior art references, because the number of such patents was too small for statistical testing.¹⁸⁵ Far more telling, however, was a comparison of the proportion of patents in each set having no references to nonpatent prior art, that is, to articles in periodicals and other types of documents. The data revealed that 62.1% (621 of 1,000) of general patents cited no nonpatent prior art, whereas only 32.2% (352 of 1,093) of the Internet business method patents cited none, a difference that is not just significant but truly striking. In sum, using various methods, we found that Internet business method patents cited substantially more prior art than did the average patent during the late 1990s.

183. Extreme values on the right side of the distribution characterized all of the other variables we studied, except for the number of inventors. The distributions for other variables were not characterized by bimodality, however.

184. The number of references was transformed using natural logs. Based on visual inspection of the log-transformed distribution, this appeared very effective in reducing the influence of the small subset of patents with exceptionally large numbers of references. The independent groups *t*-tests using the log-transformed values provides a valid test of the central tendencies (e.g., mean) of each distribution after minimizing the effect of upper outliers.

185. In the general patent data set, 14 out of 1,000 patents (1.4%) cited no patent prior art. In the Internet business method data set, 23 out of 1,093 patents (2.1%) cited no patent prior art. In the general patent data set, 2 out of 1,000 patents (0.2%) cited no prior art of any kind. In the Internet business method data set, 3 out of 1,093 (0.3%) cited no prior art of any kind.

2. *Internet Business Method Patents Compared With Patents in Individual Technology Areas*

We also compared the number of prior art references in Internet business method patents with the number in each of the fourteen areas of technology identified in the general patent data set. Such comparisons are important because patenting strategies and practices probably show some variations among different fields.¹⁸⁶ Furthermore, previous economics research exploring possible relationships between certain patent characteristics and patent value did so within technology areas or industries.¹⁸⁷

The results of comparing the log-transformed mean number of patent, nonpatent, and total prior art references in Internet business method patents with those in each of the fourteen technology areas follows in Table 2.¹⁸⁸ In the table, “more,” “fewer,” and “equal” refer to situations in

186. As means for protecting investments in research and development, the relative importance of patents, in contrast with other means such as secrecy and lead-time, varies among industries and technology fields. *See generally* COHEN ET AL., WHY U.S. MANUFACTURING FIRMS PATENT, *supra* note 51; Bronwyn H. Hall & Rosemarie Ham Ziedonis, *The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995*, 32 RAND J. ECON. 101 (2001) (exploring patenting strategies in the semiconductor industry); Levin et al., *supra* note 51; Edwin Mansfield, *Patents and Innovation: An Empirical Study*, 32 MGMT. SCI. 173 (1986); Ariel Pakes & Margaret Simpson, *Patent Renewal Data*, in 1989 BROOKINGS PAPERS ON ECON. ACTIVITY—MICROECONOMICS 331 (Martin Neil Baily & Clifford Winston, eds., 1989).

187. *See generally* LANJOUW & SCHANKERMAN, STYLIZED FACTS OF PATENT LITIGATION, *supra* note 38 (finding that litigated patents are more valuable than unlitigated ones and that the incidence of patent litigation varies significantly among technology fields); Jean O. Lanjouw, *Patent Protection in the Shadow of Infringement: Simulation Estimations of Patent Value*, 65 REV. ECON. STUD. 671 (1998) (correlating incidence of decisions to pay patent renewal fees in Germany with simulated litigation propensity within each of four industries); Lanjouw & Schankerman, *Characteristics of Patent Litigation*, *supra* note 34 (comparing total number of claims in litigated and unlitigated patents within same 4-digit IPC); Lerner, *The Importance of Patent Scope*, *supra* note 30 (correlating the number of IPCs into which patents were placed with value of patent-owning start-up firms within biotechnology industry); Schankerman & Pakes, *supra* note 29 (identifying incidence of patent renewal decisions in the United Kingdom, France, and Germany within particular industries as an indicator of patent value); Trajtenberg, *supra* note 29 (finding citations received by a patent in later patents to be an indicator of patent value within the field of computer tomography scanners).

188. There are two major differences between comparing Internet business method patents with patents in general and comparing them with patents from individual technology areas. One difference is that the number of patents in each area is necessarily smaller than the whole, meaning that any variance must be greater to demonstrate statistical significance. The other difference is that the distribution of values in a particular technology area sometimes may be even more skewed than the distribution in general patents as a

which Internet business method patents have significantly more, significantly fewer, or an insignificantly different number of references than patents in the particular technology area, respectively. To illustrate, Internet business method patents had significantly more patent references than acoustics patents, significantly fewer patent references than medical device patents, and an insignificantly different number of patent references than automotive-related patents.

Table 2: Prior Art References:

Internet Business Method Patents Compared with Patents in Each Technology Area

	Pat Refs	NonPat Refs	Total Refs
Internet Business Methods Compared With			
Acoustics	More	More	More
Automotive-Related	Equal	More	More
Biotechnology	More	Fewer	Equal
Chemistry	Equal	More	More
Communications-Related	Equal	More	More
Computer-Related	More	More	More
Electronics	More	More	More
Energy-Related	Equal	More	Equal
Mechanics	Equal	More	More
Medical Devices	Fewer	More	Equal
Optics	More	More	More
Pharmaceutical	More	Equal	More
Semiconductors	More	More	More
Software	More	More	More

whole. Thus, the use of log-transformed means for comparison and the use of nonparametric tests is likely to be even more important because unadjusted means in some technology areas may be less reliable indicators of central tendencies than are the unadjusted means in the entire set of general patents.

It is apparent that Internet business method patents had significantly more patent references than patents in eight technology areas, an insignificantly different number than in five areas, and significantly fewer in only one. They had more nonpatent references than patents in twelve technology areas, an insignificantly different number than in one area, and fewer in one. The comparison of total references yielded similar results, with more references in eleven technology areas and an insignificantly different number in three.

These findings do not support the conventional wisdom that Internet business method patents inadequately cite the prior art, when compared either with the average patent or with patents in particular technology areas. These findings appear, in fact, to support the contrary position that Internet business method patents cite more prior art.

B. Types of Nonpatent Prior Art References

Because the universe of potential sources for nonpatent prior art is so expansive, any evidence that applicants may have simply “dumped” large numbers of inferior references into Internet business method patent applications would be more likely to show up in references to sources other than prior patents. This is especially true in view of the fact that the margin by which Internet business method patents cite more *nonpatent* prior art references than general patents far exceeds the margin by which Internet business method patents cite more *patent* prior art references than general patents. Thus far, all of our measures have been purely quantitative. Although it would be desirable to precisely assess and compare the relevance, objectivity, and reliability of information in all of the prior art in our two data sets, the reasons why this is impossible are too obvious to require explanation.

There cannot be even a rough assessment of informational relevance, objectivity, and reliability in the case of references to prior *patents*, because there is no feasible way to segregate these references into different categories on the basis of their probable informational quality. In the case of *nonpatent* prior art, however, it is possible to make at least a rough appraisal of quality, because these types of prior art references can be placed into categories based on their likely informational objectivity and reliability (but not relevance).¹⁸⁹ Such an appraisal is necessarily subjective and thus must be interpreted with caution, but it can nonetheless be worthwhile.

189. To assess the relevance of nonpatent prior art references, we would have to employ an expert in the particular area of technology (or technologies) involved in each patent.

After a trial run in which we closely studied the types of nonpatent prior art in more than 100 Internet business method patents and an equal number of general patents, we classified nonpatent references based on the nature of the sources we found in those patents. After deciding on the categories and their definitions, we categorized the nonpatent prior art in a random twenty percent sample of Internet business method patents (219 out of 1,093) and a random twenty percent sample of general patents (200 out of 1,000). Following are explanations of the eight categories of nonpatent prior art we identified and the tabular results of our comparisons.¹⁹⁰

- 1) Acad/Trade: Academic or trade books, book chapters, articles, and proceedings papers. This category represents publications of a type for which there is an *independent intermediating influence* such as one or more editors or referees to increase the probability of accuracy and objectivity. We did not differentiate between academic and trade publications, because there is much overlap, collaboration and cross-fertilization between academic and industry research efforts. We also often found them difficult to distinguish, and decided to focus on whether a publication is likely to have been independently screened for accuracy and objectivity.
- 2) Comp/Indus: Company/industry publications, press releases, web sites, and advertisements. This category includes publications that were produced by individual companies or industry groups and published with *no independent intermediating influence* to increase the probability of accuracy and objectivity. Publications in this category are, therefore, likely to be less reliable on average than those in the first category. It does not include software and software documentation, however, because we believed the latter to be sufficiently distinct from and probably more reliable than other types of publications from companies or industry groups.
- 3) Univ Pubs: Publications from universities or consortia of universities, such as those from university research labs, particular departments (such as computer science), and individual faculty, as well as theses and dissertations. Although originating with

190. When we were uncertain about how to categorize a given source or when it was of unknown origin, we did a considerable amount of research on the Internet in addition to speaking with knowledgeable colleagues to better understand what particular sources were and who was responsible for them. There naturally remained a small number that we could not characterize.

- universities, these are not peer-reviewed publications. They nevertheless are likely to be of relatively high average quality compared with many other sources.
- 4) Gov Doc: Government documents/publications and websites, except for published patent applications and search reports. This category includes U.S. and foreign government publications, plus those of international government organizations such as the WIPO.
 - 5) SW: Software programs and software documentation. These are separated from other company or industry sponsored publications because of their functional nature and obvious need for a high degree of accuracy and reliability.
 - 6) PP: Popular press. Includes not only newspapers, magazines, and other publications of general interest, but also news publications aimed at general business and legal audiences.
 - 7) Pat Apps/Search Rpts: Published patent applications and patent office search reports, such as Patent Cooperation Treaty (“PCT”) and European Patent Office (“EPO”) search reports.
 - 8) Oth: Other publications. Includes varied items such as individual web pages, but most references placed in this category are those in which we found insufficient information to determine the exact identity of the item, even after we conducted a web search of key names and terms in the incomplete reference. One example is a reference to a partial title of an item, followed by “found on the web on x date.”

1. *Internet Business Method Patents Compared With the Average Patent*

In Table 3, we present descriptive statistics based on our analysis of the different types of nonpatent prior art references in each data set.¹⁹¹

191. In Tables 3 and 4 we report only descriptive statistics. We did not formally test for statistical significance in our comparison of nonpatent prior art categories, for several reasons. First, Table 3 reflects the fact that taking a 20% patent sample in each data set and dividing the nonpatent prior art into 8 categories produced numbers too small for statistical testing, except in the case of Acad/Trad and Comp/Ind nonpatent prior art references. We possibly could test the difference between the percentage of Acad/Trade references in the two data sets, and the difference between the percentage of Comp/Ind references in the two data sets, but the results would be suspect. The reason is that, when formal statistical testing is done, each observation should be independent of other observations.

When, for example, we compared the mean number of claims in the two sets of patents, each observation was a number of claims in a specific patent. This patent was independent of other patents, one observation not influencing another observation. Or, when we compared the mean number of patent and nonpatent prior art references in the two sets of patents, each observation was a number of references in a specific patent. In the set of general patents, the assumption of independence of observations can be made with exceptionally high confidence because, when taking a random sample of one thousand patents out of a population of approximately 235,000 patents issued during a two-year period from mid-1996 to mid-1998, the chance that any two of those patents were, for example, drafted by the same attorney for the same client and perhaps deriving from the same original patent application is close to zero. The assumption of independence cannot be made with the same exceptionally high level of confidence in the case of our data set of 1,093 Internet business method patents, because we sought to include the entire population of Internet business method patents issued through the end of 1999. In this data set, there is a somewhat greater probability that, in a small number of instances, two of these patents may have originated from the same original application drafted by the same attorney for the same client. Thus, there is a chance that, in a few instances, the number of claims or references in two patents may not be totally independent of one another. After carefully examining all of the Internet business method patents to check for this possibility, however, we found that this was a rare occurrence, so the assumption of observation independence is still very strong, and our results untainted, despite the somewhat greater probability of some interdependence.

With this understanding of the importance of observation independence to valid statistical testing, one can see that the validity of such testing would be very questionable when the unit of study is the *qualitative type* of nonpatent prior art reference in a patent. Here, the unit of study is not a number of something in a patent, as compared with the number of something in another distinct patent. Instead, the unit of study is the qualitative and subjectively determined category of nonpatent prior art that has been chosen by the inventor and patent attorney for citation. There often will be hundreds of potential sources for nonpatent prior art, and no one can check them all. A perfect prior art search is impossible, and a hit-or-miss prior art search is far more typical. Within a given patent, the types of nonpatent prior art references cited by the applicant's patent attorney are almost certainly not independent of one another. The attorney's decisions, presumably

**Table 3: Types of Nonpatent Prior Art References:
Internet Business Method Patents Compared with General Patents**

	Acad/ Trade	Comp/ Indus	Univ Pubs	Gov Docs	SW	PP	Pat Apps/ Search Rpts	Oth	Total
Internet Business Method Patents									
Number	1084	308	29	34	106	206	22	30	1819
Percentage	59.6%	16.9%	1.6%	1.9%	5.8%	11.3%	1.2%	1.6%	100%
Mean	4.95	1.41	0.13	0.16	0.48	0.94	0.10	0.14	8.31
Median	1	0	0	0	0	0	0	0	3
Std. Dev.	18.88	4.21	0.47	0.73	2.09	8.56	0.50	0.56	30.72
General Patents									
Number	373	109	5	6	0	5	22	12	532
Percentage	70.1%	20.5%	1%	1%	0%	1%	4.1%	2.3%	100%
Mean	1.86	0.55	0.03	0.03	0.00	0.03	0.11	0.06	2.66
Median	0	0	0	0	0	0	0	0	0
Std. Dev.	5.61	3.30	0.23	0.24	0.00	0.19	0.56	0.65	7.74

As expected from the much larger number of overall nonpatent prior art references in Internet business method patents, these patents also have far more of each category of nonpatent reference except for published patent applications and search reports.¹⁹² A comparison of the percentages

aided by the inventor, about which items of nonpatent prior art to cite in the patent are almost certainly influenced by some common factors, such as which academic or trade journals the inventor was in the habit of checking, perhaps because of more convenient availability. Although patent examiners rarely add citations to *nonpatent* sources of prior art, the same lack of independence will characterize an examiner's selection of any such references.

192. The PTO did not begin publishing patent applications (eighteen months after filing) until March 2001. Our data thus could not include references to U.S. patent applications in either of our data sets. To our knowledge, only Japan recognized business methods as patentable subject matter at the end of 1999 (the end point of our Internet business method data). The source of search reports is typically Europe (where business

of cited prior art in each category reveals nothing particularly odd. The fact that Internet business method patents cite proportionately more references to the popular press (11.3% of total nonpatent prior art references) than do general patents (1.0 %) is arguably a detraction from the overall quality of this prior art in the former. The negative connotation of “popular press” as a source of prior art is ameliorated by two factors, however. First, the newsworthiness of business method patents during the relevant years suggests that at least some references to other developments in the field would show up in the popular press. Second, our definition of the term popular press includes some sources that are likely to be more accurate and reliable than magazines and newspapers for general readership.¹⁹³

Aggregating the different types of nonpatent prior art into two broad classes based on the probable quality of the prior art reveals even more about the general quality of prior art citations in Internet business method patents. Those that, on average, are likely to be the most objective and reliable—Acad/Trade, Univ Pubs, Gov Docs, SW, and Pat Apps/Search Rpts—can be combined into a “higher quality” category. Those that, on average, are likely to be the least objective and reliable—Comp/Indus, PP, and Oth—can be combined into a “lower quality” category. The result of the aggregation is that 70.2% of the nonpatent prior art in Internet business method patents is in the higher quality category, and 29.8% in the lower quality category. On the other hand, 76.1% of the nonpatent prior art in the average patent is in the higher quality category, and 23.9% in the lower quality category. An argument thus can be made, albeit a weak one, that nonpatent prior art in Internet business method patents was of lower quality than that cited in the average patent.

2. *Internet Business Method Patents Compared With Patents in Individual Technology Areas*

The weakness of the argument that general patents cite higher quality nonpatent prior art than Internet business method patents was prominent when we compared Internet business method patents with those in each of the fourteen individual technology areas within the set of general patents. Before proceeding, we note that triple division in the general patent data set—taking a twenty percent sample, dividing into eight categories of nonpatent prior art, and further dividing into fourteen technology areas—

method patents are not recognized) or Japan, so it is surprising that our Internet business method data set included any references at all in this category.

193. We included not only newspapers, magazines, and similar publications for the general population, but also news publications targeted at business and legal audiences.

produced numbers too small in some instances for meaningful comparison between Internet business method patents and patents in individual technology areas.¹⁹⁴ In several cases, however, the number of patents in a sample of a particular technology area and the number of nonpatent prior art references in some categories were adequate for rough comparison. Moreover, comparison becomes more meaningful if we aggregate the types of nonpatent prior art into “higher-quality” and “lower-quality” categories as we did when comparing Internet business method patents with the average patent.

In Table 4, we denote the combined group of higher quality reference types as “Higher Quality P.A.,” and denote the combined group of lower quality reference types as “Lower Quality P.A.” Beside each technology area, we note the number of patents within that area in the twenty percent sample.

**Table 4: Types of Nonpatent Prior Art References:
Internet Business Method Patents Compared with Patents in Each Technology Area**

	Higher Quality P.A.	Lower Quality P.A.
Internet Business Methods	70.2%	29.8%
General Patents	76.1%	23.9%
Chemistry (59)	96.4%	3.6%
Computer-Related (30)	61.7%	38.3%
Electronics (30)	24.5%	75.5%
Mechanics (82)	27.3%	72.7%
Medical Devices (14)	60.3%	39.7%
Optics (23)	74.0%	26.0%
Pharmaceutical (22)	97.4%	2.6%
Semiconductors (16)	86.6%	13.4%

194. The number of patents and/or the number of references in the 20% sample was too small for any kind of comparison in the acoustics (four patents), automotive-related (fourteen patents but very few nonpatent references), biotechnology (four patents), communications-related (two patents), energy-related (four patents), and software (four patents) technology areas.

The data reported in Table 4 suggest that only patents in the areas of chemistry, pharmaceuticals, and semiconductors cite nonpatent prior art of higher *quality* than Internet business method patents. The observation of probable higher quality is most meaningful in the case of chemistry patents because the number of such patents in the sample of that category is much larger (fifty-nine) than the number of patents in the pharmaceutical (twenty-two) and semiconductor (sixteen) areas. In these three areas, however, chemistry and semiconductor patents contain a substantially lower *quantity* of nonpatent prior art references than Internet business method patents, while pharmaceutical patents do not cite a significantly different *quantity* of nonpatent prior art than Internet business method patents. These findings do, however, support an argument that pharmaceutical patents (equal quantity, greater quality) and chemistry patents (lower quantity, greater quality, *and* a relatively large number of patents in the sample) cited better nonpatent prior art than Internet business method patents.

Overall, we find little basis for concluding that the nonpatent prior art cited in Internet business method patents was substantially inferior to that cited in most other kinds of patents. One should, however, take any findings from our data on the different types of nonpatent prior art with caution, because these particular comparisons are much less precise than our other findings.

C. Total Patent Claims

Claims delineate the boundaries of the property right granted to the patentee. Even without empirical support, one can logically argue that a larger number of claims may be an indicator of patent value. When a patent owner sues for infringement, she does not sue for infringement “of the patent.” Instead, she sues for infringement of one or more identified claims in the patent. Each patent has at least one *independent* claim, and usually has several.¹⁹⁵ The invention is defined in its broadest sense by the independent claims. Although § 121 of the Patent Act calls for only one distinct invention per patent,¹⁹⁶ a patent may contain two or more independent claims because (a) it is possible to claim some inventions using different formats,¹⁹⁷ (b) a patent may claim a process and a product

195. For example, the average number of independent claims in the general patent data set was 2.75.

196. 35 U.S.C. § 121 (2000).

197. Software patents are an example. A software invention may be claimed as a device (or machine) itself, a device incorporating software, a system, a process, or any combination of the above. *See, e.g.,* AT&T Corp. v. Excel Communications, Inc., 172

made from that process, (c) a patent may claim a product and a process for using the product, or (d) a patent may claim closely related versions of the invention.¹⁹⁸

Typically, a series of successively narrower dependent claims follows each independent claim. Dependent claims are narrower because they contain further restrictions or limitations on the scope of the independent claim from which they flow.¹⁹⁹ Because of the one-invention-per-patent rule, the average number of independent claims does not vary across technology areas or across time nearly as much as dependent claims. Research has shown, for instance, that a dramatic increase in the average number of total claims over a recent twenty-year period is almost entirely attributable to an increase in *dependent* claims.²⁰⁰

Because independent claims are broader, any given independent claim should encompass more potentially infringing behavior or products made by others. This does not mean, however, that the total number of claims is not an indicator of patent value. Broader claims run a greater risk of encompassing uncited prior art and being held invalid in court than narrower ones, which could still leave the patent owner with quite a few narrower claims that are valid and infringed. Thus, the larger the number of total claims, the more opportunities there are for the activities of others to infringe. This fact of more claims thus should translate into greater value for the patent owner.

F.3d 1352 (Fed. Cir. 1999) (describing a process); *In re Alappat*, 33 F.3d 1526 (Fed. Cir. 1994) (en banc) (describing a machine). The “means-plus-function” claim drafting format, which is allowed by 35 U.S.C. § 112, ¶ 6 (2000), is also used in some instances and has been employed frequently in software patents. However, even identifying, much less interpreting, claims of this type is fraught with much uncertainty and the topic is far beyond the scope of the present Article. See generally Fidel D. Nwamu, *Does Your Claim Conform To Means-Plus-Function Format Under Section 112, Paragraph Six?: O.I. Corp. v. Tekmar Co.*, 6 J. INTELL. PROP. L. 189 (1999).

198. For example, U.S. Patent No. 5,566,752 (issued Oct. 22, 1996) (“High heat density transfer device”), includes an independent claim on a diamond substrate semiconductor chip device (followed by a series of successively narrower dependent claims), and another independent claim on a module comprised of several of these devices working together (again followed by a series of dependent claims).

199. See 35 U.S.C. § 112, ¶¶ 2-5. Dependent claims may depend on an independent claim or on another dependent claim.

200. Allison & Lemley, *Growing Complexity*, *supra* note 40, at 103-04. The average number of independent claims in the general patent data set was 2.75 and the average number of dependent claims was 12.12. In contrast, the average numbers of independent and dependent claims in an identical data set from twenty years earlier were 2.09 and 7.85, respectively.

We compare only total claims in this study, because previous research finding a positive relationship between the number of claims and value used total claims as the metric and did not differentiate between independent and dependent claims.²⁰¹ Furthermore, previous research in economics has found a significant correlation between the total number of claims and the probability that the patent will be the subject of litigation.²⁰² A new study comprehensively comparing litigated and unlitigated patents finds that litigated patents have far more claims than unlitigated ones.²⁰³ That study also finds a clear association between other measures of patent value and the incidence of infringement litigation.²⁰⁴ Thus, the number of claims may be indicative of patent value.

The litigation-value relationship also makes intuitive sense. Patent litigation is one of the most, if not the most, expensive forms of litigation, and a rational patent owner would be expected to pursue litigation only when the perceived value of the patent is high.²⁰⁵ We therefore believe

201. See, e.g., Lanjouw & Schankerman, *Characteristics of Patent Litigation*, *supra* note 34, at 141.

202. LANJOUW & SCHANKERMAN, *STYLIZED FACTS OF PATENT LITIGATION*, *supra* note 38, at 18; Lanjouw & Schankerman, *Characteristics of Patent Litigation*, *supra* note 34, at 141. This research compared litigated and unlitigated patents only within the same areas of technology. The research by Lanjouw and Schankerman used the 4-digit International Patent Classification (“IPC”) to define areas of technology, and made comparisons between litigated and unlitigated patents within particular four-digit IPCs.

203. Allison et al., *Valuable Patents*, *supra* note 30 (manuscript at 22-23).

204. *Id.* One other factor that may predict a greater average propensity toward litigation is the degree of information asymmetry, or divergence in the parties’ expectations about the likely outcome. The magnitude of information asymmetry is likely to be greater in the case of emerging technologies and technologies in which patent protection is relatively new. LANJOUW & SCHANKERMAN, *STYLIZED FACTS OF PATENT LITIGATION*, *supra* note 38, at 4. To the extent that this is true, one should expect a higher probability of litigation involving Internet business method patents independent of other variables. Tracking actual litigation rates for the patents in our current data set was not within the envisioned scope of this study and, indeed could not have been because of their recency—most were issued during 1998-99.

205. See AIPLA, 2003 ECONOMIC SURVEY, *supra* note 33, for recent estimates of patent infringement litigation costs. Patent litigation may be initiated by either the patent owner or a challenger (accused infringer). When the owner sues for infringement, the defendant typically counterclaims for a declaratory judgment of non-infringement and invalidity. When patent litigation is initiated by a challenger, the challenger seeks a declaratory judgment of non-infringement and invalidity (the same thing sought by the defendant in a patent infringement suit brought by the patent owner), and the patent owner typically counterclaims for infringement. The challenger has standing to act as a plaintiff to seek a declaratory judgment only when it has been threatened with an infringement action, however, so the value-litigation relationship should apply equally to litigation initiated by a challenger.

that the potential private value of innovations perceived by patent applicants and owners will frequently produce higher actual value. More claims also cost the applicant more money. PTO fees increase with the number of claims, and attorney fees increase with the additional time necessary for drafting and prosecuting more claims.²⁰⁶ Choosing to devote these additional resources is within the control of the applicant.

The influence of patent applicants and owners on indicators of patent value also explains the commonly existing correlations among patent attributes that are used to measure patent quality and value. For example, this is not the only study to find that a set of patents having more average claims usually has more average references and more of certain other indicators.²⁰⁷ Correlation does not mean causation, of course, and there is no reason to believe that the existence of more references actually causes patents to have more claims, or vice-versa. Correlation, or “interaction,” among variables cannot be ignored, however. Suppose that, in one data set, variables X, Y, and Z are each found to be significantly different from counterpart variables in a comparison data set. If there are correlations among two or more of these X, Y, and Z variables in one data set, such interactions may diminish or even destroy the earlier finding of significance. Multivariate statistical tests of several types can be used to ascertain not only the existence of these interactions among variables within the same data set, but also whether findings of significance remain valid after accounting for the correlations. In our study, such tests revealed that the significance of differences between the number of *references* in the two data sets, and the significance of differences between the number of *claims* in the two data sets remained after accounting for the correlations between these two variables within each data set.

Federal courts apply a two-step analysis to ascertain whether they have jurisdiction over a patent challenger’s declaratory judgment action for non-infringement or invalidity: “(1) an explicit threat or other action by the patentee, which creates a reasonable apprehension on the part of the declaratory plaintiff that it will face an infringement suit, and (2) present activity which could constitute infringement or concrete steps taken with the intent to conduct such activity.” *See, e.g., Cygnus Therapeutic Sys. v. ALZA Corp.*, 92 F.3d 1153, 1159 (Fed. Cir. 1996), *overruled in part on other grounds by Nobelpharma AB v. Implant Innovations*, 141 F.3d 1059, 1068 (Fed. Cir. 1998); *see also* 28 U.S.C. § 2201 (2000) (stating that a federal district court has jurisdiction to hear a declaratory judgment action by a plaintiff when there is an independent basis for federal jurisdiction—such as the patent statute—and when there is an actual “case or controversy”).

206. We later show that the number of claims significantly correlates positively with the time patents spend in the PTO. *See infra* note 240 and accompanying text.

207. This correlation is found, and strongly so, in a new study comparing litigated and unlitigated patents. *See Allison et al., Valuable Patents, supra* note 30. These correlations existed in both the Internet business method data set and the general patent data set.

Moreover, the existence of these correlations further supports the theory that a significant part of private patent value is a self-fulfilling prophecy—“self-selected value”—because applicant perceptions of greater value are probably a common cause of both the larger number of references and the larger number of claims.

One additional fact provides yet more support for the idea that the total number of claims, like the number of prior art references, is an indicator of patent value. The number of references and claims in a patent are largely within the control of the applicant—the contributions of the examiner have only minor effects on these numbers in most cases. Other studies have found that the number of “forward citations”—the number of times a patent is itself cited as prior art in later patents—is a significant predictor of litigation propensity, and thus is also a probable value indicator.²⁰⁸ Recent research comparing litigated and unlitigated patents not only confirms the relationship between the number of forward citations and litigation propensity (and indeed finds the relationship to be even stronger than discovered in previous research), but also finds that litigated patents have far more references and claims than unlitigated patents.²⁰⁹ Thus, value indicators over which the applicant has much control are confirmed by a value indicator over which the applicant has very little direct control after the patent issues.²¹⁰

208. See LANJOUW & SCHANKERMAN, *STYLIZED FACTS OF PATENT LITIGATION*, *supra* note 38; Allison et al., *Valuable Patents*, *supra* note 30; Lanjouw & Schankerman, *Characteristics of Patent Litigation*, *supra* note 34.

209. Allison et al., *Valuable Patents*, *supra* note 30.

210. Patent owners may, however, cite their own previous patents as prior art in their later ones, and the study by Allison et al., did not exclude these self-citations. See Allison, et al., *Valuable Patents*, *supra* note 30 (manuscript at 36). Others have found, however, that more valuable litigated patents not only receive more self-citations in later patents with the same owners but also in later patents owned by others. Moreover, the greater number of self-citations in more valuable litigated patents is even more pronounced than the larger number of citations by others. *See id.* Self-citations may themselves be a value indicator by revealing that a given patent is part of a group of related ones owned by the patentee, and the whole can be more valuable to the owner than the sum of its parts.

Moreover, patent owners may have some very indirect control over how often their patents are later cited as prior art because there is evidence that their filing of infringement actions sends signals causing others to cite their patents more often. Lanjouw & Schankerman, *Characteristics of Patent Litigation*, *supra* note 34. Such signaling adds yet more value to a patent. Thus, there is no reason to exclude self-citations from the total of citations received by a patent unless one wishes to explore the differential contributions to patent value of self-citations and citations by others.

1. *Internet Business Method Patents Compared With the Average Patent*

Table 5 reports the results from our comparison of total claims in Internet business method and general patents.²¹¹

**Table 5: Total Number of Claims:
Internet Business Method Patents Compared with General Patents**

	Mean	Median	Std. Dev.	Min.	Max.
Internet Business Method Patents	26.26	20	22.45	1	375
General Patents	14.87	12	11.47	1	120

Using several statistical techniques, we found that the Internet business method patents contain a far larger number of total claims than the average patent. The results were highly significant.²¹²

2. *Internet Business Method Patents Compared With Patents in Individual Technology Areas*

We then compared average total claims in Internet business method patents with average total claims in each of the fourteen technology areas within the general patent data set. Here, too, the results were notable, with Internet business method patents having far more claims than patents in each area.²¹³ Table 6 shows the untransformed means and other descriptive statistics for claims in Internet business method patents and patents in each technology area.

211. Again, we report untransformed means in the table although we used log-transformed means in testing.

212. As before, we performed an independent-groups *t*-test using log-transformed total claims and a Wilcoxon nonparametric test, the results of both tests being $p < 0.0001$.

213. These findings were highly significant statistically except in the area of acoustics. Although Internet business method patents also had many more claims than acoustics patents, the fact that there was a relatively small number of patents in the acoustics category and that these patents had substantially more claims than other areas caused the difference to be not statistically significant.

**Table 6: Total Number of Claims:
Internet Business Method Patents Compared with Patents in Each Technology Area**

	Mean	Median	Std. Dev.	Min.	Max.
Internet Business Methods	26.26	20	22.45	1	375
Acoustics	19.24	17	9.32	7	43
Automotive-Related	14.99	13	10.03	2	55
Biotechnology	13.30	9	9.42	3	36
Chemistry	15.19	13	10.96	1	82
Communications	15.82	10	13.11	1	49
Computer-Related	16.43	12.5	14.31	1	120
Electronics	15.11	13	11.07	1	57
Energy-Related	15.83	13	9.41	3	40
Mechanics	13.20	11	9.25	1	55
Medical Devices	17.48	14	12.57	1	61
Optics	16.83	14	14.53	1	120
Pharmaceutical	14.68	11	12.93	1	82
Semiconductors	15.34	12.5	11.57	2	64
Software	17.75	12	17.82	1	120

D. Number of Inventors

Prior research has not suggested using the average number of inventors as an indicator of patent value. One can argue plausibly, however, that the number of inventors may be related to the amount of costly resources devoted to the inventive enterprise. As already noted, it is likely that, on average, patent applicants who perceive greater innovation value commit greater resources to obtaining stronger and broader patents, and that this resource commitment may translate into greater actual patent value. Various human resource costs presumably rise with the number of inventors. The number of inventors may also affect patenting costs in several ways. A larger number of inventors can mean increased attorney fees because patent attorneys must interview and work with the inventors; while an attorney may not need to work closely with each of a group of

multiple inventors in every case, the average number with whom the attorney must work surely increases as the total increases. The attorney may have to spend more time determining who the true inventors actually are,²¹⁴ and there may be more uncertainty in the determination of various facts such as the best mode for carrying out the invention.²¹⁵ A larger number of inventors likely affects litigation costs, as well. The presence of

214. Inventors must make an oath in the patent application that they are the true inventors. 35 U.S.C. § 116 (2000). A patent attorney must determine whether those named as inventors really are such, and whether anyone else should also be named as an inventor. Incorrect inventorship may be corrected if inadvertent, but the deceptive omission of someone who should be named, or inclusion of someone who should not be, invalidates the patent *See id.* § 256. It can be far from obvious who is entitled to assert inventor status. The fact that one is entitled to inventor status if, when he is working with another inventor, he makes a significant contribution to conception in at least one of the patent's claims, but not if he assists with the inventive process but does not make such a contribution, coupled with the common complication of multiple patents on related inventions, can cause inventorship questions to be quite difficult. *See, e.g., Trovan, Ltd. v. Sokymat SA*, 299 F.3d 1292 (Fed. Cir. 2002). In *Trovan*, the Federal Circuit remanded the question of inventorship to the district court because of an error caused by incorrect claim interpretation. Claim interpretation, a dicey proposition itself, is prerequisite to resolution of this and most other issues in patent litigation. The court observed:

When two or more persons make an invention jointly, they must apply for a patent jointly. Co-inventors need not physically work together or at the same time, make the same type or amount of contribution, or make a contribution to the subject matter of every claim of the patent. Because conception is the touchstone of inventorship, each joint inventor must generally contribute to the conception of the invention. Conception is the formation in the mind of the inventor, of a definite and permanent idea of the complete and operative invention, as it is hereafter to be applied in practice. Conception is complete when one of ordinary skill in the art could construct the apparatus without unduly extensive research or experimentation. An inventor may solicit the assistance of others when perfecting the invention without losing any patent rights. However, the basic exercise of ordinary skill in the art, without an inventive act, does not make one a joint inventor.

Id. at 1301-02 (internal quotation marks and citations omitted).

215. In addition to disclosing enough detail in the patent application to enable an ordinarily skilled practitioner in the art to make the invention and put it into practice ("enablement"), the inventor(s) must also disclose the best mode contemplated for carrying out the invention. 35 U.S.C. § 112 ¶ 1. The best mode requirement has both subjective and objective elements. First, the requirement applies only if the evidence indicates that the inventor(s) had in fact contemplated, or envisioned, a best mode (or "preferred embodiment") at the time the application was filed. Second, the disclosure of the best mode must be objectively adequate. *See, e.g., Chemcast Corp. v. Arco Indus. Corp.*, 913 F.2d 923 (Fed. Cir. 1990). As the number of inventors increases, so does the possibility of disagreements among them about whether a particular method of putting the invention into practice is the best mode, or even whether there is an envisioned best mode at all.

more inventors means more depositions, more time spent on trial testimony, and the increase of other costs associated with litigation.²¹⁶

Furthermore, prior research has found that inventions in certain fields, namely, pharmaceuticals, biotechnology and chemistry, which require more advanced science and engineering and thus more expensive inventive processes than average, also tend to involve more than the average number of inventors. Conversely, the average number of inventors is smaller in fields such as pure mechanics and acoustics that typically involve lower-level skills and less expensive inventive processes.²¹⁷ Furthermore, inventors in areas such as pharmaceutical, biotechnology, and chemistry are likely to be better paid than most because of their skill and training, and they or their employers are likely to have higher opportunity costs than average. Therefore, there is sufficient theoretical reason to believe that the average number of inventors in a group of patents could have an association with patent value to warrant statistical testing.

1. *Internet Business Method Patents Compared With the Average Patent*

Table 7 summarizes the results of our comparison of the number of inventors in Internet business method patents and general patents, using untransformed means.²¹⁸

216. More trial witnesses also may translate to larger attorney fees, and may increase many other incidental costs such as those related to travel.

217. Allison & Lemley, *Who's Patenting What?* *supra* note 40, at 2129, 2157-58. Allison and Lemley found that inventions in fields of pharmaceuticals (2.81), biotechnology (2.84), and chemistry (2.85) had significantly more inventors than the average for patents-in-general (2.25). *Id.* Mechanical inventions (1.77) and those in the area of acoustics (1.86) had significantly fewer inventors than average. *Id.* Inventions in the other fields, such as those in the clearly more complex area of semiconductors and the clearly less complex area of traditional electronics, had numbers of inventors that were not significantly different than average, and thus, neither confirmed nor disconfirmed the proposition. *Id.*

218. Because the distributions of number of inventors in both data sets looked very similar and were not heavily influenced by upper outliers as were other distributions in our study, logarithmic transformation of the means was not necessary.

**Table 7: Number of Inventors:
Internet Business Method Patents Compared With General Patents**

	N	Mean	Median	Std. Dev.	Min.	Max.
Internet Business Method Patents	1093	2.43	2	1.82	1	14
General Patents	1000	2.26	2	1.52	1	11

The average number of inventors in Internet business method patents was greater than in general patents by a statistically significant margin.²¹⁹

2. *Internet Business Method Patents Compared With Patents in Individual Technology Areas*

We also compared this variable in Internet business method patents with patents in each of our fourteen technology areas. Table 8 reports the results of this comparison.²²⁰ Because this comparison is with patents in the individual technology areas rather than with the entire sample of general patents, the differences necessarily must be larger to be statistically significant.

219. Although we are confident of the result, it must be taken with a bit of caution because it was confirmed by only one of the two follow-up tests. The Independent Groups *t*-test resulted in $p < 0.0185$. However, a Wilcoxon nonparametric test showed no significant difference ($p < 0.1756$). We also employed the Savage test, another nonparametric test that analyzes Savage scores—the expected order statistics for the exponential distribution. This test was chosen because these Savage scores tend to cluster at a single score with only a small percentage of subjects with scores in between the minimum and the maximum. This distribution of values closely approximates an exponential distribution. Unlike the nonparametric Wilcoxon test, the Savage test yielded results that were statistically significant ($p < 0.0178$).

220. As before, the table relates untransformed means. Unlike the comparison of number of inventors between Internet business method patents and patents in general, log transformation of the means was necessary when the comparison was made with patents in individual technology areas because the distributions were characterized by more extreme variations.

Table 8: Total Number of Inventors: Internet Business Method Patents Compared with Patents in Each Technology Area

	Mean	Median	Std. Dev.	Min.	Max.
Internet Business Methods	2.43	2	1.82	1	14
Acoustics	1.81	2	0.98	1	4
Automotive-Related	2.05	2	1.29	1	7
Biotechnology	2.84	2	1.91	1	11
Chemistry	2.86	2	1.69	1	10
Communications-Related	2.23	2	1.39	1	7
Computer-Related	2.25	2	1.47	1	10
Electronics	2.37	2	1.54	1	10
Energy-Related	2.70	2	2.08	1	9
Mechanics	1.79	1	1.22	1	9
Medical Devices	2.29	2	1.67	1	10
Optics	2.59	2	1.72	1	10
Pharmaceutical	2.92	2	1.94	1	11
Semiconductors	2.17	2	1.39	1	7
Software	2.52	2	1.61	1	10

Overall, these results reveal more similarity than dissimilarity. Internet business method patents had significantly more inventors than patents in one technology area (mechanics), significantly fewer inventors than patents in three areas (biotechnology, chemistry and pharmaceuticals), and an insignificantly different number of inventors than patents in the other ten technology areas. It is difficult to draw any meaningful conclusions from these data. Although Internet business method patents had significantly more inventors than patents in general, that result was not as compelling as in comparisons of other variables, and the theory largely falls apart when the number of inventors in individual technology areas is examined. We can say, though, that to the extent that the average number of inventors has something to say about the value of patents, there is

nothing in these data to suggest that Internet business method patents are clearly worth less.²²¹

E. Time Spent in the Patent Office

The time that an application spends in the PTO before patent issuance is often referred to as “time-in-prosecution” or “pendency time.” Although no prior research has sought to determine whether pendency time might be related to patent quality or value, Allison and Lemley surmised that one reason for the dramatic growth in the complexity of all aspects of patenting over time, including average pendency times, could be an increase in the value of patents to those who seek and obtain them.²²² There is little doubt that patents have become more valuable assets in the past couple of decades.²²³ Moreover, the recent study by one of the authors comparing litigated and unlitigated patents finds that more valuable litigated patents spent much more time in the PTO before issuance than did unlitigated patents.²²⁴ We will discuss some of the factors that influence pendency times and the consequences of pendency variations before turning to comparisons between Internet business method patents on the one hand, and the average patent and patents in particular technology areas on the other.

1. *The Effect of Ancestor Applications on Average Pendency Times*

For many patents, pendency time is affected by the filing of co-pending “continuing” applications. Patent law in the U.S. allows an applicant to file another application covering a closely related invention while the first application is still pending. There are three types of continuing applications. The first, commonly called a “continuation,” uses the same written description as its predecessor application, but makes changes in the claims. Continuation applications retain the priority date of the original filing as long as they are filed while at least one predecessor application in the chain is still pending, and as long as there is no material

221. Furthermore, the idea that this metric may be an indicator of patent value cannot be totally dismissed, and warrants further empirical research. However, the new study comprehensively comparing litigated and unlitigated patents finds that patents with more inventors are *not* more likely to be litigated, thus failing to support the hypothesis that the number of inventors is an indicator of value. Allison et al., *Valuable Patents*, *supra* note 30 (manuscript at 47).

222. See Allison & Lemley, *Growing Complexity*, *supra* note 40, at 98-99, 139-41.

223. *Id.*

224. Allison et al., *Valuable Patents*, *supra* note 30 (manuscript at 44).

change in the written description of the invention.²²⁵ An applicant may do this on her own initiative in order to get a shot at receiving another patent on a closely related invention—some continuation applications are abandoned and some result in their own separate patents—or she may do so in response to a rejection or limitation by the patent examiner. A patent applicant may also file a continuing application to reconfigure the claims so that they are more likely to encompass a competitor's activities, thus increasing the universe of potential infringers.²²⁶

The second type of continuing application is a “continuation-in-part” (“CIP”). Although a CIP also claims a closely related invention, the written description has been materially changed to support one or more changes in the claims. The inventor does not retain the original filing date for any claim in the CIP that draws its support from a material change in the description. Thus, when one receives a patent based on a CIP, it is possible for different claims in the patent to have different priority dates. Because the term of patent protection now dates from the original priority filing date, it is likewise possible for different claims in a patent to have different terms of protection.

The third type of continuing application is a “divisional.” The Patent Act requires that there be only one distinct invention per patent.²²⁷ A divisional application represents an election by the applicant to keep only certain claims in the instant application and to put those claims representing the separate invention in a co-pending divisional application when the examiner issues a restriction based on finding more than one invention is found in an application. Here, again, the applicant retains the original priority filing date if no material change is made in the description of the invention. In other words, the original description is reused because it was adequate to support more than one distinct invention.

An applicant may file one or more of any type of continuing application while any other application deriving from the same original application is still pending. There can be a chain of such continuing applications, or even a tree when one or more of these applications

225. 35 U.S.C. § 120 (2000). *See generally* Quillen & Webster, *Continuing Patent Applications*, *supra* note 139, at 4–6 (discussing the types of continuing applications and their effects). Retaining the earlier priority filing date can be crucial because it determines which prior art can be used against the patent applicant (or, later, the patent owner if its patent is challenged).

226. *See id.*

227. *Id.* § 121. Although representing a fundamental precept of patenting, this requirement can be interpreted by patent examiners with varying degrees of rigidity or looseness.

branches off into multiple others. Looking backward from a given patent, one can see that it sometimes has resulted from a series of prior related applications. Those related applications prior to the one that most immediately led to this patent are often called “ancestors” (or “parents,” “grandparents,” etc.). In our general patent data set, for example, patents resulted from an average of 1.535 U.S. applications. The median, however, was one application, because the majority of patents did not result from a chain.²²⁸

The PTO understates average pendency times in its reporting by only counting the time elapsed between the filing of the most immediate application and the issuance of a patent from it, even when that application is at the end of a chain. Such reporting is inaccurate for two reasons. First, the patent’s origin and duration date back to the first filing with respect to all claims supported by the original written description. Second, when a patent is based not only on the most recent application but also on one or more ancestors, examination of the most recent one should require less of the applicant and the examiner because some of the work has already been done in connection with the ancestors.

2. *Other Reasons for Variation in Pendency Times*

Variations in patent pendency times are caused by several other factors, some of which reflect system-wide changes while others reflect idiosyncrasies in particular patents. Regarding the former, there have been dramatic increases in the number of patent applications in recent years,²²⁹ straining the PTO’s resources and resulting in longer average pendency times in all technology areas.²³⁰ Beginning in mid-1995, the term of patent protection was changed from seventeen years counting from the date of patent issuance to twenty years counting from the original filing of the

228. We did not count the total number of applications in the Internet business method data set.

229. Utility Patent Applications Per Year: 2001 = 324,211; 2000 = 291,653; 1999 = 259,618; 1998 = 238,850; 1997 = 219,486; 1996 = 189,922; 1995 = 220,141; 1994 = 185,087. USPTO, 2001 ANNUAL REPORT: OTHER ACCOMPANYING INFORMATION 107 (2002), available at <http://www.uspto.gov/web/offices/com/annual/2001/01accompinfo.pdf> (last visited Nov. 25, 2003).

230. See Allison & Lemley, *Growing Complexity*, *supra* note 40, at 98-99 (average pendency times increased quite substantially from the late 1970s to the late 1990s). The fact that the PTO has had to deal with increasing numbers of applications is surely not the only reason for increasing pendencies in all technology areas during the past twenty years. Because patents have generally become more valuable assets during that period for variety of reasons, applicants also have probably spent more time in the PTO trying to get the broadest possible protection. *Id.* at 139.

application.²³¹ It has been argued that the change would reduce overall pendency times because delay in the patent office would reduce the term of patent protection and applicants would push to get through the process faster.²³² This has not turned out to be the case. If there has been any such effect at all, it has been overwhelmed by the dramatic increase in the number of applications. Moreover, it is doubtful that the prospect of losing a relatively small portion of the patent term will motivate many patent applicants to push their applications through the PTO more quickly.²³³ The majority of patents do not create market power in the first place, and even in the case of those that do, the passage of time usually witnesses the introduction of various competing alternatives that reduce whatever market power the patent may have conferred.²³⁴

Thus, proportionately small increments in the term of protection will usually be worth little, except for pharmaceutical patents. Prior research has confirmed that patents are far more important in pharmaceuticals than in other areas.²³⁵ Also, pharmaceutical patents, unlike most patents in other fields (including Internet business methods), typically maintain their value throughout the term of protection.

231. 35 U.S.C. § 154(a)(2) (stating that “such grant shall be for a term beginning on the date on which the patent issues and ending twenty years from the date on which the application for the patent was filed in the United States”). The term of seventeen years from issuance was changed to a term of twenty years from initial filing. Patents in force on June 8, 1995, or those issuing from applications still pending on that date, have the longer of these two terms. Those issuing from applications originally filed after that date have the twenty-year-from-filing term.

232. Mark A. Lemley, *An Empirical Study of the Twenty-Year Patent Term*, 22 *AIPLA Q.J.* 369, 392-93 (1994).

233. However, changing the way that the patent term is calculated undoubtedly did reduce the incentive to drag out the patenting process excessively, as had been done by a few who engaged in so-called “submarine” patenting. Submarine patenting consisted of intentionally dragging out the patenting process for an exceptionally long time so that the applicant could spring a patent as a *fait accompli* on members of an industry that had matured while the application had been lingering in the PTO. *See, e.g.*, Allison & Lemley, *Growing Complexity*, *supra* note 40, at 126 n.108.

234. *See, e.g.*, HERBERT HOVENKAMP, *ECONOMICS AND FEDERAL ANTITRUST LAW* § 8.3, at 219 (1985) (“Many patents confer absolutely no market power on their owners The economic case for ‘presuming’ sufficient market power . . . is very weak.”); HERBERT HOVENKAMP ET AL., *IP AND ANTITRUST* ch. 4 (2003 ed.); Nat’l Inst. on Indus. & Intell. Prop., *The Value of Patents and Other Legally Protected Commercial Rights*, 53 *ANTITRUST L.J.* 535, 547 (1985) (“Statistical studies suggest that the vast majority of all patents confer very little monopoly power.”).

235. *See* COHEN ET AL., *WHY U.S. MANUFACTURING FIRMS PATENT*, *supra* note 51; Levin et al., *supra* note 51.

In addition to greater use of continuing applications and the increasing backlog of applications at the PTO during recent years, a number of other factors undoubtedly affect pendency times in either systematic or idiosyncratic ways. For example, we should expect individual examiners to work at varying speeds even if they are in the same field, because of different work habits or amounts of experience. Despite the fact that the impact of examiner heterogeneity on pendency time should be mitigated by PTO rules regulating the amount of time that patent examiners may work on applications, a recent study of patent examiners apparently found some variance among examiners in the time they take to allow the issuance of patents.²³⁶ As the PTO does, the authors of this study evidently considered only the time elapsed from the filing of the last application, failing to account for the effect of ancestor applications, and it is unknown whether the same type of time variances among examiners would exist if ancestor applications had properly been taken into account. Moreover, the study evidently used very rough estimates of filing and issuance dates—the year of filing and the year of issuance—to calculate pendency times.²³⁷ This study thus does not show the extent to which examiner identity may affect patent pendency times in a meaningful or systematic way, and an exploration of this variable is beyond the scope of our own study.

Recently, however, Doug Lichtman found that individual examiners do vary greatly in their tendencies to alter claim language during prosecution, even within the same field of technology.²³⁸ This fact in itself

236. IAIN M. COCKBURN ET AL., ARE ALL PATENT EXAMINERS EQUAL? THE IMPACT OF CHARACTERISTICS ON PATENT STATISTICS AND LITIGATION OUTCOMES 10 (Nat'l Bureau of Econ. Research, Working Paper No. 8980, 2002). We use words like “apparently” and “evidently” because the authors were not very clear in explaining their findings about the effect of examiner heterogeneity on pendency times.

237. We assume that the authors of the study did not take into account the effect of ancestor applications on pendency times because they used the NBER Patent Citation Data File, which reports the filing date of only the most immediate application, not the original priority filing date in the many cases in which there is a chain of ancestors. Moreover, this database measures the elapsed time between the *year* in which this most recent application was filed and the *year* in which the patent was granted, providing a measure of pendency time so rough as to be much less useful than we might like. See BRONWYN H. HALL ET AL., THE NBER PATENT CITATIONS DATA FILE: LESSONS, INSIGHTS AND METHODOLOGICAL TOOLS 8 (Nat'l Bureau of Econ. Research, Working Paper No. 8498, 2001).

238. Douglas Lichtman, *Rethinking Prosecution History Estoppel*, U. CHI. L. REV. (forthcoming 2004), draft available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=455380 (last visited Nov. 30, 2003).

suggests that pendency times probably vary substantially from one examiner to another.

The area of technology undoubtedly has a significant impact on the time a patent spends in the PTO, probably in at least two ways. First, some technologies are just more complicated and difficult to decipher than others.²³⁹ Patents in some fields that are known to be more technically challenging, such as biotechnology and pharmaceuticals, spend much more time in the PTO than average. Second, the relative newness of some technologies or, relatedly, the recency of rules recognizing some technologies as patentable subject matter undoubtedly have an effect on pendency times in those fields. There surely must be a learning curve for inventors, patent attorneys, and patent examiners. It takes time for attorneys to develop and become comfortable with patent drafting techniques in new areas, and for examiners to understand these techniques. It takes time for the attorneys and the PTO to learn about the sources for nonpatent prior art in such areas, and for the PTO to make the sources available to examiners. Patent attorneys may encounter increased resistance from examiners not only because of uncertainty about prior art sources and claim drafting techniques, but also simply because of the vague feelings of unease that often accompany one's exposure to the unfamiliar. All of these factors are likely to create greater application backlogs in the PTO. It may also be that greater technical difficulty or the newness of the technology or the recency of its recognition as patentable subject matter tends to cause applicants in these areas to file more continuing applications. If so, we may find several causes for longer pendency times in some fields of inventive activity that sometimes work by themselves and sometimes interlink with other causes.

Finally, longer pendency times may also be due in part to applicants' perceptions of greater innovation value that lead them to devote greater resources to obtaining the strongest and broadest possible patents. First, we know that the filing of continuing applications accounts for a substantial portion of average pendency times, and that filing continuing applications adds substantially to patenting costs. Applicants perceiving greater value may tend to file more continuing applications. Second, we find evidence of sizable correlations between both claims and references on one hand, and pendency times on the other. Drafting and pursuing more claims costs more, and contributes quite significantly to the time that

239. Although patent examiners spend an average of about eighteen hours on a single application over a period of roughly two to three years, they take amounts of time ranging from eight to thirty-two in different technology areas. Lemley, *Rational Ignorance*, *supra* note 4, at 1500 & n.19.

applications spend in the PTO.²⁴⁰ Likewise, it costs more to carry out more thorough prior art searches, and larger numbers of prior art references, most of which originate with the applicant, correlate even more strongly with longer pendency times than the number of claims does.²⁴¹ These high

240. Finding a claims-time correlation confirms the intuitive. Examination focuses on the claims because they define the invention. It is the claims that must show a novel and nonobvious advance over the prior art. In both data sets, we found that the number of claims was highly correlated with pendency times from both the original priority filing date and from the filing of the most recent application.

Because these were large samples, and because there is quite a bit of “noise” in the data on years in the PTO, there does not have to be a huge correlation coefficient to be highly significant. By noise, we mean that many factors probably affect pendency times, some of which we cannot measure and some of which we may not even know about. Nevertheless, we discovered some rather large Pearson R values. The Pearson R is the correlation coefficient, a number that falls between 0 and 1, and if it is a positive number, it quantifies the extent to which increases in one variable correspond to increases in the other variable (and the extent to which decreases in one variable correspond to decreases in the other variable).

As with almost all of our other measures, we performed log transformations on the values to adjust for large right-hand skews caused by extreme upper outliers. For Internet business method patents, the results for claims and years in the PTO from the *original* filing date were: $R = 0.09$, $p = 0.003$. From the *most recent* application, the results were $R = 0.11$, $p = 0.0002$. For general patents, the results for claims and years in the PTO from the *original* filing date were: $R = 0.08$, $p = 0.01$. From the *most recent* application, the results were: $R = 0.11$, $p = 0.0004$.

241. In both data sets, the numbers of patent, nonpatent, and total prior art references were highly correlated with pendency times from the filing of the original application at the root of an application chain, nonpatent references being much more highly correlated than patent references.

Again, and for the same reasons as before, we performed log transformations on the values. For Internet business method patents, the results for references and years in the PTO from the *original* filing date were as follows. Patent References/Pendency: $R = 0.15$, $p < 0.0001$; Nonpatent References/Pendency: $R = 0.14$, $p < 0.0001$; Total References/Pendency: $R = 0.27$, $p < 0.0001$. For general patents, the results for references and years in the PTO from the *original* filing date were as follows. Patent References/Pendency: $R = 0.14$, $p < 0.0001$; Nonpatent References/Pendency: $R = 0.29$, $p < 0.0001$; Total References/Pendency: $R = 0.31$, $p < 0.0001$.

We obtained similar results when measuring years spent in the PTO from the applications leading immediately to their corresponding patents, with one major exception in the case of patent references. In the case of both Internet business method patents and general patents, the number of nonpatent references and total references were highly correlated with pendency times from the most recent application, but the number of references to prior patents was not significantly correlated. For Internet business method patents, the results for references and years in the PTO from the *most recent* filing date were as follows. Patent References/Pendency: No significant correlation; Nonpatent References/Pendency: $R = 0.10$, $p = 0.0005$; Total References/Pendency: $R = 0.16$, $p < 0.0001$. For general patents, the results for references and years in the PTO from the *most recent* filing date were as follows. Patent References/Pendency: No significant correlation;

positive correlations tell us nothing about the possible differences between Internet business method patents and general patents directly because the correlations were practically identical in the two data sets. They do seem to provide more support, however, for our self-selected value theory, particularly given that Internet business method patents did contain many more claims and references.²⁴² Third, when applicants perceive greater innovation value, they are likely to spend more time and money trying to overcome objections to claims by the patent examiner instead of giving in, again producing longer pendency times. As we see below, Internet business method patents spent much more time in the PTO than we would have expected.²⁴³

3. *Method of Comparison*

First, we measure pendency times from the filing of the original patent application. This measure is affected by the substantial portion of patents in both data sets in which one or more continuing applications were filed. In our set of 1,093 Internet business method patents, 196 (17.9%) were based on one or more ancestor applications prior to the filing of the most immediate application leading directly to the patent. In the set of 1,000 general patents, 293 (29.3%) had one or more ancestor applications. At least a substantial portion of this difference is surely attributable to the fact that Internet business method patents have not been around as long as many other types. Therefore, the set of Internet business method patents fails to capture as many patents having a history of multiple continuing applications.²⁴⁴ Also, a smaller portion of the original applications for

Nonpatent References/Pendency: $R = 0.21$, $p < 0.0001$; Total References/Pendency: $R = 0.17$, $p < 0.0001$.

242. Our additional finding that claims and references also are significantly correlated with each other provides yet more support for the self-selected value theory. This interaction, almost certainly the result of applicants' willingness to invest more in innovations they believe to be worth more, was clearly observable from a simple visual inspection of the data. We have correlation coefficients and p -values for these relationships as well, but we do not report them here. We do not wish to glaze over readers' eyes any more than we probably already have.

243. The fact that pendency times are highly correlated with numbers of claims and references means that, when we used multivariate statistical tests to account for such interdependencies, the difference in years in the PTO between Internet business method patents and general patents ceased to be significant. The correlations themselves, however, lend support to the idea of value as a self-fulfilling prophecy.

244. Although the general patent data set had earlier average filing dates than the Internet business method data set, the two sets were contemporaneous in issue dates and overall were sufficiently contemporaneous such that our comparisons of references and claims are not measurably affected by temporal differences.

Internet business method patents were filed before the time—mid-1995—when delaying the examination process started counting against the term of patent protection.²⁴⁵ Both observations would cause us to expect, *ceterus paribus*, shorter pendency times for Internet business method patents from the original filing than for other kinds of patents.²⁴⁶

Our second set of comparisons looks at pendency times counting only from the filing of the most immediate application leading directly to a given patent. This second set of comparisons thus removes the effect of continuing applications so that we may focus on other factors.

4. *Pendency Times from the Original Priority Filing Date:
Internet Business Method Patents Compared With General
Patents and With Patents in Each Technology Area*

The following table reports the untransformed means of pendency times for Internet business method patents, general patents, and patents in each technology area, as well as the results of statistical comparisons.

245. For reasons previously discussed, we doubt that the prospect of relatively small reductions in patent terms in the case of applications filed after mid-1995 caused many applicants to push through the examination process much faster. The main effect on pendency times caused by this change is the removal of “submarine patents” with exceptionally long pendencies. Thus, the fact that the Internet business method patent data set had proportionately more patents based on applications filed after this date caused us to expect shorter pendency times for them because the effects of submarine patents should have been much less (or perhaps nonexistent).

246. One might even speculate that the rapidity of change in the Internet business sector during the period we studied to have induced patent applicants to have created an incentive for applicants to speed up prosecution and get their patents issued faster.

**Table 9: Patent Pendency Times From Original Filing:
Internet Business Method Patents Compared with General Patents and with Patents
in Each Technology Area²⁴⁷**

	Yrs in PTO From Original Filing	Significance of Difference— Internet Business Methods Spend More, Less, or Equal Time in PTO
Internet Business Methods	2.67	
General Patents	2.77	Less
Acoustics	2.69	Equal
Automotive	2.19	More
Biotechnology	4.72	Less
Chemistry	3.42	Less
Communications	2.77	Equal
Computer-related	2.80	Equal
Electronics	2.38	More
Energy	2.78	Equal
Mechanics	2.29	More
Medical devices	2.78	Equal
Optics	2.85	Equal
Pharmaceutical	4.44	Less
Semiconductors	2.77	Equal
Software	3.07	Equal

Thus, Internet business method patents spent slightly less time in the PTO from the original priority filing date than general patents did (2.67

247. In the far right-hand column “more” means that Internet business method patents spent significantly more time in the PTO from the original filing date, “less” means that Internet business method patents spent significantly less time in the PTO from the original filing date, and “equal” means that there was no statistically significant difference in pendency times from the original filing date. As before, the statistical tests were done after performing log-transformations on the means to adjust for extreme upper outliers and other abnormal characteristics of the distributions. The results were confirmed by Wilcoxon nonparametric tests, with one exception we note below.

and 2.77 years, respectively). After adjusting the means to account for extreme values in the distributions, the average pendency time for general patents was longer than that for Internet business method patents, but the difference was barely significant. Moreover, the statistical significance of the difference disappeared when multivariate tests were used that accounted for the large correlations between numbers of claims and references on one hand, and pendency times on the other. We were not surprised that general patents spent a bit more time in the PTO than Internet business method patents, given the average earlier filing dates for general patents and the larger portion of general patents that were based on one or more ancestor applications. In fact, it is rather surprising that the difference between pendency times from the original filing was not substantially larger. We suspect that the newness and recency of Internet business method patents may largely explain this unexpectedly small difference. In other words, the newness and recency of Internet business method patents probably lengthened their average pendency times, while their average earlier filing dates and fewer continuing applications probably shortened their average pendency times. Thus, the effect of the former served as at least a partial offset to the effect of the latter.

In congruence with what we have found in the case of references and claims, we also think it possible that applicant perception of greater private value may have played a part in making pendency times for Internet business method patents longer than expected. Aside from the fact that perceptions of greater innovation value may have generated more references and claims, such perceptions also may have led applicants to contest examiner objections more strenuously rather than accepting them.

The comparison between Internet business method patents and those in individual technology areas reveals no statistically significant difference between Internet business method patents and those in eight of the fourteen technology areas. Internet business method patents spent significantly *more* time in the PTO than did patents in three areas: automotive-related, electronics, and mechanics. Internet business method patents spent significantly *less* time in the PTO than those in three areas: biotechnology, chemistry,²⁴⁸ and pharmaceuticals.²⁴⁹ Overall, then, there is

248. As with other comparisons, we not only used *t*-tests comparing log-transformed means to account for the effect of extreme upper outliers, but also Wilcoxon nonparametric tests. These tests were in agreement in practically every instance. In the case of chemistry patents from the original filing date, however, the tests produced contrary results. The *t*-test showed that chemistry patents spent significantly more time in the PTO than Internet business method patents, but the Wilcoxon test showed no significant difference. We chose to report the *t*-test result for several reasons, the most important of which is

no evidence that Internet business method patents were exceptional in the time they spent in the PTO from the original filing date, except that they had longer pendencies than we would have expected.

5. *Pendency Times from the Most Recent Filing Date: Internet Business Method Patents Compared With General Patents and With Patents in Each Technology Area*

If we then look at average pendency times counting only from the most immediate applications that led directly to the patents in question, we observe two interesting phenomena. First, both data sets are still characterized by upper outliers, but they are less extreme than they were when we measured pendency times from the original priority filing date, and thus are characterized by a less pronounced right-hand skew. This is entirely expected because we have removed the effect of multiple applications leading up to a substantial number of the patents in each set. Second, and very surprisingly, we found that Internet business method patents generally spent more time in the PTO than did other kinds of patents. Table 10 reports the outcomes of these comparisons in the same format as Table 9.

that it is the more conservative of the two in light of our overall assertion that Internet business method patents had much longer pendency times than expected. We do not want to be seen as making close judgment calls in favor of this assertion.

249. Despite the fact that the untransformed mean of software patent pendency times is quite a bit longer than that for Internet business method patents (3.07, 2.67), the actual difference was not significant after the appropriate adjustments for extreme values. Findings like this one increase in likelihood as distributions of data stray farther and farther from normality. The distribution of pendency times from the original filing for software patents was characterized by more extreme individual values, thus making the untransformed means less reflective of the true central tendency of this set of data. We found a lack of statistical significance not only when we used log-transformed means but also when we used a Wilcoxon nonparametric test. The number of software patents (ninety-two) was clearly a large enough sample for valid statistical testing, although it is of course true that this magnitude of difference probably would have been significant even after adjustment had the software sample size been far larger.

**Table 10: Patent Pendency Times From Most Recent Filing:
Internet Business Method Patents Compared with General Patents and with Patents
in Each Technology Area**

	Yrs in PTO From Most Recent Filing	Significance of Difference— Internet Business Methods Spend More, Less, or Equal Time in PTO
Internet Business Methods	2.39	
General Patents	2.02	More
Acoustics	2.10	More
Automotive-related	1.98	More
Biotechnology	2.60	Equal
Chemistry	2.13	More
Communications-related	2.17	More
Computer-related	2.06	More
Electronics	1.90	More
Energy-related	1.99	More
Mechanics	1.85	More
Medical devices	2.03	More
Optics	1.89	More
Pharmaceutical	2.64	Equal
Semiconductors	2.13	More
Software	2.30	Equal

The most important finding is that, when the effect of continuing applications is purged from our data, Internet business method patents spent much more time in the PTO than the average patent. This difference remained highly significant even after taking into account the strong

influence that the number of references and the number of claims each had on pendency times.²⁵⁰

Moreover, Internet business method patents spent significantly more time in the PTO from the most recent filing than patents in eleven of the fourteen individual technology areas, and an insignificantly different amount of time than patents in three areas: biotechnology, pharmaceuticals, and software. Patents in the first two of these three fields had significantly longer pendency times than Internet business method patents when counting from the original filing date, and we would have intuitively expected the same to be true when counting only from the most recent applications. They did not, however, probably because the proportion of their average pendency times from the original filing date that had resulted from continuing applications was larger than average within the set of general patents and certainly larger than for Internet business method patents. The fact that both types of patents continued to be on the high end of the set of general patents again substantiates the idea that patents are more important in these areas and applicants perceive them as having more value than in most areas. In the case of biotechnology patents, as with Internet business method patents, newness and recency could also have played a role.

When we measured from the filing of the most immediate application, the unadjusted averages for biotechnology and pharmaceutical patents certainly *looked as though* they should be significantly greater than those for Internet business method patents. However, the differences here are not statistically significant after appropriate adjustments. Even after removing the large effects of continuing applications in biotechnology and pharmaceuticals, average pendency times in these areas continued to be severely affected by a relatively small number of applications that took a very long time to issue as patents, rendering these unadjusted means a misleading indicator of the true center points in their respective distributions.²⁵¹

250. We performed log transformations on the means of the number of claims, number of prior art references, and years in the PTO, and then performed a multiple regression to determine whether Internet business method patents still spent significantly longer in the PTO from the most recent filing when accounting for the fact that numbers of references and claims heavily influenced pendency times. The *p*-value was < 0.0001.

251. Results from both *t*-tests using log-transformed means and Wilcoxon nonparametric tests confirmed the lack of significance in the differences between the means of pendency times in these two areas and those in Internet business method patents. The differences undoubtedly would have been significantly different had the sample sizes in these areas been as large as that for the set of general patents as a whole, but these samples were sufficiently large for valid statistical testing. To reiterate, the distributions of

After purging the effect of continuing applications from the pendency time data, thus focusing on the process of getting a particular patent issued directly from a particular application, the two possible explanations mentioned earlier appear to gain strength. The newness of the technology in Internet business method patents and the recency of the technology's legal recognition as patentable subject matter almost certainly played a part. The other factor—applicant perception of greater potential value leading to the commitment of greater resources to the patenting enterprise and thus increasing actual value—also remains as a probable partial explanation.

VI. SUGGESTED EXPLANATIONS FOR THE GAP BETWEEN CONVENTIONAL WISDOM AND REALITY

Most observers who have written or spoken publicly about Internet business method patents do not like them, to say the least. Our data, on the other hand, suggest that these patents were of a quality and value at least as high as most other patents. We were naturally led to wonder about the reasons for such a chasm.

A. Possible Explanations

One possible explanation is that Internet business method patents are indeed deficient in some ways, and that these deficiencies were a natural consequence of the patent system attempting to grapple with a new technology only recently recognized as patentable subject matter, and that these patents will get better with time. Although superficially appealing, this argument quickly implodes in the face of our data. We studied Internet business method patents when they were still new, compared them with patents in fields that were not so new, and could not find anything that was objectively wrong with them.

The approach used by the PTO and the courts for addressing the requirement of nonobviousness arguably could provide another explanation. Instead of looking at the invention from the perspective of subjective hindsight, PTO examiners and federal judges must use particularized prior art to demonstrate that an invention is undeserving of a patent because it represents an insufficient advance over what has gone before.²⁵² Although there is no evidence that patent examiners have

data points for pendency times in these two areas were still affected a lot by upper outliers even when measuring only from the most recent applications.

252. See, e.g., *Panduit Corp. v. Dennison Mfg. Co.*, 774 F.2d 1082 (Fed. Cir. 1985), vacated by 475 U.S. 809 (1986), remanded to 810 F.2d 1561 (Fed. Cir. 1987).

commonly used subjective hindsight to reject applications, there have been periods when the courts clearly did so. Gone are the days, however, when courts routinely invalidated patents on the ground that inventions did not reveal a “flash of creative genius.”²⁵³ The approach to the nonobviousness inquiry may explain why one can find many patents in all fields that seem to stand on very shaky ground, but it does not support an inference that patents in one area are more trivial than in other areas.²⁵⁴

One may contend, as many have, that a disproportionate share of Internet business method patents were obvious because many such methods were previously practiced but kept secret during the years in which most people thought that they were not patentable subject matter and that there was thus a failure to create prior art as occurs in most fields. The fact of the matter is, however, that businesses in almost all industries have for a long time strongly preferred trade secret protection to patents as a means for securing the best return on their research and development investments.²⁵⁵ Thus, there has always been a lot of “secret” prior art in all fields. Admittedly, the percentage of the total volume of prior art that was kept secret in the field of business methods was probably greater than the percentage of the total in many other fields because of the recency with which business methods were deemed eligible for patent protection. In the end, however, the data must speak for themselves. Internet business method patents cited far more prior art than other patents.

The possibility also exists that our analysis of the prior art cited in patents overemphasizes quantity measures and fails to adequately account for variations in the relevance and content quality of those references. We believe that we have shown, however, that the quantity measures we use have much to say about patent quality and value, and we also attempted a reasonably thorough assessment of the caliber of the nonpatent prior art cited in the patents we studied. Our quality assessment revealed nothing to indicate that the quality of nonpatent prior art in Internet business method patents was substantially inferior to that found in other patents.

253. *See, e.g.*, *Cuno Eng'g Corp. v. Automatic Devices Corp.*, 314 U.S. 84, 90-92 (1941). *See generally* 2 DONALD S. CHISUM, *CHISUM ON PATENTS: A TREATISE ON THE LAW OF PATENTABILITY, VALIDITY AND INFRINGEMENT* § 5.02 (1999).

254. Although no knowledgeable observer would wish a return to a paradigm of subjective hindsight for obviousness decisions, we personally believe that patent applicants rather than examiners should bear the burden of proof during patent prosecution. Analysis of such a fundamental overhaul of the patent examination system is beyond the scope of this Article.

255. COHEN ET AL., *WHY U.S. MANUFACTURING FIRMS PATENT*, *supra* note 51; Hall & Ziedonis, *supra* note 186; Levin et al., *supra* note 51.

B. An Information Cascade of Negative Opinion

We are still left with questions about the gap between the accepted wisdom and what we believe to have been the reality. We believe that negative opinions about business method patents swelled to a classic information cascade. Part of a large body of theoretical and empirical work on collective behavior and an outgrowth of game theory, research on information cascades confirms the fundamental notion that people react to their environment, and that this environment includes other people's reactions to *their* environments.²⁵⁶ Information can cascade to the point (the "tipping point") that people's actions are increasingly influenced by the information conveyed through the actions of others, regardless of the accuracy of that information.²⁵⁷ This is sometimes called the "bandwagon effect," as seen in fads and fashions in apparel, toys, and academic research paradigms, and also accounts for phenomena such as herd behavior in the stock market, riots in the streets, and graffiti on the walls. It is possible for the cascade to reach a point that some people base their

256. See generally Lisa R. Anderson & Charles A. Holt, *Information Cascades in the Laboratory*, 87 AM. ECON. REV. 847 (1997) (reporting that information cascades develop consistently in behavioral laboratory experimentation); Abhijit V. Banerjee, *A Simple Model of Herd Behavior*, 107 Q.J. ECON. 797 (1992) (modeling how subsequent actors increasingly rely on information signaled by the decisions of others to act, producing information cascades); B. Douglas Bernheim, *A Theory of Conformity*, 102 J. POL. ECON. 841 (1994) (explaining how thought and behavior tend to converge and form social norms); Sushil Bikhchandani et al., *Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades*, 12 J. ECON. PERSP. 151 (1998) (observing the critical role of "fashion leaders" in information cascades and that such cascades often fade because of shocks—sometimes slight ones—such as the arrival of better informed individuals or public release of new information); Bikhchandani et al., *Theory of Fads*, *supra* note 52 (showing how information cascades can often be mistaken even though individuals are acting rationally, how a few early individuals can have a disproportionate effect on the development of a cascade, and how people's beliefs and actions can depend to an increasing degree on information learned from the decisions of others); Stephan Fuchs, *A Sociological Theory of Scientific Change*, 71 SOCIAL FORCES 933 (1993) (describing movement from one area of inquiry in a scientific discipline to another area within that discipline in a way that closely resembles models of innovation diffusion by the way in which "fashion leaders" create conditions that ultimately lead to herding); Yali Peng, *Intellectual Fads in Political Science: The Case of Political Socialization and Community Power Studies*, 27 PS: POL. SCI. & POLS. 100 (1994) (using two fields of study within political science to illustrate the development and decline of intellectual fads in academic circles); Cass R. Sunstein, *On Academic Fads and Fashions*, 99 MICH. L. REV. 1251 (2001) (applying the teachings of information cascade theory to fads and fashions in academic legal theories and models). For a popular treatment, see MALCOLM GLADWELL, *THE TIPPING POINT* (2000), discussing what he calls "social epidemics," or abrupt changes in social phenomena resulting from cascading information.

257. Banerjee, *supra* note 256, at 799.

actions almost exclusively on information received in this manner rather than information they have discovered for themselves.

A combination of factors contributes to information cascades. The novelty of the phenomenon about which information is being transmitted is probably the most important. Internet business method patents were new and very different looking to almost everyone. Some of them struck informed observers as palpably ridiculous.²⁵⁸ The identity of the actors and the means by which they act and speak are also important. More than a few of those voicing adverse opinions about Internet business method patents were luminaries, whose opinions are highly regarded by others. On some occasions they spoke through prominent outlets. Then at some point the sheer volume of information caused the cascade to develop a life of its own, which is manifested by the decisions to produce even more negative commentary and reportage and to single out these patents for different treatment in Congress and in the PTO.

Many of the commentators, especially the academics, were not merely intellectual gadflies but were exceptionally capable people of keen discernment and good judgment. Although they are almost certainly not as prone to jumping on bandwagons as are the less capable, they are not exempt from long-evolved psychological tendencies.

A related theory may help further explain the persistence of the perception-reality gap. Aided by observers' repetition of a few striking examples such as Amazon.com's decision to sue Barnesandnoble.com for infringement of the patent on one-click Internet shopping check-out method,²⁵⁹ many people developed firm beliefs about these patents, namely, that they were bad. Although some people develop beliefs only after seeing substantial evidence, many others form them much more easily. Research in behavioral decision theory shows that, once people develop relatively firm beliefs, they tend to cling to them. Most people tend to then pay greater attention to further information confirming these beliefs, and even to seek out confirming evidence while paying little attention to contrary evidence. This "confirmation bias," sometimes called

258. Some probably were rather silly, because one can find many patents in all areas that seem silly. *See, e.g.*, U.S. Patent No. 6,311,639 (issued Nov. 6, 2001) ("Dog Bone Holder"). Others of this ilk are legion. Most, of course, are never asserted against anyone. We admit that a few obvious-looking Internet business method patents were asserted, but we find no reason to believe that, over time, such patents in other areas are asserted less frequently in infringement litigation than Internet business method patents.

259. We have one colleague who said that all of this is explained by two words: "one" and "click". There is an element of truth in this observation, although we think that it is overly simplistic. Besides, we could not get a long article out of that observation.

“belief perseverance,” is a widespread psychological phenomenon even among scientists and others who have been trained to do otherwise.²⁶⁰ This is not to say that beliefs never change, of course. These are simply strong tendencies.

VII. CONCLUSION

We have found no evidence to support the idea that Internet business method patents were of inferior quality or of lower value than most other patents. They were not exceptional. There is even a possibility that the opposite was true. The idea that Internet business method patents never should have been made eligible for patent protection in the first place, or that they are in fact inferior to others, led the PTO and Congress to treat them differently. We do not believe that there were any good reasons for singling them out, but we also see a much more fundamental problem. As Judge Newman observed in *In re Shrader*,²⁶¹ the concept of “business methods” is probably undefinable. Why, for example, would a method used for applying a photo-resistant layer to a silicon wafer in semiconductor chip manufacturing not be a method of conducting the

260. Such an approach is not the logical way to test a belief (or a hypothesis in statistical testing) and is diametrically opposed to sound scientific method that seeks falsification of hypotheses. For discussions of the confirmation bias, see, for example, HOGARTH, *supra* note 53, at 116-21; RICHARD NISBETT & LEE ROSS, HUMAN INFERENCE: STRATEGIES AND SHORTCOMINGS IN SOCIAL JUDGMENT 167-92 (1980); SCOTT PLOUS, THE PSYCHOLOGY OF JUDGMENT AND DECISION MAKING 231-40 (1993); Lee Ross & Craig A. Anderson, *Shortcomings in the Attribution Process: On the Origins and Maintenance of Erroneous Social Assessments*, in JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 129, 144-52 (Daniel Kahneman et al. eds, 1982). See also MASSIMO PIATELLI-PALMARINI, INEVITABLE ILLUSIONS: HOW MISTAKES OF REASON RULE OUR MINDS 120-23 (1994) (using different terminology but discussing the same subject matter).

261. 22 F.3d 290, 297-98 (Fed. Cir. 1994) (Newman, J., dissenting). In a dissenting opinion in which she responded to a request by the PTO’s Board of Patent Appeals and Interferences for guidance from the Federal Circuit on whether business methods should be excluded from patentability, a request the majority declined to take up, Judge Newman stated:

The Board also relied on the “method of doing business” ground for finding Schrader’s subject matter non-statutory under section 101. In so doing the Board remarked that the “method of doing business” is a “fuzzy” concept, observed the inconclusiveness of precedent, and sought guidance from this court. Indeed it is fuzzy; and since it is also an unwarranted encumbrance to the definition of statutory subject matter in section 101 . . . , my guidance is that it be discarded as error-prone, redundant, and obsolete. It merits retirement from the glossary of section 101.

Id. (Newman, J., dissenting) (footnote omitted).

business of making computer chips? Although the enormous difficulty of defining an area of technology for the purpose of treating it differently at the patent system entry point is not limited to business methods, it could be especially acute in this case because all patents on any kind of method or process are presumably acquired for use in a business endeavor.

While it is true that legal decision makers often must work with definitions, it is also true that the greater the premium placed on how an activity is defined at the outset, the more skill and resources will be expended by affected parties to tailor descriptions of what they are doing to either opt into or out of that definition. When a technology is defined for the purpose of differential treatment in patenting, patent attorneys know this in advance of drafting the patent application. Skillful, experienced patent attorneys can often find ways to draft patent applications to make their clients' inventions fall within or outside a defined category. This is amply illustrated by the history of software patents, in which patent lawyers made software sound like a machine with nuts and bolts at a time when courts were still struggling with the idea of software as patentable subject matter and it was necessary that the functional ideas embodied in software be represented as accomplishing a transformation in the traditional physical world.²⁶²

The problem of definition is further illustrated by the PTO's March 2000 initiative providing an additional level of scrutiny to some business method patents. The program applies only to applications for patents in classification 705 and, indeed, only to a portion of those.²⁶³ As we noted earlier in this Article, we found many relevant patents in other classes—mainly in 707 and 709—that looked very much like the ones in 705. The PTO often places patents into more than one classification, and some of the patents we found were in two or more of these three classifications. However, many do escape the confines of category 705 and are found only in 707 or 709, although they appear to be business method patents.²⁶⁴

262. See, e.g., *In re Alappat*, 33 F.3d 1526 (Fed. Cir. 1994) (en banc).

263. Those patents placed in Class 705 and one or more other classifications are subjected to additional scrutiny only if 705 is the first-listed, "main" classification.

264. U.S. Patent No. 5,659,742 (issued Aug. 19, 1997) ("Method for storing multimedia Information in an Information Retrieval System") (describing and claiming an Internet-based system for retrieving multi-media data in categories such as sports or movies and storing demographic data about users); U.S. Patent No. 5,617,565 (issued Apr. 1, 1997) ("Broadcast Interactive Multimedia System") (describing and claiming an Internet-based system for interactive television). Although both of these patents obviously cover Internet business methods, they were not placed in the classification designed for business method patents, Class 705. Instead, the PTO placed them in Class 707 for inventions pertaining to data collection and retrieval. The patents were not misclassified and do in-

Insurmountable definitional problems therefore reduce the effectiveness of any attempt to provide stricter scrutiny for business method patents, or indeed for almost any other type of subject matter.

The PTO's differential treatment of business method patent applications is likewise misguided as a matter of sound patent policy because transaction costs are likely to be increased by the definitional gerrymandering of patent lawyers. The PTO's action probably did serve an end that is worthwhile to the agency itself—easing public and congressional pressure from those who thought that business method patents were uniquely a problem—but it does not seem to serve broader policy ends. The one positive result of the PTO's experiment is that the experience gained in the effort could prove valuable were the agency ever given the resources and motivation to enhance scrutiny of patent applications in all fields of technology.

When we say that Internet business method patents were of a quality and value at least as high as most others, and arguably even higher, we do not mean that these patents were a godsend. We do not know whether they are likely to contribute to social welfare over the long term. We do claim, however, that efforts to treat them differently upon entry to the patent system will be ineffective at best and counterproductive at worst.²⁶⁵ When

deed belong in 707. *See also* U.S. Patent No. 6,009,458 (issued Dec. 28, 1999) (“Networked Computer Game System with Persistent Playing Objects”) (describing and claiming an Internet-based system for making network game-playing available to users); U.S. Patent No. 6,009,457 (issued Dec. 28, 1999) (“Distributed Real-Time Communications System”) (describing and claiming an Internet-based system for offering virtual jam sessions by musicians). Again, these two patents clearly involve Internet business methods, but were not placed in Class 705. Both were placed, properly so, in Class 709, the subject of which is computer communication networks, and the second one was also properly assigned to 707. These represent additional examples of business method patents that were properly classified, but not in 705.

265. This statement should not be read as suggesting that everything the PTO and the courts do should or can be the same for all patents on all occasions. Whether a patent should be granted and upheld depends mainly on whether the invention represents a novel and nonobvious advance over the prior art in the field. Questions such as which prior art should be considered, whether an invention is a nonobvious advance over the prior art, and whether the patent discloses enough information are all viewed through the lens of an ordinarily skilled practitioner in the particular art. When one realizes that an ordinarily skilled practitioner may range from an experienced mechanic or electrician to a person with a Ph.D. and much experience in molecular biology or computer science, the conclusion is inescapable that not all rules can be applied exactly the same in every case. This truism is very different, however, from saddling different types of subject matter with different rules before they enter the system. The former represents necessary flexibility in the application of rules while the latter invites manipulation of the rules themselves. *See, e.g.,* Dan L. Burk & Mark A. Lemley, *Is Patent Law Technology-Specific?*,

we express the opinion that the PTO's initiative to examine them more thoroughly was misguided, we fully understand that the agency may have done so under duress. If business method patents were indeed of low quality and value, the conclusion is inescapable that a great many patents of other kinds are also questionable. We make no attempt here, however, to analyze all of the likely frailties in the patent system as a whole, because others have done this and will continue to do so.²⁶⁶

17 BERKELEY TECH. L.J. 1155, 1185-90 (2002) (exploring the idea that use of the ordinarily skilled person standard necessarily results in differences in the application of some of the basic patentability rules).

266. And besides, we've been working on this project a very long time, and we're tired.

ENDING PATENT LAW'S WILLFULNESS GAME

By Mark A. Lemley[†] and Ragesh K. Tangri[‡]

ABSTRACT

Patent law's doctrine of willful infringement turns the attorney-client privilege on its head by requiring companies confronted with a patent first to obtain a written opinion of counsel and then to disclose that opinion in court. The willful infringement game is a costly one with perverse consequences for patent law. The complex rules of the game cause attorneys to withhold information and candid advice from their clients because they know the advice may later have to be disclosed to the other side. The current rules also infect pre-litigation advice, essentially making it impossible for a competent lawyer to advise a client that a competitor's patent should be avoided. Perhaps worst of all, they discourage engineers and companies from reading patents in the first place, thereby undermining the disclosure function that is at the foundation of the patent system.

This Article proposes that the best approach to reducing these perverse effects is first to redefine willful infringement as adopting a technology with knowledge that it was derived from the patentee, and second to adjust the premium charged for it. Changing the focus of the inquiry to the time of adoption would reduce the number of accused infringers that would be required to seek an opinion of counsel and minimize the scope of the disclosure for the rest. Adjusting the damages premium by limiting willful infringement damages awards to attorneys fees in the bulk of cases will ensure that the willfulness doctrine fulfills its purpose of compensating the patentee while not over-detering competitors in the marketplace. Narrowing the willfulness doctrine and limiting the amount of damages awarded are consistent with the purposes of patent law and will put an end to a longstanding and dangerous game.

© 2003 Mark A. Lemley and Ragesh K. Tangri

[†] Elizabeth Josslyn Boalt Chair in Law, University of California at Berkeley (Boalt Hall); of counsel, Kecker & Van Nest LLP.

[‡] Partner, Kecker & Van Nest LLP. The views expressed in this Article are our own, not those of our clients or colleagues.

We thank Chris Cotropia, Daralyn Durie, Rose Hagan, Scott Kieff, Doug Lichtman, Chris Martiniak, Joe Miller, Lee Van Pelt, Elizabeth Rader, and participants in the Advanced Patent Law Institutes at the Berkeley Center for Law and Technology and the University of Texas School of Law for their helpful suggestions.

TABLE OF CONTENTS

I.	INTRODUCTION	1086
II.	THE WILLFULNESS GAME.....	1089
III.	THE SCOPE OF PRIVILEGE WAIVER.....	1094
	A. The Waiver Can Vary by Time	1094
	B. The Waiver Can Vary by Subject Matter	1096
	1. <i>Attorney-Client Versus Work Product Privilege</i>	1096
	2. <i>Opinion Counsel Versus Trial Counsel</i>	1098
	C. Implications of the Various Rules	1099
IV.	PROBLEMS WITH THE WILLFULNESS GAME	1099
	A. Undermining Patent Disclosures	1100
	B. A Trap for the Unwary	1102
	C. Polluting Pre-Litigation Advice	1103
	D. The Choice of Counsel	1105
	E. Polluting Litigation Advice and Settlements.....	1106
V.	WHAT IS TO BE DONE.....	1109
	A. Abolishing Willfulness.....	1109
	B. Eliminating the Requirement to Waive Privilege.....	1113
	C. An Objective Standard	1115
	D. Redefining Willfulness.....	1116
	E. Adjusting the Willfulness Premium	1122
VI.	CONCLUSION	1124

I. INTRODUCTION

The attorney-client privilege lies at the heart of the legal system. Attorneys can give honest advice to their clients because they know that this advice will remain confidential. Attorneys cannot breach the privilege, even in extreme circumstances. They have an ethical obligation to maintain client confidences.¹ Clients, in turn, can feel comfortable telling their lawyers the truth because they know that this information will remain confidential. Without the attorney-client privilege, the adversary system would be unrecognizable.²

1. The Model Rules of Professional Responsibility impose upon lawyers a duty of confidentiality that can be breached only in a few extreme cases. MODEL CODE OF PROF'L RESPONSIBILITY DR 4-101 (1981); MODEL RULES OF PROF'L CONDUCT R. 1.6 (Discussion Draft 2002).

2. See *Upjohn Co. v. United States*, 449 U.S. 383, 389 (1981) (discussing the values served by the attorney-client privilege); *Quantum Corp. v. Tandon Corp.*, 940 F.2d 642, 644 (Fed. Cir. 1991) (calling the attorney-client privilege a "basic, time-honored privilege [warranting] careful consideration"). On the role of attorney-client privilege, see generally CHARLES W. WOLFRAM, *MODERN LEGAL ETHICS* 250-96 (1986). Cf. Stephen McG. Bundy & Einer Richard Elhauge, *Do Lawyers Improve the Adversary System? A General Theory of Litigation Advice and its Regulation*, 79 CALIF. L. REV. 313 (1991) (discussing the ways in which the attorney-client privilege distorts the litigation process).

Patent law turns the attorney-client privilege on its head. Patent law punishes willful infringers by subjecting them to treble damages.³ An odd set of legal rules stemming from patent law's effort to determine what constitutes willful infringement effectively requires companies confronted with a patent first to obtain a written opinion of counsel and then to disclose that opinion in court. To do that, the accused infringer will have to waive its attorney-client privilege in virtually every case. Even worse, the law puts the question whether an accused infringer will have to waive privilege in the hands of the patent owner, who can send a carefully crafted letter putting a potential defendant on notice of the patent. A patent owner thus triggers the obligation to obtain a written opinion of counsel without actually threatening to sue anyone. In turn, accused infringers who are aware of these rules respond to such letters by obtaining a sort of pseudo-legal advice that both they and their attorneys recognize to be a construct. Both plaintiffs and defendants are playing a costly game. Part II explains the legal rules leading to the game.

If the game progresses to the litigation stage and the defendant proffers the opinion it obtained as a defense to alleged willful infringement, the court must confront questions of privilege waiver. All courts to have considered the issue have required defendants accused of willful infringement to waive privilege to some degree. Courts have split widely on the breadth of that waiver, however. Some have created a limited waiver, focusing on the advice the accused infringer received at the time it began the allegedly infringing activity. Others have required a broader waiver of privilege, extending to all subject matter discussed in the attorney's opinion. Still other courts have held that the waiver extends to litigation counsel, who presumably must turn over their communications relating to trial preparation and strategy. Finally, a significant number of courts have held that the waiver of privilege extends not only to attorney-client communications, but also to attorney work product, even if not communicated to the client. Part III explores these cases in detail.

Part IV explains the perverse consequences of the broad waiver of privilege in patent law. Some of these consequences affect litigation—lawyers and clients who know that the lawyer's advice will be disclosed to the other side will behave differently, withholding information and candid advice from each other. But other consequences extend beyond litigation. They infect pre-litigation advice, essentially making it impossible for a

3. 35 U.S.C. § 284 (2000). The statute does not specifically refer to trebling for willfulness, but courts have found willfulness sufficient to qualify for treble damages. *See, e.g., Jurgens v. CBK, Ltd.*, 80 F.3d 1566, 1570 (Fed. Cir. 1996).

competent lawyer to advise a client that a competitor's patent should be avoided. The rules of the willful infringement game set traps for the unwary, who may not realize the consequences of failing to obtain the necessary written opinion of counsel. They interfere with a client's ability to choose counsel. And they discourage engineers and companies from reading a competitor's patents at all, thereby undermining the disclosure function that is at the foundation of the patent system.

Part V offers solutions. One possibility is to abolish the willfulness rule entirely. We ultimately reject this approach because we worry that ordinary patent damages alone would be insufficient to deter infringement optimally in many cases. Another possibility is to abolish the rule that requires disclosure of opinions of counsel. While a good idea, this option would not solve the problems created by the willfulness game, because many defendants will still need to rely on the opinion of counsel in order to disprove willfulness.

Instead, we think the better approach to willful infringement is first to redefine it as *adopting* a technology with knowledge that it was derived from the patentee, and second to adjust the premium charged for it. Many of the problems with the willfulness rules stem from the fact that willfulness is an ongoing inquiry. The ongoing nature of the inquiry adversely affects a defendant that develops or adopts a technology in good faith but later learns it is infringing a patent. Changing the focus of the inquiry to the time of adoption is consistent with the ordinary understanding of willfulness outside of patent law and would help end the willfulness game. An independent developer could never be a willful infringer, and thus would neither need to obtain nor disclose in court a written opinion of counsel merely because it later learned of a patent. By contrast, an accused infringer would need advice of counsel if it was aware of a patent and affirmatively sought to design around that patent. Such an accused infringer therefore would have to waive privilege. But since only the accused infringer's intent *at the time of adoption* would matter, the scope of privilege waiver would be limited to communications at the time of adoption and would not infect the advice given by litigation counsel.

Finally, Part V also analyzes the economics underlying the amount of the damage premium imposed based on a finding of willfulness. As a result of that analysis, we suggest changing that premium by limiting it to awarding attorneys fees to successful plaintiffs.

II. THE WILLFULNESS GAME

Increased punishment for willful violations is common to most intellectual property regimes. Unlike copyright and trade secret law, patent law punishes not only those who copy from the intellectual property owner but also those who develop the same invention independently.⁴ However, like other forms of intellectual property, patent law treats those who intentionally infringe a right more harshly than those who act innocently.⁵ In patent law, this harsher treatment takes the form of judicial discretion to enhance damages and award attorney fees if the court determines a defendant's infringement to be willful.⁶

Identifying willful infringers naturally requires the law to define willfulness. Patent law follows the rest of intellectual property⁷ by defining willful conduct to require awareness that the act was illegal.⁸ Requiring

4. See, e.g., ROBERT P. MERGES ET AL., *INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE* 24 (3d ed. 2003).

5. Copyright law enhances statutory damages for willful infringement. 17 U.S.C. § 504(c)(2) (2000). Trade secret and trademark law permit trebling of damages for willful misappropriation. See, e.g., 15 U.S.C. § 1117 (2000) (specifying damages for willful trademark infringement); UNIF. TRADE SECRET ACT § 3(c) (2000) (specifying damages for willful trade secret infringement). Copyright, trademark, and trade secret law, unlike patent law, also include criminal penalties for willful infringement. See 17 U.S.C. § 506 (setting forth federal criminal copyright infringement penalties); 15 U.S.C. §§ 1114, 1116-18, 1120, 1122, 1125 (setting forth federal trademark infringement penalties); 18 U.S.C. §§ 1831-39 (2000) (setting forth federal criminal trade secret infringement penalties); Cal. Penal Code § 499c (West 1999) (setting forth California criminal trade secret infringement penalties).

6. 35 U.S.C. § 284 (permitting enhancement of damages for willful patent infringement); *id.* § 285 (permitting attorney fees for willful patent infringement). Courts have emphasized that an award of enhanced damages and attorney fees is discretionary, not mandatory, when willful infringement is found. *Cybor Corp. v. FAS Techs.*, 138 F.3d 1448, 1461 (Fed. Cir. 1998) (en banc). Indeed, one recent study found that only about a third of willfulness cases involved enhanced damages, and that that number was declining over time. Jennifer L. Knabb & Michael J. Jeffords, *Trends in Patent Infringement Damages*, 21 IPL NEWSL. 22, 24 (2003).

7. See Dane S. Ciolino & Erin A. Donelan, *Questioning Strict Liability in Copyright*, 54 RUTGERS L. REV. 351, 406-07 (2002); R. Anthony Reese, *Historical Development of Mental State Considerations in Copyright Infringement* (Nov. 24, 2003) (unpublished manuscript, on file with author) (observing that copyright law historically defined willfulness to require not only intent to act but awareness that the act was illegal).

8. Contributory infringement, like willful infringement, imposes a knowledge requirement. See, e.g., *Aro Mfg. Co. v. Convertible Top Replacement Co.*, 377 U.S. 476, 488 (1964) (requiring a showing that the infringer knew the product it made "was both patented and infringing" to satisfy the knowledge requirement in the context of contributory infringement); see also *Ryco, Inc. v. Ag-Bag Corp.*, 857 F.2d 1418, 1428 (Fed. Cir. 1988) (defining the test for willfulness in the context of patent law as whether a defen-

awareness of illegality departs from the willfulness rule in other kinds of tort cases⁹ in which intentional action can be willful even if the actor did not know that it was illegal. A defendant must be aware the patent exists and understand that the patent is both valid and infringed in order to be a willful infringer.

Defendants may become aware of a patent in several ways. First, an accused infringer may discover a patent on its own. A corporation can find a relevant patent during a patentability search for a new invention; or if one of its engineers regularly reads new patents in her field, observes a patent number marked on a competitor's product,¹⁰ or even peruses a news article. Second, a company may learn of a patent by having the Patent and Trademark Office cite it as prior art while the company is prosecuting its own patent application. Finally, a patent owner can put a competitor on notice of the existence of a patent by sending the competitor a letter identifying the patent and giving some indication of how it relates to the competitor's business.¹¹

Notice of the existence of a patent does not in itself suffice to establish willfulness, however. To be willful, an accused infringer must understand

dant could have "any confidence that a court might hold the patent invalid or not infringing"). *But see* Sandisk Corp. v. Lexar Media, Inc., 91 F. Supp. 2d 1327, 1335 (N.D. Cal. 2000) (suggesting, incorrectly in our view, that knowledge of the patent and of the infringing activity will suffice for contributory infringement liability).

9. The law outside of intellectual property generally treats conduct as intentional if it was done deliberately, even if the actor is not aware that the conduct is wrongful. It is a fundamental maxim of criminal law, for example, that a mistake of law does not excuse. *See* Stuart P. Green, *Plagiarism, Norms, and the Limits of Theft Law: Some Observations on the Use of Criminal Sanctions in Enforcing Intellectual Property Rights*, 54 HASTINGS L.J. 167, 185 n.67 (2002); Douglas Husak & Andrew von Hirsch, *Culpability and Mistake of Law*, in ACTION AND VALUE IN CRIMINAL LAW 157-74 (Stephen Shute et al. eds., 1993); Dan M. Kahan, *Ignorance of the Law Is an Excuse—But Only for the Virtuous*, 96 MICH. L. REV. 128, 135 (1997).

10. On the role of marking in patent law, see Roger D. Blair & Thomas F. Cotter, *Strict Liability and Its Alternatives in Patent Law*, 17 BERKELEY TECH. L.J. 799 (2002).

11. These letters are normally quite circumspect, because a letter that actually comes out and says that the competitor is infringing the patent will give the recipient a reasonable apprehension that they will be sued, and therefore permit the recipient to sue for declaratory judgment first. *See, e.g.,* Scholle Corp. v. Blackhawk Molding Co., 133 F.3d 1469, 1472 (Fed. Cir. 1998) (discussing what content in a threat letter will justify a declaratory judgment action). An accused infringer who sues first gets to choose a more favorable forum and go first in court. Empirical evidence has suggested that declaratory judgment patent plaintiffs are 14% more likely to win than similarly situated patent defendants. Kimberly A. Moore, *Forum Shopping in Patent Cases: Does Geographic Choice Affect Innovation?*, 79 N.C. L. REV. 889, 920-21 (2001).

that the patent is valid and that what it is doing infringes the patent.¹² Determining what a corporation thinks, particularly about questions as difficult and abstruse as patent validity and infringement, is a thorny evidentiary problem. In patent law, the Federal Circuit has effectively hinged the question of intent on whether the accused infringer obtained and believed a competent opinion of counsel. Failure to obtain such an opinion creates a presumption of willfulness,¹³ and failure to disclose an opinion in court after it was obtained creates a presumption that the opinion was unfavorable.¹⁴ While both presumptions are rebuttable, in practice they are likely

12. See, e.g., *Ryco*, 857 F.2d at 1428.

13. *Underwater Devices, Inc. v. Morrison-Knudsen Co.*, 717 F.2d 1380, 1389-90 (Fed. Cir. 1983) (creating a duty to obtain an opinion of counsel when confronted with a patent). While the Federal Circuit decision in *Read Corp. v. Portec, Inc.*, 970 F.2d 816 (Fed. Cir. 1992), observed that "the absence of such advice [of counsel] does not mandate a finding of willfulness," the Federal Circuit has made it quite clear that a potential infringer with actual notice of a patent has an obligation to take "due care" to avoid infringement. 970 F.2d at 828; see also *Crystal Semiconductor Corp. v. TriTech Microelectronics Int'l, Inc.*, 246 F.3d 1336, 1351 (Fed. Cir. 2001). To discharge this obligation, the accused infringer "usually" must obtain a written opinion of counsel in order to avoid a finding of willfulness. *Comark Communications, Inc. v. Harris Corp.*, 156 F.3d 1182, 1190 (Fed. Cir. 1998); *Electro Med. Sys. v. Cooper Life Scis., Inc.*, 34 F.3d 1048, 1056 (Fed. Cir. 1994); *Spindelfabrik Suessen-Schurr v. Schubert & Salzer Maschinenfabrik AG*, 829 F.2d 1075, 1084 (Fed. Cir. 1987) (stating that an opinion of counsel is "normally" required). And in *Smith Engineering Co. v. Eisenmann Corp.*, No. 01-1202, 2002 U.S. App. LEXIS 905 (Fed. Cir. Jan. 17, 2002), the Federal Circuit dismissed the opinion of defendant's engineering department as insufficient in an unpublished disposition.

By contrast, the cases in which the court excuses a failure to obtain an opinion of counsel usually involve a good reason why the accused infringer did not obtain such an opinion. In *State Contracting & Engineering v. Condotte America*, 346 F.3d 1057 (Fed. Cir. 2003), for example, the court refused to draw an adverse inference from failure to obtain an opinion of counsel where the defendant, a state contractor, relied upon the state's specific representation that it had a license to use the patented technology.

14. See *Eastman Kodak Co. v. Goodyear Tire & Rubber Co.*, 114 F.3d 1547, 1561 (Fed. Cir. 1997) ("[F]ailure to introduce an exculpatory opinion of counsel at trial may support an inference that such an opinion was either never obtained or was adverse."); see also *Advanced Cardiovascular Sys. v. Medtronic, Inc.*, 265 F.3d 1294, 1309 (Fed. Cir. 2001) ("A party asserting its attorney-client privilege [with respect to its legal consultations] runs the risk of having the fact-finder draw a negative inference."); *L.A. Gear, Inc. v. Thom McAn Shoe Co.*, 988 F.2d 1117, 1126 (Fed. Cir. 1993) ("We have held that the assertion of privilege with respect to infringement and validity opinions of counsel may support the drawing of adverse inferences."). That presumption can be rebutted in extraordinary circumstances. In *Wesley Jessen Corp. v. Bausch & Lomb, Inc.*, 209 F. Supp. 2d 348 (D. Del. 2002), the court credited the company's "reasonable" belief that it had successfully designed around the patent even though it refused to waive privilege and produce an opinion of counsel. 209 F. Supp. 2d at 391. Notwithstanding cases such as *Wesley Jessen*, the adverse inference is a major roadblock to defendants seeking to prove their good faith.

to have a strong impact on a jury that has just concluded that the patent is valid and the defendant an infringer. The court has also created a series of rules for determining whether an opinion is competent, the most important of which is that the opinion must be in writing.¹⁵

Because of these presumptions, any company that receives a letter putting it on notice of the existence of a patent, and anyone that runs across a potentially relevant patent, has an obligation either to stop using the invention or to obtain a competent written opinion of counsel explaining why the patent is invalid or not infringed. These opinions are expensive. Estimates we have heard put the cost at between \$20,000 and \$100,000 per patent, and notices of infringement often include more than one patent.¹⁶ Obtaining opinion letters requires some experience with the patent system. A company that doesn't know the legal rules—doesn't know how to play the game—ignores a threat of infringement at its peril, even if it believes in good faith that the accusation of infringement is frivolous. Worse still, the uninitiated might get a traditional form of legal advice: an honest opinion that acknowledges a risk of infringement. The accused infringer that does this will almost certainly be on the hook for willful infringement if it continues to make, use, or sell the product. A company that knows how to play the game, by contrast, pays its money and requests only a favorable written opinion. Lawyers will generally be able to come up with plausible arguments that the patent is invalid or not infringed.¹⁷ As long as the client

15. See, e.g., *Comark*, 156 F.3d at 1191-92 (also discussing competence of legal opinion); *Johns Hopkins Univ. v. Cellpro, Inc.*, 152 F.3d 1342, 1364 (Fed. Cir. 1998) (discussing competence of opinion in detail).

16. Matthew Powers and Steven Carlson independently reported the same estimates. Matthew D. Powers & Steven C. Carlson, *The Evolution and Impact of the Doctrine of Willful Patent Infringement*, 51 SYRACUSE L. REV. 53, 102 (2001).

17. There are almost always nonfrivolous arguments of this sort, and opinion counsel will make them because, at the end of the day, opinion counsel are lawyers working for a client and patent issues tend to be issues about which reasonable minds can differ. While the standard for what level of certainty an opinion writer must possess is not crystal clear from the Federal Circuit's pronouncements on the subject, it is obviously not a "you-must-believe-in-your-argument-beyond-a-reasonable-doubt-before-giving-an-opinion-to-the-client" standard. In fact, at least some cases suggest that it approaches a Rule 11 standard. See, e.g., *Ryco*, 857 F.2d at 1428. Patent law is also complex, both factually and legally: over 30% of district court claim constructions are reversed by the Federal Circuit, to pick just one example. Christian A. Chu, *Empirical Analysis of the Federal Circuit's Claim Construction Trends*, 16 BERKELEY TECH. L.J. 1075, 1104 (2001) (giving numbers between 30% and 38%, depending on the measure chosen); see also Kimberly A. Moore, *Are District Court Judges Equipped to Resolve Patent Cases?*, 15 HARV. J.L. & TECH. 1, 11 (2001) (finding similar numbers). Given the applicable legal standard and the inherent factual and legal complexity of patent law, it should not be surprising that good opinion counsel can almost always find some basis on which to offer, in

believes the argument (or says it does), it will be able to avoid liability for willfulness.

Importantly, the doctrine of willfulness is dynamic. The question is not simply what the accused infringer thought when it adopted the technology, but also what it thinks at every moment since that time. An accused infringer that originally acted in good faith becomes a willful infringer upon learning that it is infringing a valid patent at any time, including during litigation.¹⁸ This means that every company must play the game once it has been notified of a claim of infringement, whether or not it had ever heard of the patent (or the patentee) when it began selling the product.

Patent owners, then, can begin the game by the simple expedient of sending a letter to competitors identifying their patent.¹⁹ That competitor then generally must pay for a written opinion of counsel in order to avoid a willfulness charge in any subsequent suit.²⁰ Obtaining a qualified opinion allows the accused infringer to win the willfulness game, at least temporarily. Because willfulness is considered to be dynamic, the accused in-

good faith, a favorable opinion. In any event, in a dozen years of patent litigation neither we nor anyone with whom we have discussed the issue has ever encountered a case in which the accused infringer did not raise a claim of patent invalidity.

18. *Akeva L.L.C. v. Mizuno Corp.*, 243 F. Supp. 2d 418, 421 (M.D.N.C. 2003) (“[N]either infringement nor the exercise of due care [is] a static concept.”). Patent infringement is “a continuing tort, and an action even if innocently begun does not automatically retain its purity as circumstances change.” *Pall Corp. v. Micron Separations, Inc.*, 66 F.3d 1211, 1221-22 (Fed. Cir. 1995). As one court put it, an infringer who “continued to sell the infringing [products] after the patent issued, well after it had notice the patent was pending for a *particular* design, and even after this suit for infringement was filed . . . intentionally accepted the risk of infringement.” *Avia Group Int’l, Inc. v. L.A. Gear Cal., Inc.*, 853 F.2d 1557, 1566-67 (Fed. Cir. 1988); *see also* *Pacific Furniture Mfg. Co. v. Preview Furniture Corp.*, 800 F.2d 1111, 1114 (Fed. Cir. 1986) (permitting a finding of willful and deliberate infringement despite knowledge of previous, similar designs that might constitute prior art); *Shiley, Inc. v. Bentley Labs.*, 794 F.2d 1561, 1568 (Fed. Cir. 1986) (permitting a finding of willful infringement even though the accused infringer began marketing the product before the patent ever issued).

19. Because the law imposes a duty of due care to avoid infringement on anyone who is aware of a patent, *Ryco*, 857 F.2d at 1428, even a letter identifying the patent but not alleging infringement by a specific product may be sufficient to trigger the obligation to obtain an opinion of counsel. Thus, a patentee can impose this burden on a competitor without risking a declaratory judgment suit in return. Legislation under consideration by Congress at the time of this writing would eliminate this incongruity by giving an accused infringer the right to sue for declaratory judgment if a letter triggers a willfulness obligation. *See House Panel Explores Proposals to Improve Quality of Issued Patents*, 66 PAT., TRADEMARK & COPYRIGHT J. 406, 408 (2003).

20. Not all defendants obtain advice letters; some may take the chance that the patentee won’t actually sue, or that the suit will be unsuccessful. But there are strong incentives to obtain such a letter.

fringer may need to update the opinion over time and especially during litigation to account for new information. This further increases the cost of the game.

The cycle of threat and opinion letters imposes a deeper cost than just legal fees. Disclosing the opinion of counsel in court necessarily waives the defendant's attorney-client privilege. Part III discusses this waiver in more detail.

III. THE SCOPE OF PRIVILEGE WAIVER

There are a variety of ways in which patent courts skin the privilege cat.²¹ There are two distinct axes along which the waiver determination can vary—time and subject matter—and differing points along each axis at which intelligible lines can be (and have been) drawn. While not all courts considering the waiver question have expressly parsed it this way, the arguments underlying the various positions set forth below are the ones courts typically grapple with in making decisions about willfulness and waiver.

A. The Waiver Can Vary by Time

Courts differ on the temporal extent of the waiver. Some courts have refused to place any time limit whatsoever on the waiver. Others have required waivers that extend until litigation begins, and still others have required waiver of privilege only at the time the written opinion is rendered.

Some courts suggest that the waiver should not be cut off at any point in time, at least in the ordinary case in which the alleged infringement is continuing.²² These courts parse willfulness over time and presumably would consider defining some, but not all, infringement by the same defendant as willful, depending on when the defendant acquired knowledge about the patent. So long as the accused infringer chooses to continue engaging in the conduct, its state of mind remains relevant. Just because the accused infringer once had a good faith belief that its conduct was acceptable does not mean that this belief did not change (or should not have

21. For general discussions of the conflicting opinions on the scope of privilege waiver, see Powers & Carlson, *supra* note 16, at 85-109; Cecil C. Kuhne, III, *The Advice-of-Counsel Defense in Patent Infringement Cases: How Far Does Waiver of Work Product Extend?*, 30 PEPP. L. REV. 445 (2003).

22. If the accused infringer ceases the conduct when sued, the fact that it *later* concludes that it was infringing doesn't mean that it was originally engaged in willful infringement.

changed) based on new information.²³ Accordingly, even if the accused infringer adopted the technology in good faith, a decision to continue selling a product once the accused infringer knows it to be infringing is just as culpable as willfully adopting the technology in the first place.

A second set of courts takes a somewhat more limited view of the temporal extent of the waiver. These courts suggest that the waiver should be cut off at the time the lawsuit is filed, at least where the opinion was rendered prior to the filing of the lawsuit.²⁴ The argument for this position is that once litigation is underway, the policies underlying the attorney-client privilege and, perhaps more pertinently, the work product doctrine, overcome the policies regarding the accused infringer's continuing conduct laid out in the preceding paragraph.²⁵ These courts don't focus on the policies underlying the willfulness doctrine; from the patent law perspective there seems little reason to distinguish between pre- and post-litigation conduct. Rather, these courts have made a judgment that the policies underlying the attorney-client privilege should outweigh the reasoning of the willfulness doctrine once a lawsuit is filed.

A third set of courts adopts an even more limited view of the waiver, suggesting that the waiver should cut off at the time the written opinion is rendered.²⁶ According to this approach, the accused infringer must make a good faith evaluation of its conduct either when it decides to commence potentially infringing activity or when it first decides to continue that activity after becoming aware of the patent and the potential for infringement. The accused infringer can rely on its evaluation at those two points; it does not have to update that evaluation thereafter. It's not clear why this is so. It may be that these courts, like the litigation cutoff courts, are giving the attorney-client privilege precedence over the rationale for willful-

23. See *Akeva*, 243 F. Supp. 2d at 422-24; *Chiron Corp. v. Genentech, Inc.*, 179 F. Supp. 2d 1182, 1188 (E.D. Cal. 2001).

24. If a patentee sues without first putting the accused infringer on notice, as commonly occurs, see *Brooktree Corp. v. Advanced Micro Devices, Inc.*, 977 F.2d 1555 (Fed. Cir. 1992), litigation will be proceeding while the opinion is obtained. Thus a strict temporal cutoff would be unworkable.

25. *Motorola Inc. v. Vosi Techs., Inc.*, No. 01 C 4182, 2002 U.S. Dist. LEXIS 15655 (N.D. Ill. Aug. 19, 2002); *Electro Scientific Indus. v. Gen. Scanning, Inc.*, 175 F.R.D. 539, 544 (N.D. Cal. 1997); cf. *Solaia Tech. LLC v. Jefferson Smurfit Corp.*, No. 01 C 6641, 2002 U.S. Dist. LEXIS 14562 (N.D. Ill. Aug. 7, 2002).

26. *Allergan, Inc. v. Pharmacia Corp.*, Civ. A. No. 01-141-SLR, 2002 U.S. Dist. LEXIS 19811 (D. Del. May 17, 2002). *But see VLT, Inc. v. Artysyn Techs., Inc.*, 198 F. Supp. 2d 56, 58 (D. Mass. 2002) (drawing a rather artificial distinction between some litigation counsel work product documents and others, and requiring disclosure only of those that might contradict the opinion letter).

ness law. Alternatively, these decisions may be based on an understanding that companies in reality do not (or ought not be required to) continuously reevaluate their decisions in light of legal concerns. This latter approach is practical because it avoids some of the problems we discuss below, but it provides no good theoretical explanation for why parties shouldn't reevaluate their decisions when faced with new information.²⁷

Timing of the waiver is significant to clients because it affects an accused infringer's choice of litigation counsel. Furthermore, it affects the type of advice litigation attorneys feel free to give their clients. While any waiver of privilege will affect decisionmaking at some times, extending the waiver into the litigation process magnifies those effects.

B. The Waiver Can Vary by Subject Matter

Courts are even more divided over the question of what information must be disclosed when the privilege is waived in patent litigation. There is substantial agreement that a client must disclose *other* opinions it may have received bearing on the same patents.²⁸ The other opinions include, for example, opinions from other counsel, or prior or subsequent opinions from the counsel who wrote the opinion letter, but may not include opinions from trial counsel, as parsed below. Beyond that basic point, opinions vary. Some courts may be inclined to limit waiver to the particular issues addressed in the opinion, for instance, while others will not.

1. Attorney-Client Versus Work Product Privilege

One significant subject matter distinction is between material protected by attorney-client privilege and material protected by the attorney work product doctrine. The attorney-client privilege protects only communications between attorneys and clients, while the work product doctrine protects uncommunicated work product created during or in anticipation of litigation.

Some courts suggest that the waiver should be limited to documents of which the *client* was aware at the time the opinion was rendered (for example, documents given by the client to opinion counsel or by opinion counsel to the client). According to these courts, uncommunicated work

27. Indeed, it seems quite likely that clients do update their preferences as a business matter based on new information, though not always in the continuous way that some courts seem to assume.

28. See, e.g., *Allergan*, 2002 U.S. Dist. LEXIS 19811 (compelling disclosure of all documents related to opinions in suit).

product of opinion counsel need not be produced.²⁹ Because the willfulness inquiry focuses on the client's state of mind, all that matters is what the client knew. And what the client knew is based upon what counsel communicated to the client.³⁰ What else counsel may have known or thought is irrelevant if counsel did not disclose it—unless the opinion looked so slipshod that the client reasonably should have questioned its contents.³¹ Under this limited approach, even drafts of the opinion need not be produced, so long as they were not sent to the client.³²

Other courts take the broader position that all of opinion counsel's work product should be produced, even if it is not communicated to the client.³³ A document showing what counsel knew, even if the document itself was not communicated, may bear on what counsel communicated orally, or help a fact finder to infer something about the tenor of counsel's conversations with the client. Such material may also help the patent owner depose opinion counsel, and possibly the client, more thoroughly (for example, by seeking to use the documents to refresh recollections about what may have been discussed orally). Another reason to require disclosure of work product is based directly on the Federal Rules of Civil Procedure: even if all that is relevant to the willfulness inquiry is the infringer's state of mind, discovery is permitted into not just relevant evidence but also into material "reasonably calculated to lead to the discovery of admissible evidence."³⁴ This likely includes uncommunicated memoranda that might provide a basis for questioning counsel or impeaching a witness.

29. See *Eco Mfg. v. Honeywell Int'l*, No. 1:03-cv-0170-DFH, 2003 U.S. Dist. LEXIS 7257, at *18-24 (S.D. Ind. Apr. 11, 2003); *Steelcase Inc. v. Haworth, Inc.*, 954 F. Supp. 1195, 1198-20 (W.D. Mich. 1997); *Thorn EMI N. Am., Inc. v. Micron Tech., Inc.*, 837 F. Supp. 616, 620-22 (D. Del. 1993).

30. Cf. *Thorn EMI*, 837 F. Supp. at 620-21. The willfulness inquiry will always extend beyond those communications to some extent. The court will inquire whether counsel was responsibly selected, whether the opinion letter itself is accurate and complete in its recitation of facts known to the client, and whether the letter is facially reasonable and diligent in its legal analysis. Evaluation of these issues will often require some discovery beyond the bare fact of what was communicated to the client.

31. *Steelcase*, 954 F. Supp. at 1198-99; *Thorn EMI*, 837 F. Supp. at 622-23.

32. Cf. *Cordis Corp. v. SciMed Life Sys.*, 980 F. Supp. 1030, 1034 (D. Minn. 1997).

33. *Akeva L.L.C. v. Mizuno Corp.*, 243 F. Supp. 2d 418, 425 (M.D.N.C. 2003).

34. FED. R. CIV. P. 26(b)(1); see *Beneficial Franchise Co. v. BankOne, N.A.*, 205 F.R.D. 212, 218 (N.D. Ill. 2001) (discussing waiver of post-suit attorney-client and work-product privileges); *Electro Scientific Indus. v. Gen. Scanning, Inc.*, 175 F.R.D. 539, 544-45 (N.D. Cal. 1997).

2. *Opinion Counsel Versus Trial Counsel*

Both the communications-only and the work product courts divide on the question of which lawyers must disclose the information in question. Some courts suggest that any documents that are related to the subject matter of the case and in the possession of any counsel, including trial counsel, must be produced.³⁵ This view is based on the strongest version of the arguments for continuing willfulness and the relevance of documents in discovery. If willfulness is measured on a continuing basis and if work product can shed light on the substance of oral communications, it may make sense to require disclosure of all communications and all work product on an ongoing basis, even during litigation.

Other courts, by contrast, draw a sharp distinction between communications and work product generated by opinion counsel and that generated by trial counsel. These courts hold that opinion counsel must waive work product and attorney-client privilege but that trial counsel's work product and communications with the client need not be disclosed.³⁶ These courts may be motivated in part by a belief that willfulness should depend primarily on decision making by the client at the time the product is adopted. They also observe that the possibility of settlement will be reduced if trial counsel and client cannot candidly discuss the weaknesses in the client's case.³⁷

Still other courts have taken a middle position between limiting waiver to opinion counsel and applying it to trial counsel. Some courts exempt documents reflecting the opinion work product of trial counsel from the waiver, in effect requiring disclosure of all communications from all counsel, but work product only from opinion counsel. This position balances full disclosure with the belief that trial counsel in particular must have privacy to prepare effectively for litigation.³⁸ Some courts temper this rule by

35. *E.g.*, *Akeva*, 243 F. Supp. 2d at 422-24; *Novartis Pharm. Corp. v. Eon Labs Mfg.*, 206 F.R.D. 396, 399 (D. Del. 2002); *Electro Scientific*, 175 F.R.D. at 544.

36. *See, e.g.*, *Dunhall Pharm., Inc. v. Discus Dental, Inc.*, 994 F. Supp. 1202, 1205 (C.D. Cal. 1998) (finding there is a waiver of all pre-suit privileges and post-suit attorney-client privilege but no waiver of post-suit work product privilege). This approach necessarily assumes that opinion counsel and litigation counsel are different; these courts would presumably not draw a distinction if a client hired the same lawyer or firm to write the opinion and to litigate the case.

37. For further discussion of this point, see *infra* notes 61-62 and accompanying text.

38. Here, too, the court presumes that opinion counsel is not also trial counsel; indeed, most courts have endorsed a broad waiver of the attorney-client and work product privileges where the accused infringer used the same firm as both opinion and trial counsel. *Dunhall Pharm.*, 994 F. Supp. at 1205; *see also Mosel Vitelic Corp. v. Micron Tech-*

exempting from waiver documents reflecting the opinion work product of trial counsel *unless* they contradict, cast doubt upon, or are inconsistent with the opinion letter.³⁹ This position represents a compromise between the arguments in the two preceding paragraphs.

C. Implications of the Various Rules

We take no position here on which of the various positions laid out above is most correct under prevailing law. Our point for purposes of this Article is somewhat different: precisely because there are forceful arguments supporting a variety of positions on the scope of waiver issue, and because there is no Federal Circuit precedent on point,⁴⁰ accused infringers necessarily live in fear and doubt once they make the decision to interpose the "advice of counsel" defense. And because failing to waive the privilege and rely on the opinion of counsel creates an inference that the infringement was willful under current patent law, there are very strong structural incentives to rely on the advice of counsel at trial.

IV. PROBLEMS WITH THE WILLFULNESS GAME

The waiver of privilege associated with the willfulness game imposes significant costs on accused infringers, and indeed on society more generally.⁴¹ Most of these costs inhere in the system, regardless of how broadly

nology, Inc., 162 F. Supp. 2d 307, 312-13 (D. Del. 2000); *Electro Scientific*, 175 F.R.D. at 545; *Clintec Nutrition Co. v. Baxa Corp.*, No. 94-C-7050, 1996 U.S. Dist. LEXIS 4001, at *7 (N.D. Ill. Apr. 1, 1996); *cf.* *Michlin v. Canon, Inc.*, 208 F.R.D. 172, 174 (E.D. Mich. 2002) (observing that the work product waiver analysis would have been simplified if opinion counsel were not also trial counsel, and rejecting an effort to screen two different lawyers in the same firm from each other).

39. *See Beneficial Franchise Co.*, 205 F.R.D. at 218; *Micron Separations, Inc. v. Pall Corp.*, 159 F.R.D. 361, 365 (D. Mass. 1995).

40. Powers & Carlson, *supra* note 16, at 93. While Federal Circuit's expected decision in *Knorr-Bremse Systeme Feur Nutzfahrzeuge GmbH v. Dana Corp.*, 344 F.3d 1336 (Fed. Cir. 2003), may touch on the issue of the scope of privilege waiver, it is unlikely to resolve the issue directly. That court has restricted the scope of its inquiry to the propriety and extent of the current inference of willful infringement when an accused infringer *refuses* to waive privilege. *Dana Corp.*, 344 F.3d at 1336. It may not tackle the issue of the scope of the privilege waiver at all.

41. We are not the first to criticize the application of the willfulness doctrine, though many commentators focus on aspects other than the ones we discuss. *See* Robert O. Bolan & William C. Rooklidge, *Imputing Knowledge to Determine Willful Patent Infringement*, 24 AIPLA Q.J. 157 (1996); John Dragseth, *Coerced Waiver of the Attorney-Client Privilege for Opinions of Counsel in Patent Litigation*, 80 MINN. L. REV. 167 (1995); Powers & Carlson, *supra* note 16, at 106-12; Edwin H. Taylor & Glenn E. Von Tersch, *A Proposal to Shore Up the Foundations of Patent Law that the Underwater Line Eroded*, 20 HASTINGS COMM. & ENT. L.J. 721 (1998); Shashank Upadhye, *Understand-*

the waiver of privilege is interpreted. Naturally, however, the broader the waiver, the more the willfulness game distorts the behavior of lawyers and counsel. These costs fall into five basic categories.

A. Undermining Patent Disclosures

The first significant cost associated with the willfulness game is its interference with the disclosure function of patent law. Although patent policy presumes that the public learns from patents, the willfulness game creates a strong incentive not to read patents. This problem occurs irrespective of whether any notice letters are ever sent. The mere existence of the willfulness doctrine in its current form means that any time an individual or company learns of a patent that might be relevant to its products, the company is in trouble. No matter how the patent is discovered, the company must spend tens of thousands of dollars to obtain an opinion, then forego some or possibly all of its attorney-client privilege with respect to the evaluation of the patent (and possibly with respect to the eventual litigation concerning it, too). The company would be much better off if it had never encountered the patent.

It's not always possible to remain ignorant, of course. A company that receives a notice of infringement will simply have to play the willfulness game. But companies can also be hooked into the game if their engineers run across patents that seem to cover their technology, either in the course of their scientific reading or because they notice and investigate the patent numbers marked on many products.⁴² Once a company becomes aware of a patent, it has an obligation to obtain a written opinion of counsel or risk later being held a willful infringer. To avoid this significant cost, in-house patent counsel and many outside lawyers regularly advise their clients not

ing Willfulness in Patent Infringement: An Analysis of the "Advice of Counsel" Defense, 8 TEX. INTELL. PROP. L.J. 39 (1999). Perhaps the most interesting criticism comes from Alan Herda:

Except for two spiny anteater species found in New Guinea and Australia, the duck-billed platypus is the only mammal that can lay an egg. This makes the duck-billed platypus a strange creation. Likewise, the doctrine of willful patent infringement, as it relates to defending willfulness, is also a strange creation that lays an egg.

Alan N. Herda, *Willful Patent Infringement and the Right to a Jury Trial*, 9 TEX. WESLEYAN L. REV. 181, 182 (2003). *But see* Jon E. Wright, Comment, *Willful Patent Infringement and Enhanced Damages—Evolution and Analysis*, 10 GEO. MASON L. REV. 97 (2001) (defending the current rules).

42. *See* 35 U.S.C. § 287(a) (2000) (requiring that patentees mark products with patent numbers in order to recover damages).

to read patents if there is any way to avoid it.⁴³ What you do know will certainly harm you, they reason, so it is generally better not to know.⁴⁴ Thus, from the perspective of a potential infringer, ignorance is bliss.

From society's perspective, however, ignorance is decidedly not bliss. One of the fundamental justifications for the patent system is the dissemination of ideas that would otherwise remain secret. Indeed, courts have gone so far as to hold that "patent rights are given in exchange for disclosing the invention to the public."⁴⁵ Many basic rules of patent law are de-

43. See, e.g., Taylor & Von Tersch, *supra* note 41, at 737 ("As matters now stand many companies discourage employees from reading patents. This presumably lessens the chance that the company will be found to have knowledge of a patent. However, this defeats the basic purpose of the patents [sic] laws, dissemination of information."); Dennis Fernandez, *Move Over Letterman: Top 10 Most Common IP Management Mistakes for New Companies*, PAT. STRATEGY & MGMT., July 1, 2003, at 3. ("Additionally, in many cases it may be appropriate for companies, as a matter of policy, to discourage looking at issued patents owned by other entities so as to avoid awareness of potentially infringed patents.")

44. Cf. DOUGLAS ADAMS, *THE RESTAURANT AT THE END OF THE UNIVERSE* 35 (1982) (describing "peril-sensitive sunglasses" that go entirely opaque at the first sight of danger, "and thus prevent you from seeing anything that might alarm you").

45. *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 535 U.S. 722, 736 (2002); see also *J.E.M. Ag Supply, Inc. v. Pioneer Hi-Bred Int'l*, 534 U.S. 124, 142 (2001) ("The disclosure required by the Patent Act is the *quid pro quo* of the right to exclude."); *Bonito Boats, Inc. v. Thunder Craft Boats, Inc.*, 489 U.S. 141, 150-51 (1989); *Kewanee Oil Co. v. Bicron Corp.*, 416 U.S. 470, 481 (1974); *W.L. Gore & Assoc. v. Garlock, Inc.*, 721 F.2d 1540, 1550 (Fed. Cir. 1983) ("Early public disclosure is a linchpin of the patent system."). Economists and legal scholars have also argued that disclosure is a key function of the patent system. See, e.g., Donald S. Chisum, *Anticipation, Enablement, Obviousness: An Eternal Golden Braid*, 15 AIPLA Q.J. 57 (1987) (stating that disclosure is "a primary purpose" of the enablement requirement); Suzanne Scotchmer & Jerry Green, *Novelty and Disclosure in Patent Law*, 21 RAND J. ECON. 131 (1990) (assessing the costs and benefits of the patent system by relying on the value of disclosure); Manuel Trajtenberg, *A Penny for Your Quotes: Patent Citations and the Value of Innovations*, 21 RAND J. ECON. 172, 174 (1990) (arguing that the most-cited—and therefore presumably most-read—patents are also the most valuable ones). And, as another astute observer has noted:

Another aspect of what we think of as "the patent" which should not be forgotten is that it is not only a grant of right to exclude from the government; simultaneously, it is a publication, making (in principle at least) a full public disclosure of the invention due to section 112-1. So even if it does not go into the public domain during the patent term, the public gets the advantage of knowing what the invention is and how to practice it. ("Literae patentēs" = "open letters," in short form, "patents.") That's my Saturday night lecture.

Janice M. Mueller, *A Rich Legacy*, 14 BERKELEY TECH. L.J. 895, 900 (1999) (quoting e-mail from Judge Giles Rich, Circuit Judge, Ct. of App. for the Fed. Cir., to Janice Mueller, Assistant Professor, The John Marshall Law School (Sept. 6, 1998)).

signed with an eye toward making sure that the patentee discloses her invention to the public and that the public gets the benefit of that disclosure. Most obviously, the patent system promotes this goal by requiring that every patentee describe her invention in such “full, clear, concise and exact terms” as to enable anyone of ordinary skill in the art to make and use the invention.⁴⁶ In addition, the rules of priority and statutory bars serve the disclosure function by encouraging inventors to forego trade secret protection in favor of patents and file for patent protection early.⁴⁷ By waiting too long to file a patent application or inventing without giving the world the benefit of the invention, inventors lose not only their own rights to file for a patent but also the ability to prevent a second inventor who *does* give the world the benefit of the invention from obtaining her own patent.⁴⁸ All of these rules presuppose that potential infringers actually read the patent disclosure. If they don’t, and instead take their lawyers’ advice and avoid patents in order to escape the taint of willfulness, the patent system’s goal of disclosure is frustrated. The fact that the patent system discourages searching for prior patents contrasts starkly with trademark law, which premises willfulness in part on a *failure* to search for prior marks, considering that failure a form of willful blindness.⁴⁹

B. A Trap for the Unwary

As noted above, experienced patent lawyers often advise their clients to avoid reading patents in order to avoid liability for willfulness. However, not all clients will get—or follow—such advice. Companies without patent counsel or with little exposure to the patent system might naïvely assume that it is a good idea to read patents to see what is out there before inventing, or even while bringing to market a product they already have invented but which could be improved. Others might not heed their lawyers’ advice or, more likely, might be unable to do so. A company with

46. 35 U.S.C. § 112 ¶ 1 (2000).

47. *See id.* § 102(b) (barring a patent if the application is filed more than a year after the inventor’s publication or commercial activity); *id.* § 102(g) (refusing priority in a patent dispute to one who has “abandoned, suppressed or concealed” the invention).

48. This results from judicial interpretation of 35 U.S.C. § 102(b). *See* 35 U.S.C. § 102(g) (denying prior art status to inventions that have been “suppressed” or “concealed”); *W.L. Gore & Assoc.*, 721 F.2d at 1549-50 (holding that commercial but secret prior use by a third party does not constitute prior art sufficient to bar a second inventor from patenting the invention). As a result, courts have granted patents to second developers who, unlike the true inventor, gave the public the benefit of the invention by disclosing it in a patent. *See Gillman v. Stern*, 114 F.2d 28, 30 (2d Cir. 1940); MARK A. LEMLEY ET AL., *SOFTWARE AND INTERNET LAW* 201-03 (2d ed. 2003).

49. *Int’l Star Class Yacht Racing Ass’n. v. Tommy Hilfiger U.S.A.*, 146 F.3d 66, 71 n.2 (2d Cir. 1998).

ten thousand employees will find it hard to police the behavior of its engineers. Even if the company were to engage in an education program (“Stop! Don’t read that patent!”), it is quite likely that some engineers will run across patents in the course of their work and read them. Because these engineers may not know they are doing anything of legal significance, they are unlikely to do what the patent willfulness game demands—immediately seek a written opinion of counsel. And if the engineers have read patents and the company has *not* obtained the advice of counsel, the company is presumptively a willful infringer under existing Federal Circuit precedent. That precedent draws an inference of willfulness from the failure to obtain and disclose a written opinion of counsel in litigation.⁵⁰ The existence of the willfulness game thus serves as a trap for the unwary. No matter how hard a company tries, there will likely come a day when the behavior of its engineers makes it liable for willfulness.

C. Polluting Pre-Litigation Advice

Even for those companies sophisticated enough to be aware of the willfulness game, the existence of the game distorts legal advice. The Federal Circuit’s legal rules regarding disclosure of opinions of counsel box companies who seek legal advice into a corner. If a company does not obtain a written opinion of counsel, it is presumed to be a willful infringer.⁵¹ If the company obtains an opinion of counsel but doesn’t waive privilege in order to disclose that opinion once litigation ensues, the court will infer that the opinion was adverse to the company’s interests, and once again presume that it was a willful infringer.⁵² Since companies know they will have to disclose the opinions they obtain, they place a heavy premium on obtaining a written opinion of counsel that is favorable to them—one that concludes that the patent is invalid, not infringed, or both. Indeed, the importance of the letter reading favorably for the accused infringer’s case is so great that no competent patent counsel would send written advice to a client with the bad news that they likely infringe a valid patent except under extraordinary circumstances. Opinion letters often simply remain silent on issues where the news is not good, occasionally opining only on noninfringement and ignoring validity.

50. See cases cited *supra* note 13. The Federal Circuit has agreed to reconsider this practice en banc. *Knorr-Bremse Systeme Feur Nutzfahrzeuge GmbH v. Dana Corp.*, 344 F.3d 1336 (Fed. Cir. 2003).

51. See *supra* note 14 and cases discussed therein.

52. See *Eastman Kodak Co. v. Goodyear Tire & Rubber Co.*, 114 F.3d 1547, 1561 (Fed. Cir. 1997) (“[F]ailure to introduce an exculpatory opinion of counsel at trial may support an inference that such an opinion was either never obtained or was adverse.”).

Opinions of counsel, in short, aren't unvarnished legal advice. They exist for show.⁵³ The fact that the pre-litigation advice a client receives is skewed so heavily in favor of the client's public position undermines the real value of legal advice. A company facing the threat of patent litigation—or even one that stumbles across a relevant patent—would like to know what its realistic chances are of prevailing in a lawsuit. In our experience, clients often ask lawyers for odds on who will win a suit. To make an informed decision about whether to take a license, change the product, or risk infringement liability, companies must be able to rely on their lawyers to provide honest advice that explains where the company is vulnerable and which arguments are strong.

But honest advice is precisely what the willfulness game discourages. Clients instead get opinions of counsel written with litigation in mind. If they believe those written opinions, they are more likely to litigate cases they should settle based on unduly optimistic advice.⁵⁴ Or they may ignore the opinions, treating them as the advocacy showpieces they often are, in which case they must make their own assessment of their litigation prospects without the aid of their lawyers. Neither outcome is conducive to good business decision making.

There is a third possibility, perhaps the most realistic as a practical matter. Patent lawyers may write one thing down and tell the client something different orally. This is a dodge that seems to give the client what it wants—accurate legal advice—without exposing the client to liability. But it is problematic for a number of reasons. Most obviously, it puts the lawyer in a compromising position once the privilege is waived. After providing contradictory oral advice, the lawyer will either have to lie on the stand or to testify truthfully and undermine the client's argument that the

53. The Federal Circuit has tried to cabin this practice by permitting a fact-finder to disregard an opinion "written in anticipation of litigation" rather than as part of genuine advice. *Therma-Tru Corp. v. Peachtree Doors, Inc.*, 44 F.3d 988, 997 (Fed. Cir. 1995). This merely completes the Catch-22 of the court's willfulness game. *All* opinion letters are written in anticipation of litigation; no client who really wanted legal advice rather than a willfulness defense would play the elaborate game the court has created.

54. A related problem concerns non-litigation uses of patent opinions. One experienced opinion drafter who investigated what uses were made of opinion letters told us that more than half of opinions were ultimately used not in litigation, but in connection with venture financing or acquisition-related due diligence. We have some privilege-related concerns about sharing such a letter with outside financiers. But even leaving those concerns aside, the willfulness game means that venture capitalists and acquirers doing due diligence are relying on opinion letters that don't accurately reflect the likely outcome of patent litigation. The distortion of financier decisionmaking is a further social cost to the willfulness game.

opinion letter served as the basis of a good faith belief in noninfringement. Lawyers may refuse to give oral advice that undercuts a letter or may shade that advice in order to avoid finding themselves in such a situation. In addition, oral advice is harder to communicate within a client company, and may have less force with a company's board of directors than a written opinion of counsel. This sort of duplicity represents an imperfect workaround from the client perspective. It is also unethical. Feasible workaround or not, it is hardly the sort of behavior the legal system should encourage.

D. The Choice of Counsel

The impact of the waiver of privilege not only affects legal advice, but also affects the choice of litigation counsel.⁵⁵ As noted in Part III, courts are sharply divided as to the scope of the waiver of privilege resulting from the decision to disclose an opinion of counsel.⁵⁶ While some courts limit the scope of the waiver to the opinion itself and communications that surround it, other courts go further. They may permit discovery into the communications between *litigation* counsel and the client about issues considered in the opinion of counsel on the theory that those communications offer insight into whether the client continues to believe the opinion as litigation proceeds.⁵⁷ Some courts go further still, permitting discovery not only of communications between litigation counsel and the client but also of the litigation attorney's work product.⁵⁸ This broad reading of the privilege waiver is a disaster for litigators. It may force lawyers to turn over "smoking gun" memos involving the weaknesses in their case. Lawyers will have to sort their files and permit the other side access to at least part of those files.

The waiver of privilege gives smart clients a strong incentive to hire different lawyers at different law firms to do their opinion and litigation work. First and most obviously, because the client has no real choice but to waive privilege, it is a singularly bad idea to hire the same lawyer to write the opinion and litigate the case. That lawyer will end up on the stand, making it difficult to prepare and litigate the case and undermining his credibility with the jury. His communications with the client will also

55. The willfulness rules may also affect the choice of opinion counsel. Court decisions have set out guidelines for what constitutes a competent opinion of counsel, and those opinions strongly emphasize the importance of hiring lawyers with patent expertise. See *Acoustical Design, Inc. v. Control Elecs. Co.*, 932 F.2d 939, 942 (Fed. Cir. 1991). These guidelines do affect the choice of counsel, but in a way that seems unobjectionable.

56. See *supra* notes 21-40 and accompanying text.

57. See cases cited *supra* note 25.

58. See cases cited *supra* note 26.

be subject to discovery and cross-examination—an experience that can be embarrassing at best and undercut the case at worst. Even if a different lawyer at the same firm litigates the case, that lawyer's obligation to examine and defend other attorneys at the firm can create problems. More importantly, some courts have drawn a line between opinion counsel and litigation counsel, requiring waiver of attorney-client privilege from the firm that prepared the opinion but not extending waiver beyond that firm.⁵⁹ In these courts it is particularly important that a client separate litigation counsel from opinion counsel in order to insulate litigation counsel from the problems of privilege waiver.

Our firm performs only litigation, not opinion work, and so we may actually benefit from the effective disqualification of patent firms that do opinion work from subsequent litigation work. But we do not think this disqualification is good for clients. Companies often have existing relationships with counsel whom they trust, and they may be quite surprised to learn that by turning to those lawyers for advice they have made it more difficult to rely on them in litigation. Hiring a second lawyer and bringing her up to speed involves some duplication of effort. Particularly in smaller cases, clients may resent being told by a newly-hired lawyer that the first thing they need to do is hire a second lawyer to do much the same investigation. Finally, some clients, particularly those without substantial experience with the willfulness game, will make the mistake of not hiring separate counsel, putting those clients at a significant tactical disadvantage in litigation. The willfulness game thus distorts the client's freedom to choose counsel.

E. Polluting Litigation Advice and Settlements

The distortions imposed by the willfulness game do not end once litigation begins. Because the court considers willfulness on an ongoing basis, the relevant question is not merely what the client believes at the time it learns of the patent, but whether that belief changes over time. And often it will. Opinion counsel necessarily conducts only a relatively quick, limited investigation into validity and infringement. By contrast, litigation counsel takes a more comprehensive approach once a lawsuit is filed. (Recall that an expensive opinion may cost \$100,000. A moderately expensive litigation may cost forty times that.) With such further investigation, new arguments will develop and others will look less attractive. This is particularly true once the district court judge construes the patent claims⁶⁰ because the way the claims are interpreted may well foreclose arguments on

59. See cases cited *supra* notes 39-40.

60. See generally *Markman v. Westview Instruments, Inc.*, 517 U.S. 370 (1996).

which the opinion substantially relied. Further, the Federal Circuit may well have changed the governing legal rules since the opinion was written. In extreme cases, clients must obtain an entirely new opinion of counsel based on the new information or legal developments.

Because willfulness is an ongoing inquiry, litigation counsel must be careful not to do or say anything to undermine the client's confidence in opinion counsel's written opinion. If something undermines the client's confidence, the client can no longer rely on that opinion in good faith. As a result, litigation counsel may be reluctant to tell the client anything that undermines the written opinion. Litigation counsel may feel constrained to make arguments that aren't likely to win, simply because the written opinion relied on them. Similar to the problems with pre-litigation advice, litigators may not be able to tell clients the unvarnished truth, which may affect their ability to litigate cases effectively. Indeed, in some cases litigators tell their in-house legal counterparts of problems with the case, but then make it clear that in-house counsel can't share those problems with the ultimate decisionmakers, because if the ultimate decision maker knew of the problems that knowledge would undermine the company's confidence in the opinion letter.⁶¹ The problem is particularly acute in settlement discussions. To settle a case, litigators must be able to convince the client of the weaknesses in its own case. But doing so may undermine the opinion letter and leave the client with the problem of willfulness liability if the case doesn't settle. When litigators can't tell the client about the shortcomings of the case, or if they can tell the inside counsel but not the ultimate decision maker, cases become much harder to settle before trial.

Courts have sought to mitigate the harms of the waiver of privilege by bifurcating the willfulness and infringement inquiries, either by staying willfulness discovery (and therefore necessarily bifurcating the issues at trial) until the jury has found that a valid patent was infringed, or by permitting discovery to proceed but bifurcating the issues at trial.⁶² If willful-

61. It is the company decision makers that must ultimately have a good faith belief in order to avoid infringement.

62. The Federal Circuit has indicated that a bifurcated trial, at least, is the preferred approach to the "dilemma" of an accused infringer who must choose between asserting the privilege and defeating willfulness. *Quantum Corp. v. Tandon Corp.*, 940 F.2d 642, 643-44 (Fed. Cir. 1991). Although the court strongly suggested bifurcation, it refused to require it. *Id.*

In our experience, courts do not have a uniform practice as to staying or not staying willfulness discovery, and a significant number of courts do not stay such discovery. *Compare Plasmanet, Inc. v. Apax Partners*, No. 02 Civ. 9290 (BSJ) (THK), 2003 U.S. Dist. LEXIS 13533 (S.D.N.Y. Aug. 5, 2003) (bifurcating and staying), *with Dents-*

ness is tried at a separate time and no discovery into the opinion of counsel is taken unless and until the patent is found valid and infringed, the accused infringer will be able to litigate the underlying case without the inconvenience of having counsel subject to discovery. But bifurcation is far from a perfect solution, especially where only trial is bifurcated and discovery is not stayed. Bifurcating the trial but keeping discovery comprehensive does not do away with the problem of privilege waiver and the resulting tendency to skew legal advice; it just delays the problem. Counsel who know they may well be required to turn over their files will still be careful not to say or do anything that would ultimately undermine their client's willfulness case. Nor does bifurcation solve any of the other problems associated with the willfulness game.

Bifurcation may also be impractical.⁶³ Courts already require bifurcation and sometimes trifurcation of antitrust counterclaims,⁶⁴ and some bifurcate damages from liability.⁶⁵ Since virtually every case involves a willfulness claim, requiring bifurcation of such claims would threaten to divide every patent case into three, four, or even five separate trials. Further, to be effective, willfulness bifurcation must delay discovery until after the first trial ends. This means that the court will have to select a new jury to try the willfulness issue. Doing so not only imposes administrative costs—the second jury will have to be brought up to speed on the facts of the case—but also presents potential Seventh Amendment problems.⁶⁶

ply Int'l, Inc. v. Great White, Inc., Civ. A. No. 1:CV-99-1346, 2000 U.S. Dist. LEXIS 13108, at *9 (M.D. Pa. Sept. 1, 2000) (refusing to stay discovery).

63. See *Akeva L.L.C. v. Mizuno Corp.*, 243 F. Supp. 2d 418, 421 n.4 (M.D.N.C. 2003) (criticizing bifurcation as impractical); *F&G Scrolling Mouse, LLC v. IBM Corp.*, 190 F.R.D. 385 (M.D.N.C. 1999) (same).

64. See *Hunter Douglas, Inc. v. Comfortex Corp.*, 44 F. Supp. 2d 145 (N.D.N.Y. 1999) (ordering trifurcation); 1 HERBERT HOVENKAMP ET AL., IP AND ANTITRUST § 11.3b6(B) (Supp. 2003) (discussing bifurcation of antitrust claims).

65. See, e.g., *Plasmanet*, 2003 U.S. Dist. LEXIS 13533, at *6 (discussing "a line of cases which recognizes that patent cases are particularly amenable to bifurcation on issues of liability and damages").

66. The Reexamination Clause of the Seventh Amendment prevents a jury from reexamining facts already found by the first jury. See, e.g., Patrick Woolley, *Mass Tort Litigation and the Seventh Amendment Reexamination Clause*, 83 IOWA L. REV. 499 (1998). It is hard to see how a willfulness jury could avoid injecting its own view of validity and infringement into the determination whether the accused infringer had a good faith belief that the patent was invalid or was not infringed.

This problem would be avoided if there were no right to a jury trial on willfulness. Alan Herda has argued that there is such a right, however. Herda, *supra* note 41, at 182.

V. WHAT IS TO BE DONE⁶⁷

The present structure of the willfulness doctrine distorts legal advice, interferes with settlements, and frustrates the disclosure function of the patent system. There are several possible solutions to these problems. These options are abolishing willfulness altogether, eliminating the requirement that accused infringers disclose opinions of counsel, and changing the way willfulness is defined. In our view, the last approach, while far from a perfect solution, best comports with the policies of patent law while preserving the policies underlying the attorney-client privilege.

A. Abolishing Willfulness

The simplest way to solve the problems described in the previous section would be to abolish the doctrine of willfulness altogether. Without the fear of treble damage liability for willful infringement, companies would have no reason to disclose their opinions of counsel, and hence no obligation to waive the attorney-client or work product privileges.⁶⁸ As a result, they could make informed decisions in reliance on the unbiased advice of counsel without worrying about disclosing that advice in open court. They would be free to use the same lawyer to provide advice and to litigate the case, eliminating wasteful duplication of effort. And they would have no reason to discourage their engineers from reading patents in order to learn either of technical advances or possible legal impediments to a particular project. Furthermore, because the law of willfulness is entirely a judicial creation, courts could arguably modify or abolish it without waiting for Congress to legislate.⁶⁹

Abolishing the willfulness doctrine altogether, however, could create two related problems. First, patentees could be undercompensated: if they recover only their lost profits or a reasonable royalty (of which more be-

67. With apologies to V.I. Lenin.

68. Patent infringement is a strict liability offense. *Fla. Prepaid Postsecondary Educ. Expense Bd. v. College Savings Bank*, 527 U.S. 627, 646 (1999) (holding patent infringement does not require proof of intent to infringe). This means that proving innocent development or good faith belief in the invalidity of a patent will not protect an accused infringer from liability. All it does under current law is preclude enhanced damages for willful infringement.

69. The patent damages statute provides for enhanced damages beyond lost profits or a reasonable royalty, but makes no mention of willfulness. 35 U.S.C. § 284 (2000). If willfulness were to be abolished, that statute might be rendered a nullity unless there were other cases in which courts might enhance damages. The obvious alternative to willfulness for enhanced damages is a case in which the statutory remedy undercompensates for the infringement.

low), they are still worse off by at least the amount of legal fees they have expended to obtain those remedies.

Second, abolishing willfulness damages entirely could create problematic incentives for companies that are considering infringing. The remedies for ordinary patent infringement include a permanent injunction to prevent future infringement and damages designed to compensate the patent owner for past infringement.⁷⁰ Those damages are calculated as the greater of the patentee's lost profits or a reasonable royalty for the defendant's use.⁷¹ Proving lost profits is rather difficult in patent cases, however. The patentee must make a detailed economic showing of demand for the patented good, the absence of noninfringing substitutes, evidence that the patentee would have made the sales if the defendant had not (taking into account the patentee's capacity to produce the products), and the profit the patentee makes per unit.⁷² Courts require sophisticated economic analysis of how the price of a good would have interacted with the quantity sold in the absence of infringement and will not award lost profits from eroded prices without substantial economic evidence.⁷³ And for one significant class of patentees—patentees that do not manufacture the patented product, but merely license the technology—lost profits will never be available.⁷⁴ The result is that even some patentees that compete directly with infringers will not be able to show lost profits and will receive only a reasonable royalty as compensation for infringement.

70. *Id.* §§ 283-284 (providing respectively for injunction and damages). Preliminary injunctions are sometimes available in patent cases, but the standards for granting such injunctions are high and so they are quite rare today. *See Amazon.com, Inc. v. Barnesandnoble.com, Inc.*, 239 F.3d 1343, 1365-66 (Fed. Cir. 2001) (reversing grant of preliminary injunction where there was any serious question whether the patent was valid).

71. 35 U.S.C. § 284. For a discussion and critique of patent damages rules, see Roger D. Blair & Thomas F. Cotter, *Rethinking Patent Damages*, 10 TEX. INTELL. PROP. L.J. 1 (2001).

72. *See Panduit Corp. v. Stahl Bros. Fibre Works, Inc.*, 575 F.2d 1152, 1158 n.5 (6th Cir. 1978) (setting out the standard test for proof of lost profits in patent cases).

73. *See, e.g., Crystal Semiconductor Corp. v. TriTech Microelectronics Int'l, Inc.*, 246 F.3d 1336, 1358-60 (Fed. Cir. 2001) (emphasizing need for economic analysis of how markets would react to patented technology); *Grain Processing Corp. v. Am. Maize Prods. Co.*, 185 F.3d 1341, 1356 (Fed. Cir. 1999) (demanding "reliable economic proof" that the patentee would have made the sale but for the infringement, considering both the price that would have been charged and the possibility that the infringer might shift to a noninfringing substitute).

74. This is because they will not be able to show that they would have made the infringer's sales themselves but for the infringement, as the lost profits law requires. *Amstar Corp. v. Envirotech Corp.*, 823 F.2d 1538, 1542-43 (Fed. Cir. 1987).

The fact that damages in many patent cases will be limited to a reasonable royalty creates problems of deterrence. Specifically, it may underdeter willful infringers. A reasonable royalty is the amount an accused infringer would have had to pay the patentee to license the patent in the first place. In the absence of treble damages for willful infringement, if a company knows that retroactively paying a reasonable royalty is the worst that will happen if it is found to infringe, it may be more willing take its chances in court.⁷⁵

This is not to say that patent law would provide *no* deterrence if willfulness damages were abolished altogether. The court will generally set the royalty at a somewhat higher rate than the parties would have bargained for *ex ante* because the jury has determined that the patent is valid and infringed.⁷⁶ Further, the prospect of paying \$2 million or so in legal fees may provide additional deterrence.⁷⁷ And most importantly, infringers generally will not want to make substantial investments in assets specific to the infringing technology if they know that they will ultimately be enjoined from using that technology after trial. While patent litigation takes a very long time—the average patent case that goes to court is decided 8.6 years after the patent is issued⁷⁸—and a significant number of patents are

75. Willful infringers may also be motivated by the prospect that they can escape personal liability through the limited liability rules of corporation law, and that the corporation itself may be able to enter bankruptcy and avoid its debts. But trebling damages won't solve these problems.

76. *See, e.g.*, *Fromson v. Western Litho Plate & Supply Co.*, 853 F.2d 1568, 1576 (Fed. Cir. 1988); *Stickle v. Heublein, Inc.*, 716 F.2d 1550, 1563 (Fed., Cir. 1983); *Panduit*, 575 F.2d at 1158 (“The setting of a reasonable royalty after infringement cannot be treated . . . as the equivalent of ordinary royalty negotiations among truly ‘willing’ patent owners and licensees. . . . [T]he infringer would have nothing to lose, and everything to gain if he could count on paying only the normal, routine royalty non-infringers might have paid.”). This may be a significant difference. As Carl Shapiro points out, a patent does not itself confer a right to exclude competitors but merely a right to try to exclude them. Carl Shapiro, *Antitrust Limits to Patent Settlements*, 34 RAND J. ECON. 391 (2003). *Ex ante* royalty negotiations will take into account the possibility that a patent will be held invalid or not infringed. A royalty agreement over a valid, infringed patent will tend to produce a significantly higher royalty rate.

77. *See* AM. INTELL. PROP. LAW ASS'N (AIPLA), 2003 REPORT OF THE ECONOMIC SURVEY 22 (2003) (surveying members and reporting the cost of patent litigation to average \$2 million where \$1-25 million was at stake in the case, and nearly \$4 million where more than \$25 million was at stake). In large patent cases fees and costs may exceed those estimates substantially.

78. *See* John R. Allison & Mark A. Lemley, *Empirical Evidence on the Validity of Litigated Patents*, 26 AIPLA Q.J. 185, 236 tbl.12 (1998).

invalidated in court,⁷⁹ cases involving deliberate copying of a patent will likely be resolved more quickly than average. These cases will also more likely be resolved in the patentee's favor, all other things being equal, unless all of the defendant's significant witnesses are willing to lie and able to do so convincingly. The risk of being shut down by an injunction that will render a defendant's manufacturing investments useless may be a powerful reason not to infringe.⁸⁰

Still, were willfulness to be abolished altogether, it seems likely that companies would, at the margins, be more willing to infringe a patent. Particularly where a patent is in the last half of its term, companies might take their chances that 1) the court will strike down the patent, 2) the patent term will end before the case does, preventing the court from issuing an injunction, or 3) the benefits of having been in the market during the pendency of the lawsuit are worth any damages.

This will be particularly true for those defendants who know they are copying from a patentee that does not manufacture products based on the patent, as those defendants know for sure they will owe only a royalty and not their victim's lost profits. It is worth noting that non-manufacturing patent owners may not be the inventors the system has the strongest interest in protecting. While some inventors may have perfectly legitimate reasons for not manufacturing their inventions (such as individuals or small companies without the resources to devote to building a physical plant), the wide availability of venture capital means that, in most industries, a patentee who wants to enter the market can find a way to do so. Accordingly, many non-manufacturing owners are holdup artists or "trolls" who are in the business of litigation, not innovation.⁸¹ Arguably these trolls contribute little or nothing to society, while substantially increasing the social costs of the patent system.⁸² A number of commentators have suggested that patent law should give less protection to non-manufacturing

79. *Id.* at 205-06 (finding that 46% of all litigated patents were invalidated); Kimberly A. Moore, *Judges, Juries and Patent Cases: An Empirical Peek Inside the Black Box*, 99 MICH. L. REV. 365, 385-86 (2000) (finding that patentees lose 42% of the subset of cases that make it as far as trial).

80. See ROBERT PATRICK MERGES & JOHN FITZGERALD DUFFY, *PATENT LAW AND POLICY* 1079 (3d ed. 2002) (making this point).

81. The most notorious such patent owner was Jerome Lemelson, who extracted more than \$1 billion in royalties from his 500+ patents without ever making a product. See Mark A. Lemley & Kimberly A. Moore, *Ending Abuse of Patent Continuations*, 84 B.U. L. REV. (forthcoming 2004) (discussing abuse of the patent system by Lemelson and others).

82. See *id.*; Michael J. Meurer, *Controlling Opportunistic and Anti-Competitive Intellectual Property Litigation*, 44 B.C. L. REV. 509 (2003).

patent owners than it currently does.⁸³ Abolishing willfulness altogether might serve that goal by rewarding patentees who participate in the market, and therefore can claim lost profits, while giving less relief to other patentees. Limiting non-manufacturing patent owners to such a remedy would effectively create a doctrine of “efficient infringement,” permitting more efficient companies to infringe and pay damages that amount to less than the profit they can make by infringing.⁸⁴

In sum, while the possibility of treble damages for willful infringement provides some additional deterrence to unscrupulous copyists who might otherwise decide to infringe a patent deliberately, there are other significant incentives that deter copying, and the current system may provide the greatest marginal protection to those patentees who are least worthy of such help. We will return momentarily to the question of how optimally to deter deliberate copyists while minimizing the damage to the disclosure goal of patent law.⁸⁵

B. Eliminating the Requirement to Waive Privilege

If we keep the doctrine of willfulness, a second possibility is to abolish the effective requirement that accused infringers obtain and then disclose a written opinion of counsel. As we explained in Part III, companies do this because patent law permits fact finders to infer willfulness from failure to obtain an opinion of counsel, and further permits them to infer that an opinion was adverse if it was obtained but was not disclosed. Since, as we noted in Part IV.B, essentially all written opinions obtained by companies facing a charge of infringement are favorable, accused infringers generally decide to disclose the opinion rather than have the court draw an adverse inference that is almost certainly factually incorrect.⁸⁶ The inference, then,

83. See, e.g., Michelle Armond, Comment, *Introducing the Defense of Independent Invention to Motions for Preliminary Injunctions in Patent Infringement Lawsuits*, 91 CALIF. L. REV. 117 (2003); Julie S. Turner, Comment, *The Nonmanufacturing Patent Owner: Toward a Theory of Efficient Infringement*, 86 CALIF. L. REV. 179 (1998).

84. See Shapiro, *supra* note 76, at 392-93 (arguing that infringement by a more efficient company is likely if damages are not multiplied).

85. See *infra* notes 102-105 and accompanying text.

86. While it is certainly possible that defendants who wish to withhold an opinion of counsel do so because they have received the rare negative opinion letter, it is far more likely that they are reluctant to disclose the opinion either because they don't wish to waive the privilege, or because the opinion, while favorable, is obsolete in view of subsequent discovery, prior art searches, or claim construction rulings.

One common reason for wanting to withhold even a favorable opinion is that the arguments counsel wants to make at trial contradict those made in the opinion. For example, an opinion letter might interpret the claims of a patent narrowly in concluding that there is no infringement, while at trial the accused infringer might want to forego that

isn't really a reasonable estimate of the facts, but rather a threat designed to force the disclosure of information the accused infringer would rather not give up. And it is a threat that would be inconceivable in other areas of litigation. For instance, no court would infer from a party's failure to call an expert witness that the party must have hired an expert who gave a negative opinion.

As a result of these concerns, some have suggested that the solution to the willfulness game is to eliminate the inferences the courts draw from an accused infringer's failure to obtain or disclose a written opinion of counsel.⁸⁷ The Federal Circuit sitting en banc has taken a case that directly presents these issues.⁸⁸ Eliminating the adverse inference from nondisclosure may give accused infringers greater freedom of action. They will be somewhat less likely to rely on the opinion of counsel and thereby suffer a waiver of the attorney-client privilege, particularly in cases pending before a court that reads the scope of the privilege waiver broadly. Accordingly, accused infringers will have more freedom to choose counsel without worrying about disqualification, and to obtain accurate legal advice during litigation. Eliminating the adverse inference from failure to *obtain* an opinion at all would also ameliorate the problems of skewed pre-litigation advice and the trap awaiting those unaware of the requirement.⁸⁹ And it seems consistent with our treatment of willfulness in other areas of law. Criminal law and intentional torts likewise inquire into the defendant's state of mind but, unlike patent law, do not require the defendant to waive the attorney-client privilege.

infringement defense because a broader understanding of the patent would render it invalid in light of newly discovered prior art not available to the lawyer who wrote the opinion.

87. See Powers & Carlson, *supra* note 16, at 108; M. Patricia Thayer & Elizabeth A. Brown, *Tendering Advice of Counsel In Patent Litigation: Damned If You Do, Damned If You Don't*, 3 SEDONA CONF. J. 111 (2001), available at <http://www.hewm.com/use/articleDetails.asp?articleID=745>. The American Intellectual Property Law Association has taken this position, for example. Brief for Amicus Curiae American Intellectual Property Law Association at 7-10, Knorr-Bremse Systeme Feur Nutzfahrzeuge GmbH v. Dana Corp., 344 F.3d 1336 (Fed. Cir. 2003) (en banc) (No. 01-1357).

88. Knorr-Bremse Systeme Feur Nutzfahrzeuge GmbH v. Dana Corp., 344 F.3d 1336 (Fed. Cir. 2003) (en banc).

89. Neither approach would encourage competitors to read patents, however, because they would still risk treble damages liability by doing so. Only abolishing willfulness outright would do that. Eliminating the inferences would reduce the cost of learning about a patent, however, by removing the privilege waiver risk and possibly by removing the need for a written opinion altogether. They may therefore make it somewhat more likely at the margins that lawyers will countenance their clients reading patents.

However, while eliminating the adverse inference from failure to disclose an opinion of counsel may be a good idea,⁹⁰ it is unlikely to solve the problems of privilege waiver in more than a few cases. Unlike the ordinary criminal or tort case, there is no question in the average patent willfulness dispute about whether the defendant was aware of the patent and is making the product in question. Rather, the disputed issues are whether the patent is valid and the defendant's product infringes it. Even in the absence of a legal instruction to presume willfulness, the trier of fact is likely to presume willfulness as a factual matter. This is because the very same trier of fact will just have ruled against that defendant at trial on both validity and infringement.⁹¹ The defendant's best hope of avoiding liability for willfulness is to rely on the advice of counsel. As a result, even without the coercive pressure of the adverse inference, many accused infringers will want to disclose a favorable opinion letter and, accordingly, will have no choice but to waive privilege.

C. An Objective Standard

Doug Lichtman has suggested to us that many of the problems of willfulness could be solved if willfulness were tested objectively rather than subjectively. Under this approach, it would not matter whether an accused infringer who was aware of a patent actually believed that the patent was valid and infringed, only what a reasonable person in the accused infringer's position would have thought was relevant. Replacing the current

90. We are less persuaded that the court should reverse the adverse inference it draws when a company that learns of a patent fails to obtain guidance from counsel at all. While eliminating this inference would obviously have some substantial benefits, as noted in text, a company that is aware of a patent but makes its own decision to build a product without legal advice seems to us to be acting at least recklessly, if not willfully. Patent law is an arcane subject full of counterintuitive rules and not easily accessible to laymen. An engineer's rough sense that the patent is obvious or that what she is doing is different enough may not have any actual basis in patent law. We would, therefore, be inclined to treat the failure to obtain legal advice as evidence of willfulness, though not determinative evidence.

91. This is necessarily the case in any willfulness dispute, since willful infringement presupposes infringement of a valid patent. Sometimes the questions will be determined by the factfinder at the same time, and sometimes sequentially, however. *See supra* notes 62-66 and accompanying text (discussing bifurcation).

Evidence of a legitimate effort to design around the patent might help to rebut this *de facto* presumption. But at least one court has held that if an accused infringer refuses to waive privilege, it is not entitled to rely on evidence of its design-around efforts because it would be unfair for the patentee not to have the evidence with which to evaluate those efforts. *Advanced Cardiovascular Sys. v. Medtronic, Inc.*, 265 F.3d 1294, 1309-10 (Fed. Cir. 2001). Under this approach, there will be little evidence indeed that a non-waiving defendant can present to show that its conduct was not willful.

subjective inquiry into willfulness with an objective one has some attractive features. It would do away with the requirement to obtain and disclose an opinion of counsel, since the accused infringer's state of mind would not matter. It would therefore prevent the distortions of legal advice we described above. The Federal Circuit's en banc order in *Knorr-Bremse* suggests that the court may be willing to consider treating willfulness as an objective standard; it asks the parties to brief the question whether proof of a substantial though ultimately unsuccessful defense to infringement should negate a finding of willfulness.⁹²

Although an objective approach has substantial advantages over the current willfulness game, it is not really consistent with the concept of willfulness. Rather, such a standard would effectively punish accused infringers who raised unreasonable defenses, whether or not they knew those defenses were unreasonable. Further, as a practical matter it would require a jury to determine the relative quality of legal arguments without substantial experience in the range of such arguments. Patent law already has such a rule—it awards attorneys fees to prevailing patentees in exceptional cases.⁹³ The point of willfulness is to focus on culpable conduct, which seems to require evidence of bad intent. The objective approach is not so much a bad idea as one that is orthogonal to the purposes of willfulness doctrine. It is really better thought of, therefore, as a variant of the proposal to abolish willfulness altogether, one that would enhance damages for some willful infringers and for some defendants who were not willful but whose arguments were weak.

D. Redefining Willfulness

A final solution to the willfulness game is to take a closer look at what constitutes willfulness. Patent law today defines willfulness on an ongoing basis. A company accused of infringement that adopts a product without knowledge of a patent may become a willful infringer once it learns of the patent, if it continues to sell the product.⁹⁴ Indeed, even an accused infringer that has properly obtained and relied in good faith upon an opinion of counsel may be a willful infringer if it later discovers facts that undermine that reliance. The result is that even companies that adopt a technology in perfect good faith and exercise due care to avoid infringement when they later find out about a patent may ultimately be found to be willful infringers.

92. *Knorr-Bremse*, 344 F.3d at 1336.

93. 35 U.S.C. § 285 (2000).

94. See *supra* note 18 and cases cited therein.

Arguably this is a rather artificial definition of willfulness. It requires companies to make their product adoption decisions on a continuous basis, reevaluating whether to sell the product every time they learn new information. While in one sense this is an accurate portrayal of what companies do—business decisions are always subject to second-guessing as new information comes in—in a more important sense it is an unrealistic burden to impose. Deciding to produce a particular product involves making a series of investments in innovation, personnel, and in physical plant.⁹⁵ This is particularly true if the company deciding to produce a product develops that product on its own.⁹⁶ Once those investments have been made, the decision to continue producing is no longer made on a blank slate. A company that has invested \$3 billion in a new semiconductor fab⁹⁷ or \$800 million in developing, testing, and marketing a new drug⁹⁸ or even a few million dollars in a more mundane invention, can hardly be expected to throw those product-specific investments away every time the company is confronted with one of the more than two million patents currently in force in the United States.⁹⁹ While those investments are sunk costs, once

95. See David J. Teece, *Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy*, 15 RES. POL'Y 285, 288-90 (1986) (discussing such "co-specific assets").

96. This is quite possible even if someone else has patented the invention, because patent law makes even independent development illegal. See, e.g., Fla. Prepaid Postsecondary Educ. Expense Bd. v. College Savings Bank, 527 U.S. 627, 645 (1999). For an argument that patent law should not do so, but rather should permit independent development to serve as a defense, as it does in trade secret and copyright cases, see Stephen M. Maurer & Suzanne Scotchmer, *The Independent Invention Defense in Intellectual Property*, 69 ECONOMICA 535 (2002); see also Armond, *supra* note 83, at 117; John S. Liebovitz, Note, *Inventing a Nonexclusive Patent System*, 111 YALE L.J. 2251 (2002).

97. Mark LaPetus, *Leading-Edge Fab Costs Soar to \$4 Billion*, Silicon Strategies, at <http://www.siliconstrategies.com/article/showArticle.jhtml?articleId=10801238> (Mar. 10, 2003); see also Steve Lohr, *World-Class Chip, but a Fragile Business*, N.Y. TIMES, Aug. 4, 2003, at C1 (noting that fabs cost two to three billion dollars each).

98. Estimates of the average cost of drug development and testing range from \$150 million to \$500 million; the latter is the industry's figure. Compare PHARM. RESEARCH AND MFRS. OF AM., WHY DO PRESCRIPTION DRUGS COST SO MUCH . . . AND OTHER QUESTIONS ABOUT YOUR MEDICINES (2000), available at <http://www.phrma.org/publications/publications/brochure/questions>, with Public Citizen, *Rebuttals to PhRMA Responses to the Public Citizen Report, "Rx R&D Myths: The Case Against the Drug Industry's R&D 'Scare Card'"*, at http://www.citizen.org/congress/reform/drug_industry/corporate/articles.cfm?ID=6514 (created Nov. 28, 2001). A recent estimate is even higher—over \$800 million per drug. See Gardiner Harris, *Cost of Developing Drugs Found to Rise*, WALL ST. J., Dec. 3, 2001, at B14.

99. The Information Products Division of the U.S. Patent and Trademark Office reports the number of utility patents issued each year, which has increased from 70,860 in 1986 to 166,039 in 2001, with further increases in 2002 and likely in 2003. Because the

they have been made the cost of continuing to sell a product is far less than the cost of starting the development and investment process all over again for a new product. To make such a company liable for willful infringement merely because it came to believe at some point during litigation that the patent claim against them might have merit seems unfair.¹⁰⁰ Imposing liability may also impose a substantial social cost in the form of wasted investment, since companies with significant potential financial exposure in a patent suit will be deterred from litigating close cases if their financial exposure will triple as a result. And it will likely discourage design-arounds by increasing the sanction for unsuccessful design-around efforts. This would result in a significant cost to society. Not only can design-arounds have collateral social benefits, but the law explicitly encourages efforts to design around a patented invention.¹⁰¹

The real goal of willfulness seems more modest. Willfulness law is designed to deter unscrupulous copyists from taking advantage of the patent disclosure in order to copy the patentee's invention and rush it to market.¹⁰² A much more circumscribed definition of willfulness, one that fo-

patent term ran for seventeen years from issue until the mid-1990s, all patents issued before this date in 1986 have yet to expire. Totalling these patents leads to roughly two million unexpired patents, depending on how many have issued in 2002 and 2003. U.S. PAT. & TRADEMARK OFF., STATISTICAL REPORTS AVAILABLE FOR VIEWING, at <http://www.uspto.gov/web/offices/ac/ido/oeip/taf/reports.htm#PSR> (last visited Oct. 28, 2003).

100. Indeed, the Federal Circuit has rejected in the strongest possible terms the idea that *patentees* should have to change their behavior merely because they are faced with a claim that they behaved inappropriately. In *Kingsdown Medical Consultants, Ltd. v. Hollister, Inc.*, 863 F.2d 867 (Fed. Cir. 1988), the district court found inequitable conduct based in part on the patentee's continuation of the lawsuit after its error in prosecution was called to its attention. The federal circuit said:

[A] suggestion that patentees should abandon their suits, or disclaim or reissue, in response to every charge of inequitable conduct raised by an alleged infringer would be nothing short of ridiculous. The right of patentees to resist such charges must not be chilled to extinction by fear that a failure to disclaim or reissue will be used against them as evidence that their . . . intent was deceitful.

Kingsdown Med., 863 F.2d at 875-76. The court to date has shown no similar solicitude for patent *defendants* who wish a chance to argue their case without being labeled willful as a result.

101. See Matthew J. Conigliaro et al., *Foreseeability in Patent Law*, 16 BERKELEY TECH. L.J. 1045 (2001) (discussing the law and policy arguments in favor of design-arounds).

102. See, e.g., *Seymour v. McCormick*, 57 U.S. 480, 488-89 (1853) (describing the "wanton and malicious pirate" as the target of willfulness law). Indirect evidence of this can also be found in Federal Circuit decisions that find intentional copying to be a factor in aggravation of willful infringement. See, e.g., *Electro Med. Sys. v. Cooper Life Scis., Inc.*, 34 F.3d 1048, 1058 (Fed. Cir. 1994) (denying enhanced damages, despite willful-

cuses only on the accused infringer's state of mind at the time it initially adopts the product, can provide appropriate deterrence. An accused infringer that copies from the patent or the patented product should probably be liable for willful infringement, at least in the absence of evidence of a good faith belief that it was not infringing or that the patent was invalid. By contrast, an accused infringer who independently develops the invention, or even one who copies in good faith, doesn't seem particularly culpable merely because it later obtains a copy of the patent or discovers that the legal advice it obtained was faulty. Indeed, Matt Powers and Steven Carlson have reviewed the historic evidence in detail and concluded that, before the creation of the Federal Circuit, willful infringement meant deliberate copying, not continued use of an independently created invention.¹⁰³

We suggest, therefore, that willfulness be redefined to focus only on the accused infringer's state of mind at the time she adopts the product in question.¹⁰⁴ This narrower definition of willfulness would not only comport with the core purposes behind the willfulness requirement, but would solve a significant number of problems with the willfulness game. First, and most obviously, limiting the inquiry into state of mind to the time of the adoption decision would render all subsequent legal advice irrelevant. Courts would have no basis for compelling litigation counsel to waive the attorney-client or work product privileges. As a result, the problem of polluting litigation advice and settlement negotiations would be eliminated entirely. Second, because accused infringers would not have to disclose the later opinions of trial counsel, clients would be free to choose opinion counsel as trial counsel without fear of losing the attorney-client or work product privileges for litigation advice.¹⁰⁵ Third, redefining willfulness in

ness, where the infringer did not engage in misconduct or copy the patented product); *Amsted Indus. v. Buckeye Steel Castings Co.*, 24 F.3d 178 (Fed. Cir. 1994).

Doug Lichtman has argued to us that imposing extremely high damages remedies makes patent law more of a property rule and therefore encourages the negotiation of licenses. We are not persuaded. Patent law is already enforced by means of a property rule; injunctions are standard in virtually all cases once infringement is found. Trebling damages will raise the cost to an accused infringer of challenging a patent, and will therefore presumably deter some challenges to patents and increase patentee licensing revenue. But there is no reason to think that simply shifting the median damage award upwards will encourage more settlements, rather than just settlements at a higher price.

103. Powers & Carlson, *supra* note 16, at 55-79.

104. Powers and Carlson make a similar suggestion among their several arguments for reform and we endorse their approach. *See id.* at 107-08.

105. It would probably still be unwise to choose the same lawyer or firm to write the opinion and litigate the case, because that lawyer would have to testify as to the opinion,

this way would eliminate the costs associated with obtaining additional opinions of counsel as new facts come to light.

In addition, the refocused willfulness doctrine we propose would have one significant collateral benefit—it should reduce the utility of “submarine patents.” Submarine patents are obtained after the product they are intended to cover has been put on the market by a competitor. They are based on a continuation application drafted specifically to cover the targeted product.¹⁰⁶ Submarine patents have been much criticized,¹⁰⁷ and they are particularly problematic precisely because of the current willfulness doctrine. Having been drafted to cover the particular device in question, they present an expensive challenge for even the most creative opinion counsel. Moreover, because the accused infringer necessarily has invested the costs to develop the product without being aware of the patent (because the patent did not exist during the product’s development), the accused company will be faced with the particularly sharp dilemma of either risking a likely willfulness finding or walking away from its prior investment. Under our proposed reformulation, no submarine patent could ever support a willfulness claim against its intended target, which seems a substantial virtue.

This narrower definition of willfulness would not eliminate the need for written opinions of counsel altogether. A company that sought to make a new product would have to obtain such an opinion if it were aware of a patent that arguably read on that product. As a result, companies would still be faced with the problem of skewed pre-litigation advice and would still risk being trapped into a willfulness finding if they fail to obtain an opinion of counsel. Importantly, however, the *number* of circumstances in which an accused infringer would need to obtain such an opinion should decline sharply. Companies would only need an opinion of counsel if they were actually aware of a patent at the time they decided to adopt a product potentially covered by the claims of that patent. Companies that developed a product independently would, by definition, not be willful infringers, even if they later learned of the patent. Similarly, those that copied a product in the marketplace without knowledge that it was covered by a patent

and in doing so may tend to undermine his own case. But at least the risk of losing privilege would be gone.

106. For discussions of submarine patenting, see, for example, Stuart J.H. Graham, *Hiding in the Patent’s Shadow: Firms’ Use of Secrecy to Capture Value from New Discoveries* (Working Paper, 2002); Mark A. Lemley, *An Empirical Study of the Twenty-Year Patent Term*, 22 AIPLA Q.J. 369 (1994); James W. Morando & Christian H. Nadan, *Silent Enemies*, RECORDER, May 4, 1994, at 10.

107. See, e.g., Lemley & Moore, *supra* note 81.

would not be willful infringers, though the common practice of marking products with patent numbers would make this a smaller set of cases.¹⁰⁸ Our anecdotal experience in litigation suggests that a significant number, perhaps even a majority, of all patent cases involve independent development. In none of these cases would the accused infringer have to obtain an opinion of counsel. Therefore none of these accused infringers would have to worry about skewed pre- or post-litigation advice, traps for the unwary, or limiting their choice of counsel.

This narrower definition of willfulness reduces but does not eliminate the incentive for clients to avoid reading patents. A company developing a new product would still have reason to avoid learning of patents that might cover that new product because running across such a patent would trigger their obligation to obtain a written opinion of counsel. A company's incentive to avoid reading patents would be somewhat lessened, however, because the consequences of doing so would be reduced. The company could be liable for willful infringement only for new products adopted with knowledge of the patent, not for existing products that it later happens to discover are infringing. Further, the costs of reading patents would be less than under the present system, since the scope of the privilege waiver would be substantially reduced. Nevertheless a significant number of companies would likely continue to be deterred from having their engineers read patents by the prospect of having to pay up to treble damages if found liable of willful infringement, coupled with the knowledge that in most cases the only viable way to fend off a finding of willful infringement is to obtain an opinion and then waive privilege.

That outcome not only frustrates the disclosure goal of patent law, but it also potentially reduces the total compensation paid to inventors. As noted above, there are many situations in which it is rational for a product maker to want to license an invention and for the inventor to grant a license. Such a mutually beneficial transaction is less likely to occur, however, if the would-be producer remains unaware of, and thus unable to request a license to, the invention because she is busily not reading patents. Accordingly, in the next part we return to the question whether, in addition to adjusting the temporal dimension of the willfulness inquiry, it is also necessary to adjust the financial dimension.

108. Marking or some other form of notice to the accused infringer is necessary to recover damages in the ordinary case. 35 U.S.C. § 287(a) (2000). As a result, most patentees mark their products.

E. Adjusting the Willfulness Premium

Current law permits enhancing damages upon finding willfulness but does not require it.¹⁰⁹ Indeed, the Federal Circuit has held that district courts properly exercise their discretion by refusing to enhance damages for willfulness where the defendant's conduct does not look particularly culpable, even where the jury finds the conduct to be willful.¹¹⁰ Because the definition of willfulness sweeps so broadly, courts must determine whether even willful infringers have in fact behaved reprehensibly. Narrowing the definition of willfulness means that those who are found to be willful infringers would be truly culpable parties that have intentionally copied an invention without a good faith belief that they had a right to do so. Willfulness so defined ought to be deterred, and its victims ought to be made whole. Thus, were the courts to define willfulness to include copying with knowledge of a patent, we think it should compel a finding of enhanced damages. The question is: By how much should damages be enhanced?¹¹¹

Our conceptual answer is: By the least amount possible needed to deter deliberate copyists and make the victims whole. We reach this answer because we believe first, that there are substantial negative consequences to over-deterrence and second, that there are significant deterrents to deliberate copying other than a willfulness penalty. As to the former, over-deterrence works against the disclosure goal of patent law by discouraging legitimate efforts to design around a patent at the margin.¹¹² It also discourages engineers from reading patents, which not only reduces the spillover benefits of invention¹¹³ but tends to reduce the number of mutually-

109. See, e.g., *Stryker Corp. v. Davol Inc.*, 234 F.3d 1252, 1259 (Fed. Cir. 2000); *Cybor Corp. v. FAS Techs.*, 138 F.3d 1448, 1461 (Fed. Cir. 1998) (en banc).

110. See, e.g., *Odetics, Inc. v. Storage Tech. Corp.*, 185 F.3d 1259, 1274 (Fed. Cir. 1999).

111. Under current law, courts consider deliberate copying, bad faith litigation conduct, size and financial condition, and the closeness of the legal issues as relevant factors in deciding how much damages should be enhanced. See *Amsted Indus. v. Buckeye Steel Castings Co.*, 24 F.3d 178, 183-84 (Fed. Cir. 1994).

112. On the economic benefits of design-arounds, see, for example, Conigliaro et al., *supra* note 101 and cites therein.

113. On the economic benefits of spillovers from "leaky" intellectual property rights, see, for example, Dietmar Harhoff, *R&D Spillovers, Technological Proximity, and Productivity Growth—Evidence from German Panel Data*, 52 SCHMALENBACH BUS. REV. 238 (2000). See also RUSLAN LUKACH & JOSEPH PLASMANS, MEASURING KNOWLEDGE SPILLOVERS USING PATENT CITATIONS: EVIDENCE FROM THE BELGIAN FIRM'S DATA (CESifo, Working Paper No. 754, 2002). Sector-specific productivity is directly and positively related to the level of spillover. Harhoff, *supra*, at 238.

beneficial voluntary licensing transactions, thereby hurting patentees as well. As to the latter, deliberate copyists will, on the whole, fare worse on the merits of their liability cases, will have to pay their lawyers substantial fees in addition to any damages award, and will be subject to a permanent injunction that will deprive them of return on their investment in infringing physical plant.¹¹⁴

As noted, the principal basis for imposing a willfulness penalty in the first place is to make sure that intentional copyists wind up worse off than if they had not copied, and that the patentees whose work was copied wind up no worse off. Even if the willfulness penalty achieves these goals when imposed, intentional copyists will, as a systemic matter, have a slight incentive to copy, because the penalty will not always be imposed. Some copyists may escape detection by the patentee and the system will necessarily decide some cases incorrectly, thereby allowing a stray copyist here and there to escape. Accordingly, the willfulness penalty ("WP") should be set as follows: $WP = ((\text{gains from infringement}) \times (1/\text{probability of detection})) + 1$.¹¹⁵

We suggest that, for most cases, awarding successful willfulness plaintiffs their attorney fees will serve as an adequate willfulness penalty.¹¹⁶ An

114. Some infringers—even deliberate copyists—may be able to escape an injunction in rare cases if their infringing product serves some important social purposes. *See Vitamin Technologists, Inc. v. Wis. Alumni Research Found.*, 146 F.2d 941, 944-47 (9th Cir. 1944) (refusing to grant injunction against cheap method of enhancing food with vitamin D); *Milwaukee v. Activated Sludge, Inc.*, 69 F.2d 577, 593 (7th Cir. 1934) (refusing to grant injunction that would require closing the City of Milwaukee's sewage treatment plant). Those cases are the exception rather than the rule, however, and are most likely to be applied where the patentee does not itself manufacture the patented product. *Foster v. Am. Mach. & Foundry Co.*, 492 F.2d 1317, 1324 (2d Cir. 1974).

115. Thus, if the gains from infringement were \$1,000,000, and there was a 0.2 probability of detection, the willfulness penalty would be \$5,000,001; if the chance of detection rose to 0.9, the willfulness penalty would be only \$1,111,112.

116. We say "most cases" because in cases in which detection is particularly unlikely, the penalty should be increased. Cases involving process patents or obvious efforts by the infringer to conceal its actions, such as by marking its infringing goods with the patentee's patent number and otherwise trying to pass off its goods as if they were the patentee's, would fall into this category.

It is possible that the trebling of current willfulness law corresponds, intentionally or otherwise, to the treble damages of the Clayton Act. The measure of damages in the Clayton Act was selected on the theory that, because antitrust conspiracies are hard to detect, the system needs to punish severely those that are detected. *Cf. United States v. Rabinowich*, 238 U.S. 78, 88 (1915). The Supreme Court in *Rabinowich* stated:

For two or more to confederate and combine together to commit or cause to be committed a breach of the criminal laws is an offense of the gravest character, sometimes quite outweighing, in injury to the public,

award of attorney fees will help reduce the likelihood that the patentee is undercompensated, and given that fees can be nontrivial, provide a modicum of monetary deterrence against copying. Of course, providing any willfulness penalty does not completely end the willfulness game, and therefore does not maximize the disclosure value of patents. Companies considering designing around a patent may still want to obtain and disclose an opinion of counsel in order to avoid paying attorney fees should they lose a subsequent patent suit. But simultaneously refocusing the willfulness inquiry on the time of adoption (thereby greatly limiting the scope of and costs imposed by the privilege waiver), and reducing the penalty itself would significantly increase the likelihood that companies will allow their engineers to read patents, and somewhat increase the likelihood that they forego the charade of opinion-getting altogether.

Patent law already provides for an award of attorneys fees in “exceptional cases.”¹¹⁷ It might seem odd as a matter of statutory interpretation to interpret two different statutory sections, one dealing with enhanced damages and the other with attorneys’ fees, to provide the same remedy. But the oddity can be resolved by focusing on the purpose of the two different sections—to deter subjectively willful misconduct in the case of § 284, and to increase remedies in cases that lacked an objectively reasonable defense in the case of § 285. These cases might overlap—an infringer might both be improperly motivated and lack a substantial defense—but they are analytically distinct.¹¹⁸

VI. CONCLUSION

In an effort to deter willful infringers, patent law has developed a bizarre complex of rules that has unintended consequences far outweighing the good of the rules themselves. Companies forced into the willfulness game must give up their legal rights and spend substantial amounts of money to defeat a charge of willful infringement. Imposing this burden distorts legal advice, impedes settlement, affects the choice of legal coun-

the mere commission of the contemplated crime. It involves deliberate plotting to subvert the laws, educating and preparing the conspirators for further and habitual criminal practices. And it is characterized by secrecy, rendering it difficult of detection, requiring more time for its discovery, and adding to the importance of punishing it when discovered.

Id. There is no similar problem in detecting most patent infringement, however.

117. 35 U.S.C. § 285 (2000).

118. Alternatively, as we explained above, § 284 could be interpreted to achieve a different purpose altogether—to correct for inadequate compensatory damages awards.

sel, creates traps for the unsophisticated, and undermines the disclosure function of the patent system. Solving these problems requires the courts to end, if not the willfulness doctrine itself, at least the game-playing that surrounds it. Courts can begin by undoing the perverse rules that compel an accused infringer to obtain a written opinion of counsel and, having obtained that opinion, to disclose it. But eliminating these rules won't do away with the heart of the problem. That requires reconceptualizing willfulness based on the understanding that it is the copying of an invention, not merely competitive business conduct, that the law is designed to target. A narrower willfulness doctrine with a more limited financial penalty attached to it will more faithfully serve the purposes of patent law, and put an end to a longstanding and dangerous game.

THE INFORMATION SEMICOMMONS

By Robert A. Heverly[†]

ABSTRACT

We think of information as property; law and economic structures, we argue, make it so. But this should not be the end of our inquiry. If we believe information is property, we must ask: What *kind* of property is information? While at times common use of information, even privately owned information, is accepted under the guise of fair use, private ownership of information forms the bedrock of our understanding of the information property regime. Common uses, when allowed, are perceived as infringements on the private owner's property rights.

This Article argues that this perception is mistaken: we may think of information as owned, but ownership need not be based in a purely private ownership scheme. Information ownership is instead a semicommons, a property model that explicitly recognizes the dynamic relationship and interdependence of private and common property uses.

Using the example of peer-to-peer file sharing, this Article applies semicommons theory to information, and shows that this application has broad implications for decisionmaking in the digital age. Through the lens of the semicommons, we realize that common uses are part of the very structure of an accurately described information ownership regime. As such, common uses increase the overall societal benefits that flow from information creation and should not only be tolerated, but encouraged.

Information is not a private property regime: it is a semicommons.

TABLE OF CONTENTS

I.	INTRODUCTION	1128
II.	TECHNOLOGICAL TENSION: PEER-TO-PEER FILE SHARING.....	1132
III.	THE CURRENT DEBATE: THE TENSION BETWEEN PRIVATE AND COMMON USES .	1139

© 2003 Robert A. Heverly

[†] Lecturer in Law and Director of the LL.M. Programme in Intellectual Property Law at the Norwich Law School, University of East Anglia, England. The majority of work on this Article was completed while the author was a Resident Fellow at the Information Society Project at Yale Law School. The following people have provided invaluable assistance and comments on this Article: Henry Smith, Jack Balkin, Caio Mario da Silva Pereira Neto, Nimrod Kozlovski, Shlomit Wagman, James Grimmelman, Markus Gehring, participants in the Graduate Workshop at Yale Law School, and participants at a Faculty Development Workshop at Albany Law School. Robert received his LL.M. from Yale Law School and his J.D. from Albany Law School of Union University.

IV. THE INFORMATION OWNERSHIP QUESTION	1147
A. Information.....	1147
B. Owning Information: A Framework for Inquiry	1151
C. Owned Information	1154
1. <i>Introduction</i>	1154
2. <i>Property and Rights in Brief</i>	1155
3. <i>Information Ownership</i>	1156
4. <i>Conclusion</i>	1157
D. Economics of Information.....	1157
E. The Nature of Information	1158
F. Final Thoughts.....	1161
V. INFORMATION OWNERSHIP AS A SEMICOMMONS.....	1161
A. Semicommons Theory.....	1161
B. The Information Semicommons.....	1164
1. <i>Private and Common: Information Use Dynamics</i>	1165
2. <i>Information "Creation" and Information Dynamics</i>	1167
3. <i>Strategic Behavior in the Information Semicommons</i>	1172
a) Identifying Strategic Behavior in the Information Semicommons	1172
b) Avoiding Strategic Behavior	1178
c) The Digital Revolution and Strategic Behavior.....	1181
C. Conclusion.....	1183
VI. THE SEMICOMMONS EFFECT.....	1183
A. Introduction.....	1183
B. The Normative Effect of the Semicommons Approach: Acknowledging Reality	1184
C. Peer-to-Peer Revisited.....	1185
D. Using Content and Distribution Level Interactions in Decisionmaking	1187
VII. CONCLUSION	1188

I. INTRODUCTION

When the curtains open, information is at center stage. Economically, in terms of speech, creativity, entertainment, and public discourse, information is the focus of much of life today. Our responses to the increased importance of information have predictably been shaped by our experience.

Specifically, we believe that information creators and distributors, the holders of copyrights, patents, trademarks and trade secrets, *own* the information they have created or seek to protect. Statutory provisions and judicial interpretations support this belief. In the modern case, information ownership is viewed as something akin to the ownership of land—as an asset to be used efficiently. Within the legal and economic structures that define the ownership of information, we treat information as property.¹

1. This Article does not advocate the position that information is *in fact* property, but rather that through our perception of it as such, we have failed to go further and ask the questions necessary to tell us what kind of property information should be. Those

Our conception of private ownership of information may be subject to some caveats,² but the general and overarching theme of information ownership is the private use and control of information.

However, when discussing the ownership of information, we generally fail to ask what *kind* of ownership applies to information. This question is particularly relevant to peer-to-peer file trading applications: networked systems that allow users to exchange digital media files, the content of which is held as property by someone else.³ Control over distribution—also known as the right to exclude others from using or distributing—is one of the key rights that private users have over their property.⁴ Peer-to-peer file sharing appears to shred this right, providing an apparent technological end-run around the private rights of copyright owners.

It should thus not be surprising that lawsuits followed the widespread deployment of peer-to-peer technology.⁵ The lawsuits were based on, and decided upon, our perception that information is property. On the whole, the legal disputes show how firmly ingrained the private notion of information ownership is in our legal, cultural, and societal psyche. The plaintiffs argued that peer-to-peer users were stealing their property, whereas users argued that they were making legally acceptable use of the property. The courts decided the cases based on the private use rights of the information owners.

This fact must not be understated: the legal decisions issued by courts involving peer-to-peer applications have been based on a conception of information as private property.⁶ As a result, courts have shut down poten-

critical of the notion that information is property will quite expectedly reject the notion that information is a particular kind of property. Thus, the basic argument—that information is property—is described, but not stridently defended, so as to reach the important question of the nature of information as property when we perceive it as such.

2. Such caveats might include, for example, fair use. *See* 17 U.S.C. § 107 (2000).

3. Peer-to-peer file trading systems are used here as one particularly trenchant example of the problems inherent in our current perception of information as property, but they are by no means the only one.

4. *See, e.g.*, Thomas W. Merrill, *Property and the Right to Exclude*, 77 NEB. L. REV. 730 (1998) (arguing that the right to exclude others is the most important of all property rights).

5. *See, e.g.*, *Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd.*, 259 F. Supp. 2d 1029 (C.D. Cal. 2003); *In re Aimster Copyright Litig.*, 252 F. Supp. 2d 634 (N.D. Ill. 2002); *A&M Records, Inc. v. Napster, Inc.*, 114 F. Supp. 2d 896, 913 (N.D. Cal. 2000) [hereinafter *Napster I*], *aff'd in part, rev'd in part*, 239 F.3d 1004 (9th Cir. 2001) [hereinafter *Napster II*], *remanded to No. C 99-05183*, 2001 U.S. Dist. LEXIS 2186 (C.D. Cal. Mar. 5 2001) [hereinafter *Napster III*], *aff'd*, 284 F.3d 1091 (9th Cir. 2002) [hereinafter *Napster IV*]; *see also infra* notes 19-23 and accompanying text.

6. *See* cases cited *supra* note 5.

tially liberating, freeing, equalizing technological applications that take full advantage of the decentralized nature of networks such as the Internet.

But this Article does not argue that such technologies should be spared solely because of alternative or empowering uses that are not currently at the core of their operations. Rather, it argues that the trading of copyright protected files via peer-to-peer networks is itself a valid reflection of the nature of the information ownership regime. Indeed, it argues that such use of peer-to-peer systems should be accepted and endorsed by the legal system under certain circumstances—some of which are reflected in current doctrine.

The implications of such an argument are broad and far-reaching. They would require the consideration of new policy implications when deciding whether to extend private uses and protections in information, affect judicial decisionmaking relating to use of existing information, and—as in the case we will study here—have implications for common use of information that is subject to copyright protection. Trademark, copyright, trade secret, and patent law would all be affected, as would all the arenas of economic, political, and legal decisionmaking that touch information.

To sustain such claims, to support the call for such a shift in paradigm, this Article makes a number of descriptive claims relating to information and information ownership. The positive claims begin with the idea that judicial decisions ordering peer-to-peer software to preclude the sharing of protected files, or holding software creators liable for infringement that occurs using their products and supporting networks, are based on the understanding of intellectual property as *private property*, with common uses viewed as infringements on the rights of the property owners. Even fair use is viewed within the structure of property as a defense to infringement; use is not a right per se of common users, but rather can be advanced as a defense once infringement has been proven or acknowledged. Fair use is measured against the private exploitation rights of the information owner.

This view of information as private property, of fair use as a perhaps barely tolerable infringement on private users' rights, though widely accepted, is incorrect. Information ownership can better be described as a semicommons,⁷ a form of ownership that acknowledges the dynamic rela-

7. See Henry E. Smith, *Semicommon Property Rights and Scattering in the Open Fields*, 29 J. LEGAL STUD. 131 (2000). Semicommons theory has much to offer in this regard, as a way of explaining how private and common uses of information might co-exist, in contrast to what in the past has been portrayed as a primarily antagonistic relationship.

tionship between private and common uses. Within such a conception, various novel arguments should be accepted in defense of peer-to-peer networks and other technological developments that in the current model would be discarded. For example, in the case of peer-to-peer networks, users often argue that peer-to-peer network sharing of copyright protected files increases the total number of sales of music and video products.⁸ In the current “information as private property” regime, such claims are only peripherally relevant to our analysis of infringement, even if empirically proven.⁹ However, under a scheme informed by an understanding of information ownership as a semicommons, such claims would be directly relevant to the question of whether uses were fair, and whether infringement had occurred or was facilitated by a particular technology.

The purpose of this Article is to establish that, within our perception of information as owned, such ownership is not purely private, but is instead a semicommons, and to examine further the implications of this conclusion. Part II begins the Article in earnest, discussing the peer-to-peer networking arena as an example of the private property view of information ownership in practice, and laying out the issues involved, the arguments raised, and the prizes at stake.

Part III discusses more fully the current debate over information ownership and details the nature of the debate as one fraught by notions of what is termed “purely private” ownership. Part III also discusses the normative implications this private property paradigm has on the information ownership debate.

Part IV shows what no one seems to contest: within the relevant legal framework, information is viewed as something that can be—and is—owned. This section notes that the legal structures that lead to private ownership are based primarily on the perceived need to provide the incentives necessary to make information creation and distribution economically viable.¹⁰ It further lays the necessary foundation for a discussion of the dynamic relationship between common and private uses of information, and offers a description of the nature of information, both in market and non-market terms, showing that inherent elements of information—especially its intangibility, general substitutability,¹¹ and nature as a public

8. *See Napster II*, 239 F.3d at 1018.

9. *See id.*

10. The question of whether this perceived need is in fact accurate is beyond the scope of this Article.

11. Substitutability is used here not in the economic sense of a good that can be perfectly replaced with another good for market purposes, but rather to illustrate the notion that a user may accept different expressions of the same information based on his or her

good—pose problems not found in other kinds of property. These problems do not undermine the conclusion that information is viewed as property in current legal doctrines, but instead highlight the differences that must be considered when thinking about information as property.

Part V advances a new framework through which information ownership can be viewed: the semicommons. The hallmark of a semicommons is the explicit recognition of a dynamic relationship between the common and private uses of semicommons property, such that their coexistence maximizes wealth to an extent not possible under either a purely common or a purely private scheme. An integral part of the maximization equation is that the benefits derived from the existence of the semicommons property scheme must outweigh the costs of prohibiting strategic behavior. Various methods—based in the legal, technological, architectural, and normative arenas—are used to prevent strategic behavior in the semicommons. Part V first evaluates the application of the semicommons theory in the context of the open fields of Medieval England as developed by Henry Smith, and then explores how information can also be viewed as a semicommons. It then discusses methods for avoiding strategic behavior for the information semicommons.

Part VI explores the ramifications of viewing property as a semicommons. Semicommons theory provides greater explanatory power in the current debate over private versus common uses of information than do other available theories. Returning to the peer-to-peer file sharing, Part VI shows by example the effectiveness of the semicommons theory.

Part VII concludes that the semicommons theory can illuminate the issues and the stakes involved in making important normative choices in ways that simply accepting private property theories cannot, and can make a substantive contribution to the ongoing debate over information ownership.

II. TECHNOLOGICAL TENSION: PEER-TO-PEER FILE SHARING

The case study presented here involves one particular technology, the peer-to-peer file sharing application, that allows users to connect across the Internet for the purpose of sharing files of various types.¹² Internet us-

needs. That is, the information contained within the expression may be similar, and thus the varied expressions of it may all be acceptable to various users. *See infra* notes 51-53 and accompanying text.

12. Peer-to-peer software applications do not generally discriminate by file type, but in respect of file types are considered a neutral platform for exchange.

ers hailed peer-to-peer as a boon because of its facilitation of the direct transfer of files in a nearly costless fashion across a decentralized platform made up of the users themselves.¹³ Peer-to-peer technology promised users the ability to download files from and share files with other users. The act of sharing was at once seen as expressive—users chose files to share that they valued and thought others would value—and as contributing to a community of users.¹⁴ Peer-to-peer options increased, and users joined in increasing numbers.

Peer-to-peer applications also drew the attention of the owners of the intellectual property rights to the music and video files being shared. The owners sued to stop what they viewed as the illegal distribution of files via the peer-to-peer networks, but in place of suing the individual users who were making the files available,¹⁵ they sued the peer-to-peer networks themselves.¹⁶ In one notorious case, this led to the “closing” of the peer-to-peer application and the end of its usefulness to its users.¹⁷ Other peer-to-peer networks responded by locating outside of the United States (ostensibly to avoid the jurisdiction of U.S. courts), and designing networks over which they allegedly had little or no control.¹⁸

13. See, e.g., LAWRENCE LESSIG, *CODE AND OTHER LAWS OF CYBERSPACE* (1999) (stating that “[p]eer-to-peer is the next great thing for the Internet”) [hereinafter LESSIG, *CODE*]; see also PEER-TO-PEER: *HARNESSING THE POWER OF DISRUPTIVE TECHNOLOGIES* (Andy Oram ed., 2001); Philip Agre, *P2P and the Promise of Internet Equality*, 46 *COMM. ACM* 39 (2003).

14. See Kathy Bowrey & Matthew Rimmer, *Rip, Mix, Burn: The Politics of Peer-to-Peer and Copyright Law*, 7 *FIRST MONDAY* 8 (Aug. 2002), available at http://firstmonday.org/issues/issue7_8/bowrey/index.html; see also CLAY SHIRKY ET AL., 2001 *P2P NETWORKING OVERVIEW: THE EMERGENT P2P PLATFORM OF PRESENCE, IDENTITY, AND EDGE RESOURCES* (2001); Jintae Lee, *An End-User Perspective on File-Sharing Systems*, 46 *COMM. ACM* 49 (2003).

15. In 2003, rights owners started suing individual users as well. See *RIAA's Lawsuits Meet Surprised Targets*, *WASH. POST*, Sept. 10, 2003, at E1 (describing the Recording Industry Association of America's lawsuit strategy that targeted 261 individual users).

16. See *Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd.*, 259 F. Supp. 2d 1029, 1031 (C.D. Cal. 2003); *Napster I*, 114 F. Supp. 2d 896, 896 (N.D. Cal. 2000).

17. See *Napster IV*, 284 F.3d 1091, 1091 (9th Cir. 2002); see also Brenda Sandburg, *Federal Judge Plays Requiem for Napster, Shuts Down Site*, *LEGAL INTELLIGENCER*, July 13, 2001, at 4.

18. For example, a Dutch court recently held that Kazaa, another peer-to-peer application, was not in violation of copyright laws in operating under a model not substantially different from that of Napster. See *'Dutch Napster' Wins Reprieve*, *BBC NEWS*, Mar. 28, 2002, at <http://news.bbc.co.uk/1/hi/business/1899725.stm> (last visited Mar. 25, 2003); see also *Grokster*, 259 F. Supp. 2d 1029, 1046 (C.D. Cal. 2003) (finding Grokster and Kazaa not liable for contributory infringement).

Regardless of the nature of the peer-to-peer networks and the complaints raised about them, the legal action taken by those who saw themselves as the owners of the information being traded have brought what they view as an effective end to open peer-to-peer network applications.¹⁹ There are, in the information owners' views, no legitimate purposes for such applications other than to undermine ownership in information or, more specifically, to steal their property.²⁰ The courts have viewed the disputes as involving technology that facilitates theft of private property, or property in information created by federal statute.²¹

Therein lies the tension. On the one hand, peer-to-peer file sharing systems have the potential to empower individuals, to provide distribution networks for individual works, expressions, creations and beliefs, and to allow all of this to occur in a decentralized and end-user controlled system.²² At the same time, information creators, owners, distributors, and

19. "Open" here is an indication of the network's non-affiliation with any particular content. There is some sign that various content—for example, information—owners are planning to launch their own peer-to-peer networks that will allow for distribution of their own content. Such systems are significantly different from systems that allow the exchange of any type of file. See *Chasing the Bottleneck*, RED HERRING, Oct. 15, 2001, at <http://www.redherring.com/Article.aspx?f=articles%2farchive%2fmag%2fissue106%2fattach.xml> (reporting that these new peer-to-peer networks should be successful because they "borrow" end-user bandwidth and thus reduce costs for distributors).

20. See Compl. for Contributory and Vicarious Copyright Infringement, Violations of Cal. Civil Code Section 980(a)(2), and Unfair Competition at 12, *A&M Records, Inc. v. Napster*, 114 F. Supp. 2d 896 (N.D. Cal. 2000) (No. C99-5183-MHP) ("Napster is not developing a business around legitimate MP3 music files, but has chosen to build its business on large-scale piracy."), available at http://news.findlaw.com/hdocs/docs/napster/riaa/napster_complaint.pdf (last visited Dec. 7, 2003).

21. See, e.g., *Grand Upright Music, Ltd. v. Warner Bros. Records, Inc.*, 780 F. Supp. 182, 183 (S.D.N.Y. 1991). The court in *Grand Upright Music* stated:

'Thou shalt not steal' has been an admonition followed since the dawn of civilization. Unfortunately, in the modern world of business this admonition is not always followed. Indeed, the defendants in this action for copyright infringement would have this court believe that stealing is rampant in the music business and, for that reason, their conduct here should be excused. The conduct of the defendants herein, however, violates not only the Seventh Commandment, but also the copyright laws of this country.

Id. (citations omitted). But see *Dowling v. United States*, 473 U.S. 207 (1985) (holding that the shipping of "bootlegged" records did not constitute shipping stolen goods in violation of federal law).

22. See generally Yochai Benkler, *From Consumers to Users: Shifting the Deeper Structures of Regulation Towards Sustainable Commons and User Access*, 52 FED. COMM. L.J. 561 (2000) [hereinafter Benkler, *From Consumers to Users*] (arguing that

publishers accuse peer-to-peer networks of serving primarily to undermine their property rights.²³

The potential of the peer-to-peer design should be apparent: it enables users to connect with each other, allows small creative artists to distribute their works outside of the traditional channel of distribution controlled by large, powerful media corporations.²⁴ It also provides an avenue for distributing, sharing, or collaborating in text works, whether traditional novels, scientific research, or new categories of works.²⁵ The technology has the ability to enhance autonomy by giving individuals greater control over their own lives, creativity, and choices.

But in practice, the technology took a different turn. Users of early peer-to-peer file sharing networks chose to share two primary types of files: audio and video. In general, these files were converted or extracted from media files held by peer-to-peer participants.²⁶ Thus, users were distributing files on which others held enforceable copyrights.

“[t]he emergence of the digitally networked environment makes possible the development of a robust, open social conversation in which all can participate as peers”).

23. See, e.g., Howard P. Goldberg, Note, *A Proposal for an International Licensing Body to Combat File Sharing and Digital Copyright Infringement*, 8 B.U. J. SCI. & TECH. L. 272, 273 (2002) (“Computer users, through the digital file sharing of copyrighted works, are able to infringe upon a copyright holder’s exclusive rights.”). To combat peer-to-peer networks’ alleged encroachment on their rights, information owners have developed DRM systems. See generally Lionel S. Sobel, *DRM as an Enabler of Business Models: ISPs as Digital Retailers*, 18 BERKELEY TECH. L.J. 667 (2003) (describing DRM technologies and their implications). These systems are designed to preclude the easy copying of digital works and are themselves protected by a legal regime designed ultimately to ensure protection for the underlying creative works. See 17 U.S.C. § 1201 (2000). DRM systems combine software or code protection with hardware protection. Digital Video Disks (“DVDs”), for example, use a system called Content Scrambling System (“CSS”) to decode the video and audio data encrypted on the disk. See *Universal City Studios, Inc. v. Corley*, 273 F.3d 429, 436 (2d Cir. 2001). DVD player manufacturers must license the right to implement CSS on their systems, and in this way content manufacturers can ensure that media content is protected. See *id.* at 436-37.

24. See, e.g., HASSAN M. FATTAH, P2P: HOW PEER-TO-PEER TECHNOLOGY IS REVOLUTIONIZING THE WAY WE DO BUSINESS (2002); see also MICHAEL MILLER, DISCOVERING P2P (2001).

25. See *Upgrading the Internet*, ECONOMIST, Mar. 24, 2001, at 32. The newest incarnation of the Internet, known as Internet2, is designed to facilitate the same types of collaborative efforts that peer-to-peer applications also promised to bring. *Id.*

26. Audio files, for example, can be easily “extracted” from audio compact discs by computer software applications that convert the CD audio wave files to MP3 format. MP3 files are smaller in size than audio wave files, thus they are easier to send over networks and store on computer hard drives. Video files can either be “captured,” where users attach a VCR to their computer and “record” the analog output of the video as a

Initially, this led to very few problems, as peer-to-peer technology was flying largely under the radar of the major copyright holders.²⁷ However, the number of peer-to-peer end users and the number of technological options for file-sharing increased. In the late 1990s, artists and publishing companies began launching attacks on file-sharing and started public relations campaigns that appealed to users' ethical and legal consciences to refrain from copyright infringing activities.²⁸

Eventually, when the publishing and distributing industries, as well as some elements of the creative community,²⁹ felt they were losing the battle, or more particularly, losing business and sales to online peer-to-peer sharing, they began to bring lawsuits against the networks themselves. One of the earliest lawsuits sought an injunction against Napster, a successful peer-to-peer application, requesting that the court shut down the company's network operations.³⁰ The District Court issued the injunction.³¹

Napster, in danger of bankruptcy if the injunction was allowed to stand, appealed to the Ninth Circuit.³² As it had below, Napster argued that its users were not direct infringers, and, in the alternative, even if they were, Napster was not liable under copyright doctrine for assisting or failing to stop its users' infringing activities.³³ The Ninth Circuit, however, agreed largely with the analysis adopted by the District Court. Specifically, the court found that by trading copyrighted files over the Internet, Napster users were directly infringing the protected rights of the record companies.³⁴

digital file, or "ripped," where users extract digital files from DVDs or other forms of digital video.

27. See Matthew Fagin et al., *Beyond Napster: Using Antitrust Law to Advance and Enhance Online Music Distribution*, 8 B.U. J. SCI. & TECH. L. 451, 457-64 (2002) (detailing the background and history of peer-to-peer development and the content owners' reaction to it).

28. *Id.*

29. Not all artists felt that file-sharing was a negative. See Stephanie Green, *Reconciling Napster with the Sony Decision and Recent Amendments to Copyright Law*, 39 AM. BUS. L.J. 57, 65 (2001) ("Artists who support Napster see the potential for Napster's method of music distribution which gives more return to the artist directly by eliminating the middleman and the stranglehold of the big record labels.").

30. *Napster I*, 114 F. Supp. 2d 896, 896 (N.D. Cal. 2000).

31. *Id.* at 927.

32. *Napster II*, 239 F.3d 1004, 1004 (9th Cir. 2001).

33. *Id.*

34. *Id.* at 1014.

Napster raised the fair use defense on its users' behalf under three separate theories, all of which proved unsuccessful.³⁵ The first, or "sampling," theory was based on the notion that users download the music in an effort to decide whether to buy it.³⁶ The second, or "space-shifting," theory was based on the idea that users who already own a licensed copy of the recording would download the MP3 version from Napster so as to avoid the need to rip the audio track themselves.³⁷ The third, or "permissive distribution," theory was based on the idea that some artists and recording companies have decided to make their works available online and without the requirement of a purchased license, and thus redistribution is implied under the companies' chosen method of distribution.³⁸

The court rejected the first two of Napster's fair use theories, and found the third irrelevant for purposes of the record companies' action.³⁹ In rejecting the first two theories, the court reviewed each of the four statutory elements of fair use: the purpose and character of the use, the nature of the copyrighted work, the amount and substantiality of the portion used, and the effect of the use on the potential market for the work or the value of the work.⁴⁰

For our purposes, the Ninth Circuit's analysis of the last statutory factor is the most relevant. In discussing the effects of Napster users' activities on the market for the plaintiffs' works, the District Court had discounted but not excluded evidence offered by Napster in the form of an expert opinion that Napster users' activity was beneficial to the recording

35. *Id.* at 1019.

36. *Id.* at 1018.

37. *Id.* at 1019.

38. *Id.*

39. *Id.* (holding that the permissive distribution claim was irrelevant because the plaintiff record companies were not alleging that such permissive distribution violated their rights under copyright law, and thus any permissive distribution was not relevant to their claims).

40. *Id.* at 1014-15. The Ninth Circuit agreed with the District Court's determination that Napster users' activities militated against a finding of fair use under each of the four tests. *Id.* First, under the "purpose and character" test, the courts found that the nature of the users' activity was not transformative, but rather was a mere "retransmission [of the original works] in another form." *Id.* at 1015. Furthermore, although the users were generally not making money from each other, their use was commercial. *Id.* Second, both courts found that the works were creative and were thus the exact type of works that copyright is intended to protect. *Id.* at 1016. Third, users making complete copies of copyrighted sound recordings. *Id.* Finally, users were impairing the market as to sales of traditional sound recordings, and were also displacing the market for the copyright holders in the digital realm. *Id.* at 1016-17.

industry.⁴¹ Even without the derogation of the defendant's evidence, however, the Ninth Circuit noted that "lack of harm to an established market cannot deprive the copyright holder of the right to develop alternative markets for the works."⁴² Note that, in effect, this makes nearly *any* evidence of benefit to the market for an information product irrelevant where it occurs in a market in which the information owner has not already established itself, or possibly even where it has. In such a case, the common use that has been established prior to the private owner either seeing or taking advantage of a new market for its goods acts as a "barrier" to the owner's entering the new market.⁴³ In short, the court noted that benefits to the record companies from the common use would not be sufficient to bring the use within the fair use provisions of copyright law. According to the Ninth Circuit,

[t]he District Court . . . correctly noted that 'any potential enhancement of plaintiff's sales . . . would not tip the fair use analysis conclusively in favor of defendant.' We agree that increased sales of copyrighted material attributable to unauthorized use should not deprive the copyright holder of the right to license the material.⁴⁴

Having determined that the Napster users' infringing activities were not protected by the fair use defense, the Ninth Circuit upheld the District Court's finding that Napster was potentially liable for both contributory and vicarious copyright infringement. This finding was based on the fact that Napster knowingly provided the method by which files were traded and had some ability to police its own system which it refused to fully exercise. While the potential for liability was sufficient to uphold the preliminary injunction, the Ninth Circuit did modify the constraints of the injunction, though not in a way significant enough to allow Napster to continue to operate as it had in the past.

On remand, the District Court modified its injunction to bring it in line with the Ninth Circuit's decision, and required Napster to remove files from its search index where it received "reasonable notice" of the existence of infringing files. The District Court also adopted a "zero tolerance" requirement, whereby Napster's failure to stop all trading of files "owned" by others would not be excused. After three months of court-administered monitoring, the District Court found Napster in violation of the modified

41. *Id.* at 1017.

42. *Id.*

43. *Id.*

44. *Id.* at 1018.

order, and in accordance with its zero tolerance policy, directed it to shut down. The Ninth Circuit upheld the shutdown order on appeal, accepting explicitly the District Court's adoption of the zero tolerance requirement.

The Napster case is a clear example of courts and litigants viewing information protected by copyright as private property. Fair use is ultimately the use of *someone else's property*, a notion that colors and channels debate and legal decisions regarding use of information. Peer-to-peer software has the potential to invigorate and support a variety of new methods of production and distribution of digital goods, and in so doing to invigorate and empower creative people the world over. But because of how we view information ownership, because our conception is one that pits private users against common users, where private users are expected to seek compensation for each and every use of their "property" as such,⁴⁵ Napster was ultimately shut down by court order.⁴⁶ The only way that the Napster system could have continued in operation was if it was modified to operate according to the needs, not of its common users, but of the owners of the information being traded via its facilities. The District Court's adoption and the Ninth Circuit's approval of the zero tolerance requirement made this clear.

III. THE CURRENT DEBATE: THE TENSION BETWEEN PRIVATE AND COMMON USES

Over time, different types of goods and means of production have been the primary focus of market transactions and exchanges, and property law has often been the center of attention in terms of distribution of wealth.⁴⁷ Each era's economic engines have been fueled by different resources and production methods, and ownership of those resources and methods has been a political and cultural issue of great importance.⁴⁸

45. This expectation is not somehow an imposed or unreasonable expectation. Market based systems expect property owners to seek remuneration, but that expectation is not the reason for owners' actions. Markets base their expectations on owners' actual desires. In other words, owners seek remuneration because they desire remuneration, not because it is expected of them.

46. See *Napster IV*, 284 F.3d 1091, 1096 (9th Cir. 2002); see also Brian Garrity, *Media Giant's Support Of Napster 'Vicarious Infringement'*, BILLBOARD, Mar. 1, 2003, at 7 (reporting that one of Napster's funders, the Bertelsmann corporation of Germany, is now being sued by a group of music publishers who allege it is also vicariously liable for Napster's actions).

47. See, e.g., KARL MARX & FREDERICK ENGELS, *THE COMMUNIST MANIFESTO* (A.J.P. Taylor ed., Samuel Moore trans., 1967).

48. For example, Rousseau touches on this theme in his second Discourse on Inequality: "[A]ll the inequality which now prevails owes its strength and growth to the de-

In today's advanced capitalist economies, markets are becoming more and more focused on information creation and distribution. Information is the primary resource in the information economy; its production is the main focus of members of the information society and its ownership is the driving force in the creation and continuation of wealth.⁴⁹ It is unsurprising that the nature of the property scheme that underlies the shift to the new economy should become a subject of dispute, especially as its establishment and continued development are certain to affect the distribution of wealth in society, both in postindustrial nations and throughout the world.⁵⁰

There are a number of questions within this setting, however, that are different in nature and degree from those that have been asked in past generations about the ownership of resources and production methods. The nature of information is significantly different from the nature of past commodities: it is intangible, non-rival, and non-exclusive. In many cases, it is also substitutable: one piece of information may be "replaced" by another similar one.⁵¹ Given this, the legal schemes and policy discussions that surround information must also be different.⁵² With the growth over the past ten years of large-scale and nearly costless methods of informa-

velopment of our faculties and the advance of the human mind, and becomes at last permanent and legitimate by the establishment of property and laws." JEAN-JACQUES ROUSSEAU, *THE SOCIAL CONTRACT AND DISCOURSES* 238 (G.D.H. Cole trans., J.M. Dent 1913); see also CONSTANTIN SONIN, *PRIVATE PROTECTION OF PROPERTY RIGHTS, INEQUALITY, AND ECONOMIC GROWTH IN TRANSITION ECONOMIES* (Econometric Soc'y World Congress, Contributed Paper No. 1300, 2000); Timothy Frye & Andrei Shleifer, *The Invisible Hand and the Grabbing Hand*, 87 *AM. ECON. REV.* 354 (1997).

49. See TREVOR HAYWOOD, *INFO-RICH-INFO-POOR* (1995); MICHAEL W. HILL, *THE IMPACT OF INFORMATION ON SOCIETY* (1999); DAVID LYON, *THE INFORMATION SOCIETY* (1988); NAT'L ACAD. SCI., *THE DIGITAL DILEMMA: INTELLECTUAL PROPERTY IN THE INFORMATION AGE* (2000); DAVID WEINBERGER, *SMALL PIECES LOOSELY JOINED: A UNIFIED THEORY OF THE WEB* (2002).

50. See, e.g., COMM'N ON INTELL. PROP. RIGHTS, *INTEGRATING INTELLECTUAL PROPERTY AND DEVELOPMENT POLICY* 25-26 (2002), http://www.iprcommission.org/graphic/documents/final_report.htm (warning that developing countries should not adopt Western strength IP protection).

51. This does not mean information forms are substitutable in the economic sense, but rather that the meaning contained within information may be transmitted or continued without the need for or use of the original expression of that information.

52. See, e.g., LAWRENCE LESSIG, *THE FUTURE OF IDEAS: THE FATE OF THE COMMONS IN A CONNECTED WORLD* (2001) [hereinafter LESSIG, *FUTURE OF IDEAS*]; MICHAEL PERELMAN, *STEAL THIS IDEA* (2002); MARK ROSE, *AUTHORS AND OWNERS: THE INVENTION OF COPYRIGHT* (1993).

tion distribution, questions of ownership and control have dramatically increased in importance.⁵³

Within these debates, the ownership of information is often treated as a given.⁵⁴ According to this view, information is clearly subject to private ownership; information is property.⁵⁵ This perception pervades discourse on information issues. Efforts to define the importance of information policies accept the notion of private ownership of information, and seek to limit the extensions of this ownership within an effort that is consistent with past attempts to control privatization of public goods.⁵⁶

These efforts, and the overall ownership debate, are based too heavily on a conception of private property. By adopting this perspective, we focus too much attention on the burdens that common and private uses of information place on each other, while failing to acknowledge the significant benefits that result from the interaction between both types of use. The reason is clear: burdens and benefits resulting from the dynamic interaction between common and public information uses do not easily fit within traditional notions of private property. In a more traditional property scheme, the private property owner has the right and, indeed, the eco-

53. See JORGE REINA SCHEMENT & TERRY CURTIS, *TENDENCIES AND TENSIONS OF THE INFORMATION AGE: THE PRODUCTION AND DISTRIBUTION OF INFORMATION IN THE UNITED STATES* (1995).

54. See, e.g., Andrew Beckerman-Rodau, *Are Ideas Within the Traditional Definition of Property?: A Jurisprudential Analysis*, 47 ARK. L. REV. 603 (1994) (concluding that common law idea protection, trade secret law, and patent law are all property based regimes predicated on the recognition of property rights in ideas).

55. See, e.g., *United States v. Wang*, 898 F. Supp. 758 (D. Colo. 1995) (holding that, in contrast with the Supreme Court finding that the National Stolen Property Act did not apply to intangible property so as to bring copyright infringement within that Act, the federal wire fraud statute did apply, as it contained no requirement that the property in question be "tangible"); see also *United States v. Riggs*, 739 F. Supp. 414 (N.D. Ill. 1990) (holding that even if wire fraud statute required tangibility, information contained in a computerized list qualifies for that status).

56. For example, during debate on the Senate floor prior to adoption of the DMCA, Senator Strom Thurmond stated: "This legislation implements the WIPO treaties to help protect the property rights of the creative community in our global environment." 144 CONG. REC. S11888 (daily ed. Oct. 8, 1998) (statement of Sen. Hatch). Senator Hatch, during the same debate, stated, "The WIPO treaties and the DMCA will protect the property rights of Americans in their work as they move in the global, digital marketplace, and, by doing so, continue to encourage the creation of new works to inspire and delight us and to improve the quality of our lives." *Id.* at 11889.

conomic imperative, to undertake proper development. Private property is thus efficiency's answer to the tragedy of the commons.⁵⁷

Ostensibly responding to the need for a system of incentives to encourage the creation of information products, legislative enactments and court decisions have expanded the private use rights of information creators.⁵⁸ Information owners have praised these actions as providing further incentives for creation of information to fuel the information economy. Common use advocates have decried them as part of an attempt to "enclose" what is and should be the commons in the name of gaining profits.⁵⁹ The debate is ongoing, and has come to a head in lawsuits such as *Eldred v. Ashcroft*⁶⁰ in which the plaintiff challenged Congress's authority to extend the length of both prospective and retrospective protection for works covered under the copyright law.

On the one side of this debate are scholars, inventors, writers, and creators who question the establishment of further protection of individual ownership of works, inventions, and creations.⁶¹ On the other, those who believe privatization of information is simply another proper step in reaching the same economic efficiency for information as we have for other commodities.⁶² The common use literature makes the case not that "information longs to be free,"⁶³ but that it is necessary and important that

57. See Garrett Hardin, *The Tragedy of the Commons*, 162 SCI. 1243 (1968); see also ELINOR OSTROM, GOVERNING THE COMMONS: THE EVOLUTION OF INSTITUTIONS FOR COLLECTIVE ACTION (1990).

58. See, e.g., Digital Millennium Copyright Act ("DMCA"), Pub. L. No. 105-304, 112 Stat. 2860 (1998) (codified in scattered sections of 17 U.S.C., 28 U.S.C., and 35 U.S.C.); Sonny Bono Copyright Term Extension Act, Pub. L. No. 105-298, 112 Stat. 2827 (1998) (codified in scattered sections of 17 U.S.C.).

59. See James Boyle, *The Second Enclosure Movement and the Construction of the Public Domain*, Conference on the Public Domain (Nov. 2001), at <http://www.law.duke.edu/pd/papers/boyle.pdf> (last visited Feb. 24, 2003); see also David Lange, *Recognizing the Public Domain*, 44 LAW & CONTEMP. PROBS. 147, 150 (1981).

60. 537 U.S. 186 (2003).

61. See, e.g., JAMES BOYLE, SHAMANS, SOFTWARE AND SPLEENS: LAW AND THE CONSTRUCTION OF THE INFORMATION SOCIETY (1996); LESSIG, FUTURE OF IDEAS, *supra* note 52; Yochai Benkler, *Through the Looking Glass: Alice and the Constitutional Foundations of the Public Domain*, 66 LAW & CONTEMP. PROBS. 173 (2003); Jed Rubenfeld, *The Freedom of Imagination: Copyright's Constitutionality*, 112 YALE L.J. 1 (2002).

62. See, Brief of Amici Curiae American Intellectual Property Law Association at 10-16, *Eldred v. Ashcroft*, 537 U.S. 186 (2003) (No. 01-618) (arguing that the Sonny Bono Copyright Term Extension Act provides authors the benefit of their copyright bargain and encourages the preservation, dissemination and exploitation of older, valuable works).

63. This is one of the early cries of hackers and other activists on the Internet, and is more rhetorical hyperbole than actual argument.

current and future generations have access to and use of the store of knowledge, and even particular expressions of that knowledge.⁶⁴ Those who view themselves as the owners of particular expressions counter that these things are their property, and in line with the concepts of property, it is entirely appropriate to provide extra, extended, or greater protection where doing so further extends the perceived economic benefits of the creative process.⁶⁵

Two macro-level perspectives are generally identifiable for information ownership: one that focuses on private ownership (and thus private control) of information in models drawn from property theory,⁶⁶ and one that focuses on common ownership (and thus common control) of information.⁶⁷ The paradigm is well established that common use exists at the expense of private use. In fact, the need for private ownership of the commons grows out of well established property conceptions. The two sides here could even be described as battling over information ownership, each setting out essentially opposite positions designed to show how the engagement of the legal structures in certain ways will affect societal and market development, as well as individual autonomy, freedom, and satisfaction. The positions are, for the most part, diametrically opposed: private use proponents advocate more private control, longer ownership terms, and more private rights in the information they claim, while common use proponents advocate less private control, shorter ownership terms, and more common use rights in information generally.

This is not an indication that each side seeks the “end” of the other. There is some acknowledgment in most of the common-advocacy literature that *some* private incentives might be necessary for information creation to occur, and in the private-advocacy literature that *some* common use is acceptable. But on the whole, the two sides are in conflict over the balance that is to be struck between the uses, and the uses themselves are viewed as antagonistic.

In other words, these debates are based on the notion that common and private uses of information are inherently and primarily conflicting. For example, providing longer lengths of time to copyright protection for information creators removes important information from the information commons. One use “takes from” the other. There is little acknowledgment

64. See, e.g., LESSIG, *FUTURE OF IDEAS*, *supra* note 52.

65. See Brief of Amici Curiae Recording Artists Coalition, *Eldred v. Ashcroft*, 537 U.S. 186 (2003) (No. 01-618) (arguing that the term limits in the Sonny Bono Act provide incentives both to current artists and to their heirs).

66. See Becker-man-Rodau, *supra* note 54.

67. See Benkler, *From Consumers to Users*, *supra* note 22.

or explicit discussion of the dynamic interaction of private and common uses. Instead, the uses are at war; gain by one is its pure gain and the other's pure loss.

This framework of antagonism is entirely consistent with historical private use versus common use discussions. Where land was used for grazing in the traditional commons, private property displaced the commons and sought to make more efficient and sustainable use of it.⁶⁸ The need to strike a proper balance between common and private uses is a recurring theme, and the current debate seems to fall neatly into existing paradigms of property and use rights.⁶⁹ In fact, the paradigm appears to be so well established that some scholars have begun to attempt to chip away at the use of property metaphors in the information arena in the hopes of recasting the argument in a different light.⁷⁰

Private use arguments are most often based on the perceived need to enhance existing incentives to create information and, more precisely, to provide greater economic advantages for engaging in such creation. While the position is also sometimes grounded in Lockean notions of property rights flowing from individual effort, the economic incentive arguments generally carry the majority of the weight in policy and legal decision-making.⁷¹ Viewed in any light, the private use perspective on information ownership is based on traditional notions of real and personal property. In other words, it is based on ownership of some "thing."⁷²

68. See Hardin, *supra* note 57, at 1243-48.

69. See Boyle, *supra* note 59.

70. See, e.g., Dan Hunter, *Cyberspace as Place and the Tragedy of the Digital Anticommons*, 91 CALIF. L. REV. 439 (2003) (arguing that the "place" metaphor used in discussing and referring to the Internet has colored our perceptions and led us to accept property based conceptions of information, and raising important issues as to the way in which metaphors affect law); see also STEPHEN L. WINTER, *A CLEARING IN THE FOREST: LAW, LIFE, AND MIND* (2001).

71. There is an important distinction to be made here between using Lockean notions when seeking legislative and judicial changes and using Lockean notions in relation to the rhetoric of public education about rights in information. The large industrialized apparatus that makes up the core of the new information infrastructure generally relies heavily on notions of efficiency and incentive in arguing for more private control over information, but sings a more moral tune when publicizing the rights it has obtained in the legislative and judicial fora.

72. What a "thing" is in this regard is an interesting question. If it is a physical thing, then intellectual property is not property at all. But many of us would agree that a sound is a "thing," though it is intangible and temporary. For our purposes here, we will include both the tangible and intangible as "things." See Henry Lanz, *The Thing-Conception in the Middle Ages*, 29 PHIL. REV. 226, 226 (1920) ("Generally speaking, there are as many thing-conceptions as there are types of civilisation and culture; every

For their part, common use advocates dispute not only the proper amount of private protection needed to properly stimulate the creation of information, but also often question the idea that incentives are necessary at all.⁷³ Even when begrudgingly accepting the private incentives argument, common use advocates seek greater liberty for users of information, including the rights to alter, to derive new works from existing ones, and to comment on and critique the works of others through satire or parody. The concept of the public domain or of the commons in information is illustrative of this point.

epoch has a theory of objectivity of its own.”); R.W. Sellars, *A Thing and its Properties*, 12 J. PHIL. PSYCHOL. & SCI. METHODS 318, 318 (1915) (“Things, for common sense, are perceived or at least capable of being perceived and are known to be independent of the perceiver so long as he does not operate upon them by means of his body.”); Leslie Stevenson, *On What Sorts of Thing There Are*, 85 MIND 503 (1976) (laying out some of the ontological complexities of discussions of things). That intangibility is not problematic to property theory is not without disagreement; see, e.g., J.E. PENNER, *THE IDEA OF PROPERTY IN LAW* (1990) (arguing that “thingness” is particularly important to property theory).

73. This is not necessarily a new idea. In the 1830s, William Leggett wrote: If the principle of copyright were wholly done away, the business of authorship, we are inclined to think, would readily accommodate itself to the change of circumstances, and would be more extensively pursued, and with more advantage to all concerned than is the case at present. It is very much the fashion of the day to deride and decry cheap publications. We are not of the number who can join in the censure. The great good which the invention of printing originally effected, was to diffuse literature, and make books accessible to myriads, who were precluded from them before, by reason of the enormous prices at which manuscript copies were sold. . . . It is spreading literature over the entire land. . . . It is penetrating with it into every nook and corner of society. It is offering its golden fruits, ay, richer than gold, to the poor and ignorant, as well as to the rich and educated. It is awakening millions of human beings to a sense of their birthright; to acknowledge that they are God’s creatures, and not beasts that perish. . . . If there were no copyright laws, all literature would take a cheap form, and all men would become readers. . . . Instead of an edition of two, or three, or five thousand copies, which never constitute, as a general rule, the maximum of a popular author’s success, twenty, thirty, and perhaps a hundred thousand would be readily disposed of. Let us withdraw our attention, for an instant, from a contemplation of the interests of authors, to consider those of mankind at large. Who can fail to see how vastly the general benefit would be promoted? . . . [I]f knowledge is power, what a vast influence it would exercise in the counsels of nations, and in directing the destinies of mankind!

WILLIAM LEGGETT, *DEMOCRATICK EDITORIALS: ESSAYS IN JACKSONIAN POLITICAL ECONOMY* 394-95 (Lawrence H. White ed., 1984) (essays originally published beginning in 1834).

Just as the concept of information as subject to private ownership has been bound up in land and physical property analysis, so has the common use view of information. Discussions of the “information commons” are rife in common use literature, though the literature does not uniformly define the public domain. One of the most accepted definitions of the public domain is information that is available for use without restriction,⁷⁴ while another expands the definition to “the range of uses of information that any person is privileged to make absent individualized facts that make a particular use by a particular person unprivileged.”⁷⁵ Another definition would allow for information to be considered as part of the commons, even if there is a fee, so long as the fee is applied without discrimination.⁷⁶

Only recently has the perception of the information commons begun to move from an asset-based understanding to a use-based understanding.⁷⁷ The shift is important in that it coincides with the argument made here that information ownership is not based in traditional regimes of asset-based property rules that provide protection for an entire “thing,” but rather is based on the notion of use of information. While the newer understanding is clearly more in line with the theory of the information semicommons, it is still based on a primarily antagonistic relationship between the private and common uses, and in that way continues the “information as private property” conception that has been the primary theory of information ownership so far.

Within the debate over the balance between private and common uses of information, it is clear that both sides are constrained within the private use/common use dichotomy that has evolved in the landed property arena. While some commentators question the validity of the application of property principles to information due to the differences in the nature of infor-

74. See Jessica Litman, *The Public Domain*, 39 EMORY L.J. 965, 968 (1990).

75. See Yochai Benkler, *Free as the Air to Common Use*, 74 N.Y.U. L. REV. 354, 362 (1999) [hereinafter Benkler, *Free as the Air*]. Note the difference between the Litman and Benkler definitions. Litman’s definition is more narrow, in that it only applies to information that is available for use by *all* people, and thus includes only unprotected information as a whole. Benkler’s definition includes information that *any* person can use, and thus includes information subject to protection, but which may also be subject to common use. Benkler’s definition includes uses protected by fair use and other copyright defenses, while Litman’s does not. For purposes of the semicommons, Benkler’s definition is closer to our understanding of the nature of information in that it recognizes explicitly that both private and common uses of information can occur contemporaneously.

76. See LESSIG, *FUTURE OF IDEAS*, *supra* note 52.

77. See Benkler, *Free as the Air*, *supra* note 75.

mation and property,⁷⁸ the point of this Article is not that the property paradigm within which the discussion is taking place is faulty in treating information as property,⁷⁹ but that it is too limited in its scope.

Our understanding of property need not be limited to a choice *between* private and common uses; it can be both at the same time. This is not true in every case, and it need not always be true in relation to all property. But when it is true, or when it must be true—as we shall see is the case with information ownership—the decision-making process in relation to it must be different from that relating to property that is subject either to private or common use, or that is subject to a tug-of-war between the two uses. This Article details the extent of this difference, along with its implications.

IV. THE INFORMATION OWNERSHIP QUESTION

A. Information

Before delving into the structure of information ownership, we must first define what is meant by *information*. The literature on intellectual property and property rights in information bypass this question entirely, even when it is apparent that the notion of information—broad or narrow—could have a distinct effect on the analysis undertaken and the conclusions reached.⁸⁰

Traditional notions of information include the concept of knowledge. For example, the Oxford English Dictionary defines information as “[k]nowledge communicated concerning some particular fact, subject, or event; that of which one is apprised or told.”⁸¹ Similarly, the American Heritage Dictionary of the English Language defines information as “[k]nowledge derived from study, experience, or instruction.”⁸²

78. See Pamela Samuelson, *Information as Property: Do Ruckelshaus and Carpenter Signal a Changing Direction in Intellectual Property Law?*, 38 CATH. U. L. REV. 365 (1989) (examining the change in paradigm from one that views information as not covered by concepts of property to one that views information as covered by concepts of property); see also Douglas G. Baird, *Common Law Intellectual Property and the Legacy of International News Service v. Associated Press*, 50 U. CHI. L. REV. 411 (1983) (arguing that freedom of information is commonplace for intellectual property law).

79. On this question the author need not, and does not, express any opinion.

80. See, e.g., Benkler, *Free as the Air*, *supra* note 75 (arguing that enclosure of the public domain in information is occurring to the detriment of First Amendment principles, but not defining information).

81. OXFORD ENGLISH DICTIONARY 944 (2d ed. 1989).

82. AMERICAN HERITAGE DICTIONARY OF THE ENGLISH LANGUAGE 899 (4th ed. 2000).

Not every conception of information, however, focuses on knowledge. In 1948, Claude Shannon adopted a definition for information designed to aid the mathematical understanding of information transmission in a communicative system.⁸³ For Shannon, information is part of a mathematical function of uncertainty and is defined as anything that lowers uncertainty.⁸⁴ This definition is quantitative in nature: Shannon was concerned with *how much* information was transferred in communication. This definition is most relevant to disciplines concerned with communication of information rates.

Richard Losee went beyond Shannon's definition and attempted to develop a discipline-independent definition of information.⁸⁵ Losee counseled that we must be careful not to be overly restrictive in our initial definition of information and warned against imbuing a definition of information with requirements of understanding, knowledge, lack of repetition, or other more particular requirements.⁸⁶ While such requirements may be useful in inquiring how knowledge is developed, transmitted or understood, they mostly serve to artificially restrict broader inquiry, such as that undertaken here. Where one is unconcerned with the specific *content* of the information at issue, or even the ability to convey a *particular* message, a broader definition is more appropriate.

Losee's domain-independent definition of information emphasizes this generality: "Information is produced by all processes and it is the values and characteristics of the process's output that are information."⁸⁷ Note that this definition is not concerned with the "truth" or "falsity" of information—the output characteristics convey information even if that information is not true.

83. Claude E. Shannon, *A Mathematical Theory of Communication*, 27 BELL SYS. TECHNICAL J. 379 (1948)

84. *Id.*

85. Robert M. Losee, *A Discipline Independent Definition of Information*, 48 J. AM. SOC'Y FOR INFO. SCI. 254 (1997).

86. *Id.* This is not to say such definitions are useless or wrong, but rather that they are more limited in their usefulness. *See, e.g., State v. Green*, 493 So. 2d 1178, 1184 (La. 1986) (focusing on the idea that information is knowledge: "In the context of the facts now under consideration it is clear the 'information' sought by the state from Walsh was his *knowledge* of Green's possession of the gun prior to its surrender to the police authorities.") (emphasis original). Considering the definition of information within certain domains—quantification for Shannon or information as knowledge for the court in *State v. Green*—is certainly appropriate, but does not necessarily leave us with a definition that is appropriate for other uses. Thus we search for a broader and more inclusive definition.

87. Losee, *supra* note 85, at 256.

For Losee's purposes, anything that happens or occurs that yields some characteristic or value produces information. We perceive the characteristic or value as information, but there is no need for actual perception for the creation of information. It is sufficient that the characteristic is capable of such perception. Thus, all processes yield information. What we do with that information—qualify it, quantify it, categorize it—is up to us. But, according to Losee, we do not need to decide whether something is information because any characteristic constitutes information.

However, for our purposes less of a focus on the process by which information comes into being may be warranted. Theories of memetics also provide useful conceptions of information that focus less on the creation process. In his book *The Selfish Gene*, Richard Dawkins argued that genes are not the only things that replicate and evolve.⁸⁸ Culture also evolves through the transfer of memes. Dawkins's meme, a unit of information, is a measurement of convenience, a "piece" of information that can stand alone for an idea or cultural phenomenon. According to Dawkins, "[e]xamples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches."⁸⁹ Meme theory provides a way of thinking about the conveyance of information. Memes exist at many levels, and may combine to make other memes.⁹⁰

These three ways of defining information—knowledge, output characteristics, memes—all have some relationship to our understanding of information, but none directly defines its contours. Given that here we are considering information ownership, we will define information as *any thing*⁹¹ *that might be acted upon based on its content*. The "thing" need not be tangible, but might be a sound, a gesture, or even a momentarily existing arrangement of clouds; each of these has the potential to convey meaning.⁹²

We must define information broadly so that we can understand the implications of providing a property scheme for the ownership of information. If we define information narrowly, we begin only within a subset of

88. RICHARD DAWKINS, *THE SELFISH GENE* 189-201 (2d ed. 1989).

89. *Id.* at 192.

90. *Id.* at 195-96.

91. The separation between "any" and "thing" here is not accidental; it allows for the emphasis on the idea that for information to be owned, it must be in existence in some form, though the form chosen is not a limiting measure for our purposes. For a more complete discussion of the "thinghood" of information, see *infra* notes 96 to 108 and accompanying text.

92. See, e.g., *State v. Fjermestad*, 791 P.2d 897, 901-02 (Wash. 1990) ("[W]e conclude that the term 'any information' as used in RCW 9.73.050 includes visual observations as well as assertive gestures.").

information ownership, and generalizing our conclusions to the broader range of information ownership becomes difficult. Note that the choice of definition has immediate implications for this discussion. Given the broad nature of the definition, it is unlikely that all information could ever be owned, regardless of reductions in transaction costs, market failure, or uncertainty. This point will have implications for the theory developed later.

The definition adopted here is consistent with Losee's conception of information as the outcome of any process. But rather than focusing on the *creation* of the output characteristic (the information), it focuses on the characteristic or the thing created itself.⁹³ The definition is not limited to those things that are *actually* acted upon because of their content (for example, that actually reach a human's senses and ultimately, her conscious awareness) but includes any thing that has the *potential* to be acted upon because of its content.

The adopted definition of information is also not limited in size, scope, or content. As with memes, one "piece" of information may well be made up of other "pieces" of information, just as one meme may be made up of other memes.⁹⁴ The whole and its component pieces are all information.⁹⁵

93. See also ANDREAS GÜNTHER, *PRODUKTHAFTUNG FÜR INFORMATIONSGÜTER: VERLAGSERZEUGNISSE, SOFTWARE UND MULTIMEDIA IM DEUTSCHEN UND US-AMERIKANISCHEN PRODUKTHAFTUNGSRECHT* (2001) (Translation assistance provided by Meike A. Heverly) (discussing various definitions of information, including an object oriented definition).

94. One example from Dawkins's exposition of memes is instructive in this sense:

So far I have talked of memes as though it was obvious what a single unit-meme consisted of. But of course that is far from obvious. I have said a tune is one meme, but what about a symphony: how many memes is that? Is each movement one meme, each recognizable phrase of melody, each bar, each chord, or what?

I appeal to the same verbal trick as I used in Chapter 3. There I divided the 'gene complex' into large and small genetic units, and units within units. The 'gene' was defined, not in a rigid all-or-none way, but as a unit of convenience, a length of chromosome with just sufficient copying-fidelity to serve as a viable unit of natural selection. If a single phrase of Beethoven's ninth symphony is sufficiently distinctive and memorable to be abstracted from the context of the whole symphony, and used as the call-sign of a maddeningly intrusive European broadcasting station, then to that extent it deserves to be called one meme. It has, incidentally, materially diminished my capacity to enjoy the original symphony.

DAWKINS, *supra* note 88, at 195-96.

95. But see *Piper & Marbury v. U.S. Postal Serv.*, No. 99-2383, 2001 WL 214217, at *14 (D.D.C. Mar. 5, 2001) ("While contracts may certainly contain information, such as the price of the goods being sold, the entire contract . . . itself cannot qualify as 'infor-

This does not require that the smaller pieces take precedence over the larger, or that the larger subsume the smaller pieces it comprises. Instead, it is simply a recognition that information inheres in a large variety of things⁹⁶ and results from myriad processes.

B. Owing Information: A Framework for Inquiry

While it seems clear that within legal and cultural understanding information is perceived as being subject to ownership, there remain questions about whether information can really be owned. These questions arise primarily because information is intangible,⁹⁷ and thus the spread of information is difficult to stop.⁹⁸

Difficulties with treating information as a thing to be owned have given rise to numerous doctrines intended to articulate a theory of information as a commodity or good that can be owned. One example is the distinction used in copyright law between an idea and a particular expression of that idea.⁹⁹ The creator of an expression of an idea has rights in that particular expression of the idea but not in the idea itself.¹⁰⁰ Though

mation' in any ordinary sense of either word."). This language, however, is not at odds with the understanding that information can be made up of other information. The Postal Service's argument in this case, as rejected by the court, was ostensibly the opposite: that once something can be identified on the whole as information, the constituent pieces of information that make up that whole are no longer cognizable for purposes of the Freedom of Information Act ("FOIA"). *Id.* While the language the court chose indicates that the whole—here, the contract—cannot be considered information, this is an indication only that it was still required under FOIA to investigate whether the constituent pieces of the contract were information, and could not simply look at the whole in this regard.

96. The "thingness" of information is also consistent with meme theory, as memes are thought to exist physically. *See* DAWKINS, *supra* note 88, at 192 ("[T]he meme for [a particular belief] is actually realized physically . . . as a structure in the nervous systems of individual[s] the world over.").

97. This is true even though a particular expression of information recorded in a medium, such as an audio compact disc, might be held in one's hand.

98. Anyone who has been the object of gossip and rumors can attest to this fact.

99. *See* Feist Publ'ns v. Rural Tel. Serv. Co., 499 U.S. 340, 349-50 (1991) ("To this end, copyright assures authors the right to their original expression, but encourages others to build freely upon the ideas and information conveyed by a work. This principle, known as the idea/expression or fact/expression dichotomy, applies to all works of authorship.") (citations omitted); *see also* Alfred C. Yen, *A First Amendment Perspective on the Idea/Expression Dichotomy and Copyright in a Work's "Total Concept and Feel"*, 39 EMORY L.J. 393 (1989).

100. *See, e.g.,* Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 547-48 (1985). The Court in *Harper & Row* stated:

No author may copyright facts or ideas. The copyright is limited to those aspects of the work—termed 'expression'—that display the stamp of the author's originality. Copyright does not prevent subse-

criticized by scholars,¹⁰¹ this so-called idea/expression dichotomy is well established in copyright doctrine and forms the basis of many judicial decisions regarding information ownership and use.¹⁰² It is one way in which the legal and economic system has attempted to grasp the nature of information in what on its face appears to be an attempt to provide incentives for its creation.

But there are many currently unresolved questions: whether ownership as an incentive to create is necessary, whether information can in fact be “owned,”¹⁰³ and whether the idea of ownership of information is a good idea in general.¹⁰⁴ As more information in a variety of forms passes easily over the Internet, ownership seems to be a more troubling concept.¹⁰⁵

quent users from copying from a prior author’s work those constituent elements that are not original—for example . . . facts, or materials in the public domain—as long as such use does not unfairly appropriate the author’s original contributions.

Id.

101. See, e.g., Amy B. Cohen, *Copyright Law and the Myth of Objectivity: The Idea-Expression Dichotomy and the Inevitability of Artistic Value Judgments*, 66 IND. L.J. 175, 231 (1990) (“[I]t is no longer necessary or valuable or even possible to dissect a work of art to uncover the universal truths or ideas which must remain freely available to all future authors.”).

102. See *Meade v. United States*, 27 Fed. Cl. 367 (1992); *Atari Games Corp. v. Nintendo of Am., Inc.*, 975 F.2d 832 (Fed. Cir. 1992); *Frybarger v. Int’l Bus. Machs. Corp.*, 812 F.2d 525 (9th Cir. 1987).

103. Contrast this question of actual ownership with the concept of perceived ownership adopted below.

104. See, e.g., Yochai Benkler, *Coase’s Penguin, or, Linux and the Nature of the Firm*, 112 YALE L.J. 369 (2002) (arguing for the recognition of a peer-based, rather than individual- or firm-based, mode of production that is facilitated by the existence of digital communications, rather than the traditional incentives allegedly provided by ownership and property rights in information) [hereinafter Benkler, *Coase’s Penguin*].

105. One recent commentator has even argued that information has not truly been property in the past, and is only now becoming true property because of the technical ability to encrypt information and to preclude those who do not pay from decoding it; it is because of the ability to exclude others in some physical form that information can become property. RANDAL C. PICKER, FROM EDISON TO THE BROADCAST FLAG: MECHANISMS OF CONSENT AND REFUSAL AND THE PROPERTIZATION OF COPYRIGHT (U. Chi. Law & Econ., Olin Working Paper No. 159, 2002), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=329340 (last visited Dec. 3, 2003). Portions of this argument have their difficulties—I have, for example, no efficient technical means to preclude someone from picking up my pencil from the desk, but I still consider it my property. I could put the pencil in a vault or a safe, but this could also be done with information. In both cases, once the vault is open, the information (and the pencil) are unprotected. Yet both can still be considered property. Technical encryption neither makes nor breaks the idea of information as property. Understanding, however, that technical encryption al-

These questions aside, this Article proceeds along a more pragmatic route for evaluating ownership. The assumption is that statutory legal structures provide the basis for the perception that something is property. Property, under this conception, evolves from a set of rules or structures intended to provide members of society with a clear idea of rights and duties in relation to certain things defined by that society. To determine, then, whether information is property, and as such can be owned, we must examine whether and to what extent the rights given in information are analogous to those given in other, more established property settings. The rights to which we will look include the right to use, the right to exclude others from access and use, and the right to transfer (also known as the right to alienate).¹⁰⁶ If we conclude that these rights can be found in relation to information, then information can be considered property, and it can be owned.¹⁰⁷

It will be critical throughout this inquiry to keep in mind our definition of information as any thing that might be acted on based on its content, and to recall that information need only have the potential to convey meaning, and that one “piece” of information may itself contain distinct pieces. Ownership of one piece does not necessarily imply ownership of its individual and component pieces, and ownership of one component piece does not imply ownership of the whole. Rather, we must be explicit

lows greater protection for information as property, a point made by the article, is certainly important to understanding the nature of information as property today.

106. This view of property is not universal. There are a number of theorists who argue that exclusion is the lynchpin to a finding that something is property. *See* Merrill, *supra* note 4, at 736 (discussing views of property and categorizing them as “single variable essentialism, multiple variable essentialism, and nominalism”). Single variable essentialists argue that the existence of the right to exclude makes something property. Multiple variable essentialists argue that the right to exclude is a “necessary but not a sufficient condition” for finding property. *Id.* This means that without the right to exclude, a thing is not property, but something more than simply the right to exclude is needed for a thing to be property. Finally, nominalism views property as whatever the legal system decides to call property; it has no form or function outside of the definition provided for it by the legal system. *Id.* at 737. While this last approach might seem similar to the one taken here (for example, states can create property by law, statute, and interpretation), this is too simplistic a view. Instead, Adam Mossoff’s recent article advocating an integrated view of property, where all the rights noted above—acquisition, use, and disposition—are important to our conception of property more accurately reflects the view espoused here. Adam Mossoff, *What is Property? Putting the Pieces Back Together*, 45 ARIZ. L. REV. 371 (2003).

107. Adam Mossoff concludes, within the integrated theory of property, that information is property, and that intangibility creates no impediment to our understanding of information as property. *Id.*

in articulating exactly which piece of information is owned when considering information as property.¹⁰⁸

C. Owned Information

1. Introduction

There is a critical distinction among the views adopted by different disciplines concerning the creation and existence of property rights. Legal scholars generally refer to the statutory, judicial, and regulatory environment when asking whether property rights exist as to a particular thing.¹⁰⁹ Economic theorists, however, generally insist that property rights exist where some person has the ability, not necessarily the right, to control the thing and to obtain or influence the income stream from it.¹¹⁰ Thus, someone who steals a car has no legal property rights in the car, but does have all or nearly all of the economic property rights in the car once it is in his possession.¹¹¹ Legal rights, then, can influence economic rights, but they cannot be entirely determinative of them. Instead, economic property rights are bestowed by physical and cultural means in addition to legal means.¹¹²

The legal doctrines of copyright, trademark, trade secret, and patent are the structures that create legal property rights in information. Within these structures, information is viewed as something that is owned, and given additional power conferred by technology, such as the ability to encrypt data or content so as to prevent those who have not paid access fees from using it, economic property rights can be gained as well. While each of the four legal regimes of information ownership creates differing, and at times overlapping, rights with respect to the information covered, each provides at least some rights that can be characterized in terms of use, ex-

108. This is not to digress into the “bundle of rights” theory of intellectual property ownership. Rather, it indicates that as with all things, information has properties, and those properties themselves may consist of information that is owned or unowned in its own and separate right. As to the bundle of rights theory and the exclusionary theories of property ownership, see Mossoff, *supra* note 106. As to things being made of other things, see Sellers, *supra* note 72.

109. See, e.g., Andrew Beckerman-Rodau, *Are Ideas Within the Traditional Definition of Property?: A Jurisprudential Analysis*, 47 ARK. L. REV. 603 (1994) (concluding that common law idea protection, trade secret law, and patent law are all property based regimes predicated on the recognition of property rights in ideas).

110. See YORUM BARZEL, *ECONOMIC ANALYSIS OF PROPERTY RIGHTS* (2d ed. 1997).

111. The cliché “possession is nine-tenths of the law” actually more accurately reflects the economic view of property rights, and not the legal view. Of course, “possession is nine-tenths of the obtaining economic benefit” is not quite as catchy a phrase.

112. BARZEL, *supra* note 110.

clusion, and alienation. It is not necessary to delve deeply into the doctrine of each of these four areas to grasp the concept that together they provide a structure for information ownership. A brief review of the relevant concepts from property law, followed by a discussion of how each is handled by information ownership laws, will suffice to prove that information is indeed subject to ownership.

2. *Property and Rights in Brief*

In addition to identifying the rights that are present when we discuss property, it is also useful to clarify property by form of ownership. There are three generally recognized structural forms of property ownership: private, public, and common.¹¹³ In private and public ownership, an identifiable person or group of persons controls the property in question. Property that is held in common is conceived of as property to which every person has equal rights—it is still property, but it is not “owned” property.¹¹⁴ In addition, some common property may be further limited, in essence controlled by a group of people, none of whom can exclude others within their group from use, but who together as a group exclude “outsiders” from use.¹¹⁵

As to purely common property, no one acquires the rights that inhere in property until it is claimed or converted from the public domain to private use.¹¹⁶ This does not mean that one cannot use it, but rather that one cannot exclude others from using it.¹¹⁷ In a limited commons, those who have control of the property can exclude “outsiders” from using it, but may not exclude others who have rights to it.¹¹⁸

In order to undertake the ultimate analysis required here, we must consider information ownership in light of our understanding of property. Thus, where property is defined by the existence of some combination of rights, the answer to the question of whether information can conceivably be viewed as property is to be found in considering the rights that information owners are granted under information ownership doctrines. The answer, as we shall see, is that information can be viewed as property.

113. See Carol Rose, *The Several Futures of Property: Of Cyberspace and Folk Tales, Emission Trades and Ecosystems*, 83 MINN. L. REV. 129, 129-44 (1998).

114. See *id.*

115. See *id.*

116. See *id.*

117. See *id.*

118. Still, a limited commons is not necessarily uncontrolled even as to those people who may not be excluded. Common property in Medieval England was often subject to restrictions and controls on its use, such as the number of sheep that could be grazed on it. See *id.*

3. *Information Ownership*

Literature concerning information often bemoans the manner in which information has become propertized and commodified.¹¹⁹ Seldom if ever, however, does this literature question that information is viewed as subject to an ownership regime created by copyright, patent, trademark, or trade secret law. That is, we believe on the whole that information is property, even though we are not necessarily happy with that conclusion.

Returning to the basic predicates of property—possession and use, alienation, and exclusion—the doctrines of copyright, patent, trademark and trade secret provide these elements in relation to information. These laws together make up the property regime applicable to information, and within this regime, information is seen as being owned. For example, copyright law provides exclusive rights to the creator (or owner) of works, including the right to exploit and alienate those works.¹²⁰ The exclusivity of the rights gives the copyright owner the legal authority necessary to exclude others.¹²¹ Patent law provides similar protections for inventions.¹²² Trademark and trade secret law add breadth to the ownership of information through the protection of marks in commerce and information that is protected by the owner from disclosure.¹²³

Note additionally that the U.S. Supreme Court has held that the doctrinal schemes that provide at least certain rights for protection of information ownership—trademark¹²⁴ and patent law,¹²⁵ for example—are property rights, and may even be subject to the protections of the Takings

119. See, e.g., Niva Elkin-Koren, *Cyberlaw and Social Change: A Democratic Approach to Copyright Law in Cyberspace*, 14 *CARDOZO ARTS & ENT. L.J.* 215 (1996).

120. See 17 U.S.C. § 106 (2000) (granting a copyright owner exclusive rights in copyrighted works, which include the right to reproduce, prepare derivatives, distribute, perform, and display).

121. The right of exclusivity is so strong it has been buttressed with additional rights in the form of the Digital Millennium Copyright Act's "anti-circumvention" provisions, designed to prevent frustration of exclusivity. See *id.* § 1201.

122. See 35 U.S.C. § 271(a) (2000) ("Except as otherwise provided in this title, whoever without authority makes, uses, offers to sell, or sells any patented invention, within the United States or imports into the United States any patented invention during the term of the patent therefor, infringes the patent.") This language creates the right to use and distribute in the patent holder, and also gives the corollary right to exclude others from using or distributing it. See also *id.* § 261 ("Subject to the provisions of this title, patents shall have the attributes of personal property.")

123. See 15 U.S.C. § 1114 (2000).

124. See *K-mart Corp. v. Cartier, Inc.*, 485 U.S. 176, 186 (1988).

125. See, e.g., *James v. Campbell*, 104 U.S. 356, 358 (1981).

Clause of the Fifth Amendment.¹²⁶ This adds further credence to the notion that information is viewed as property within legal and cultural circles.

4. Conclusion

Within the structures of copyright, patent, trademark and trade secret, there are rights provided in information that afford the same types of protections secured to real property owners. These rights are not necessarily unlimited, but nothing in property theory requires *unlimited* rights for property to be considered property. The rights systems themselves may overlap as to particular pieces of information, but this also is not fatal to our conception. A piece of land may be subject to both an easement and a lien, yet that piece of land can still be owned. Overlapping property rights schemes go more to the specificity with which we view information than to the actual ability to own information. Within the broad definition of information we have adopted here, information is clearly viewed as something that can be owned.

D. Economics of Information

Several aspects of information factor into an economic analysis of information as a commodity. Information is generally considered a public good¹²⁷ since many people who want to use a piece of information at one time theoretically can do so without other potential users being precluded from its use. One person using a piece of information does not prevent another person from using that same piece of information, either for the same or a different purpose, or at the same or a different time. This makes information non-rivalrous, meaning that it can be used without depletion.¹²⁸ In addition, it is difficult to preclude people from using information. It can be distributed in a variety of ways because it is non-rivalrous, the “original” creator or distributor may not even know distribution is occurring. This makes information non-excludable: users who have not paid for it can be kept from using it only at very high costs. Absent legal protection, information produced by one person is easily used and distributed

126. The Fifth Amendment to the U.S. Constitution provides, in part, “nor shall private property be taken for public use without just compensation.” U.S. CONST. amend. V. The Fifth Amendment was applied to intellectual property in *James v. Campbell*. 104 U.S. at 358.

127. No one seriously disputes the public-good nature of information. *See, e.g.*, Julie Cohen, *Copyright and the Perfect Curve*, 53 VAND. L. REV. 1799, 1807 n.21 (2000). What is disputed is what to do about information being a public good.

128. *Id.*

by others, without deprivation of use or provision of economic benefit to the creator.¹²⁹

This understanding of information as an economic good allows us to see that information has inherent difficulties in any attempt to “own” it outside of the legal structure. While a builder may use force to defend against occupation of a house before it is paid for, once an informational good is out there, it is difficult to prohibit anyone’s use of it without legal protection. Legal structures build in exclusivity for the purpose of creating a legal market where none would otherwise exist.

E. The Nature of Information

While the economic nature of information can be identified, not everything can be defined and described in terms of economics. It is therefore especially important to understand information’s non-economic nature. As with the economic nature of information, the most critical elements of the non-economic nature of information come from its intangibility. This intangibility leads to information’s description as a public good in economic terms.¹³⁰ But it also leads to another important element in information’s existence: slight changes in how information is expressed can lead to the meaning¹³¹ being passed along without distribution of the original piece of information. For example, one person can read a book and tell another person about it; the second person may not need to read the book in order to learn about and understand the idea behind the book. The book contains a particular expression of meaning. That meaning can be passed from person to person, even if the book itself or the precise expression contained within it is not transferred.

This ability of information to be distributed independent of its particular expression is critical to understanding how the semicommons theory applies to information as property. Not only does this separable quality underlie the public good aspects of information, namely non-excludability and non-rivalrousness, it also shows that something else is happening with information, something that requires us to think of information using a metaphor other than physical property.

129. *Id.*

130. *See id.*

131. Recall that above we adopted the definition of information as any thing that can be acted upon based on its content. *See* discussion *supra* Part IV.A.

A further aspect of information is that it is essentially timeless.¹³² While the truth of certain information may be disproved over time, or become lost or distorted, the information itself can theoretically remain perfect throughout time. If someone was to write the phrase “the sun rose at 3:54 a.m. on Tuesday, March 4, 2003,” it is easy to imagine this information being recalled exactly as it had been written originally, even 200 or more years later. While some tangible commodities may also be essentially timeless—diamonds, for example, do not necessarily deteriorate over time¹³³—there is no tangible commodity that has the potential for exact continued existence to the extent that information does.¹³⁴

This should not be read to imply that the value of the information does not change over time. Certainly it does, but it is nearly impossible to broadly define the *way* in which that value changes. Some information becomes more useful over time, other information less so. Some information that has had relatively little value may suddenly become very valuable. Information might be valuable only for a narrow purpose for a specific period of time. Applying another meteorological example, a report of the weather on a particular day might have relatively little value. However, the information could be extremely important in a case involving an automobile accident that occurred on that day.¹³⁵ The timelessness of information is thus rather unique in terms of commodities.

An additional element of information that distinguishes it from more traditional, tangible commodities, involves the ability of particular pieces of information to exist in many different forms. The same piece of information might exist in writing, in digital form, in video, in a recorded audio track, or even in someone’s mind. In contrast, tangible commodities might exist in different forms, but the exact commodity cannot. Consistent with

132. This argument is not an argument that the *value* of information is timeless, but simply that the “thing” we call information can theoretically exist without much difficulty in perpetuity.

133. Deterioration here is meant to follow the line that information is timeless as to itself, not as to value. Information goods can be depreciated in the economic sense of that term, and lose value over time, but the information itself stays the same. The same cannot generally be said for tangible goods, which require maintenance and additional expenditure at the very least to continue in existence, though some tangible items are more durable in this respect than others.

134. This does not guarantee that information will exist forever, but only that under the proper circumstances it could exist forever.

135. This discussion points out the distinction between depreciation or value, and deterioration of a good. In terms of the good, so long as it is remembered or recorded in some way, the information can be in existence forever, and can be “working” during that time. It does not physically deteriorate—again, for the very reason that it is different from most other goods, it is intangible—and so can exist in perpetuity.

the laws of physics, matter can exist in only one place at one time. Information's intangibility allows it to escape this physical restraint.

The multiplicity of forms of identical information leads to another distinction between the tangible and the intangible. Information is subject to distortion in ways that other commodities are not. For example, if someone remembers reading the statement that the sun rose at a certain time on a certain day, but recalls the details incorrectly, the original piece of information has changed or become distorted. The false belief as to what time the sun rose is still information, but it is no longer identical to the original piece of information. As such, information can lead to and create new information without replacing or destroying the original.

One final and potentially significant aspect of information as a commodity is that there is no conceivable way in which all information can be owned. The accuracy of this statement becomes clear once we consider the breadth of what is covered by our definition of information. Recall that, according to our definition, the meaning potentially conveyed need not be true or accurate; indeed, it can even be false. False information is still information, even if it is not useful in the end.

Some examples of what qualifies as information are helpful to illustrate this final point. A tree can convey information; an observer might see that the tree's leaves are falling and conclude that it is autumn. Broken glass lying next to a car with a broken window on a city street can convey the information that the car has been broken into. A flag stretched taut from a flag pole can convey the information that it is a windy day. Certainly, more staid and conventional things can convey information: books, movies, sounds, images. But even given these more formal kinds of information, it is difficult even to conceive of a world where all information could be owned, where one would be required to reimburse another when she sees the waves rolling onto the beach or hears the call of the eagle on the wind (or perhaps just as importantly, when she tells—that is, distributes—that information to others). The waves and the eagle's cry can both be acted upon due to their content, and thus meet our definition of information. But the idea that someone could possibly own that information is beyond comprehension. The existence of such a wide variety of information, in a world where essentially *everything* can convey information, makes it clear that *all* information cannot be owned, regardless of the strength of information rights enforcing technology or the manner in which the law is written regarding it. The implications of this conclusion are set aside for now.

F. Final Thoughts

We generally perceive information as being subject to ownership. Our media, judicial, and legislative discourse echoes this perception. It is clear, however, that when we pause to consider the meaning of the term information, and in turn consider the application of that meaning within present legal doctrine, not *all* information can be owned. However, for the purposes of the remainder of this Article, we will accept as true the perception that information can be owned. We do so only to allow us to engage in a discussion about the type of ownership regime that would exist if that perception were ultimately proven. It is to that task that we now turn.

V. INFORMATION OWNERSHIP AS A SEMICOMMONS

A. Semicommons Theory

In his article *Semicommon Property Rights and Scattering in the Open Fields*, Henry Smith identified and described the theory of semicommon property ownership.¹³⁶ Initially, Smith adopted a simple property model to establish the existence of the semicommons, explaining, “On this view, private property rights will emerge when they are cost effective, and efforts at definition and enforcement of property rights will proceed until marginal benefit equals marginal cost.”¹³⁷ While acknowledging that this pattern can and does become more complex where appropriate,¹³⁸ Smith focused on the costs of enforcing property rights to help avoid the quagmire of assumptions based upon zero transaction costs. Smith’s article described the conditions under which semicommon property ownership becomes efficient.

In order for semicommon property ownership to be efficient when contrasted with purely private, purely common, or hybrid property schemes, this property regime must involve a dynamic relationship between the private and the common uses, with significant impacts of one type of use on the other. Once the dynamic relationship is established, the benefits of operating on multiple scales must outweigh the costs imposed on the ownership scheme by strategic behavior and its prevention. Thus, the nature of the strategic behavior must be considered, including either

136. Smith, *supra* note 7.

137. *Id.* at 133.

138. *Id.* (“[T]ransitions to and from property regimes require contracting among participants. The institutional structure in which contracting takes place is also a key to the formation, persistence, and eventual destruction of a semicommons property regime.”).

how the benefits of the scheme outweigh the costs of the behavior or how strategic behavior is controlled or prevented.

In applying the semicommons notion to the open fields, Smith identified the dynamic relationship between private and common uses as the contrast between the economies of scale available for the grazing of animals and the private incentives required to encourage grain growing,¹³⁹ which at the time enjoyed no economies of scale.¹⁴⁰

The difficulty of the semicommons notion now becomes apparent. In any situation like the one described, individual property owners will attempt to overuse the commons to the extent that they can do so without suffering the costs themselves, presenting a classic tragedy of the commons scenario.¹⁴¹ In addition, they will attempt to both minimize the negative effects of common use on their own property and maximize the negative effects of their own private uses on the commons.¹⁴²

According to Smith, there are four potential cases of strategic distribution of harms in the semicommons: 1) commons use imposes costs on commonly owned attributes such that each actor bears a proportional amount of the costs of use; 2) commons use imposes costs on privately owned attributes, either with a) all costs being imposed on the actor's own property such that the actor bears all of the cost for the use, or b) costs are imposed on other people's privately owned attributes, such that the actor bears none of the costs of the use; 3) private use imposes costs on private

139. *Id.* at 132.

140. Smith did note that applications beyond the open fields should be considered in the future. Specifically, he noted that its application to intellectual property might be appropriate. *See id.* at 165-67 & n.124. There are differences between Smith's original, albeit brief, notation and the proposal later outlined in this Article. In Smith's example, technology developers choose to make their technology into a semicommons, as did the farmers in the open fields (which might be termed a "created semicommons"). *Id.* The argument as to the regime of information ownership is that, at least in part, it would be difficult for it to be anything other than a semicommons (which might be termed a "natural semicommons") so long as private rights in information are provided for in law. If private rights of ownership in information were terminated, a pure commons would return; whether this would be sustainable or not is not the point. Rather, the question is instead how we would describe the state of affairs that would exist in that case. It is difficult to imagine a situation in which all information is owned, or even that all use of any one piece of information can be forbidden. The conclusion is that private use and common use must coexist in information, and that because of this, information ownership is a natural semicommons.

141. *Id.*

142. This does not necessarily imply that a deceptive or "bad actor" attitude is being reflected. Rather, it is understandable to attempt to minimize costs to oneself, even if that minimization involves shifting those costs to the commons and its other users.

property, either with a) costs imposed on the actor's own property such that the actor bears all of the cost for the use, or b) costs are imposed on other people's privately owned attributes, such that the actor bears none of the cost of the use; or, 4) private use imposes costs on the commonly used attributes of the property.¹⁴³

The types of strategic behavior in the open fields revolved primarily around the private user's attempts at maximizing the value of the commons use while minimizing private costs. In this instance, the animals were the common property. Benefit could be derived from the common property if the shepherd could be persuaded to station the flock on one's private property overnight. This provided fertilization for the private owner's crops. In contrast, the daytime activities of the animals imposed costs on owners of private property, as the animals ruined crops by grazing on and walking over them. Thus, when a private owner of property could convince a shepherd to station the flock on his property overnight, but graze the flock elsewhere in the daytime, that private owner would benefit. The benefit of the commons use was the fertilization that occurred when the animals were kept in an enclosed (and ostensibly private) area overnight.¹⁴⁴ A private user, for example, an individual land owner, who could encourage, pay, or otherwise coerce the shepherd in charge of the grazing to station the flock on his property overnight would gain additional fertilization. In contrast, the daytime activities of the flock, walking and grazing, could impose costs on a private user through the acts of walking and grazing. Where a shepherd could be convinced to graze the flock elsewhere, a private owner would benefit.¹⁴⁵

The strategies set up to avoid the private strategic behavior were primarily based on altering the borders of the private use areas so as to make the costs of engaging in the strategic behaviors prohibitive. "Scattering" is the term applied to the private use scheme, where a private owner's lands were scattered throughout the entire range of the grazing land.¹⁴⁶ Scattering increased the cost of engaging in strategic behavior: it made it difficult, if not impossible, for the shepherd to identify whose land the flock

143. *Id.* at 139. Omitted from Smith's conception of strategic behavior under possibilities two and three are the two possibilities in which costs are imposed either by common use (under use type two) or private use (use type three) on *all* private users. *See id.* In such a case, though unable to strategically avoid all injury to the privately owned attribute, the actor will attempt to maximize costs imposed on the privately owned attributes of others and minimize the costs as to her own privately owned attributes. *See id.*

144. *Id.* at 149-50.

145. *Id.* at 150.

146. *Id.* at 146.

was either grazing upon in the day or enclosed in for the night. Because the lands were not clearly marked, it would have entailed the expenditure of significant amounts of effort, for the shepherd and a private user to agree to the strategic grazing or enclosure of the flock. Through scattering, strategic behavior was minimized, but the dynamic benefits that flowed from using the land for both private and common uses, continued to be valuable.¹⁴⁷

The costs for the property system of avoiding the strategic behaviors is the crucial issue in terms of recognizing an efficient semicommon property scheme. Where the benefits that flow from the dynamic relationship between the public and common outweigh the costs of avoiding the strategic behavior, the semicommons will continue to exist. Where another form of ownership becomes more efficient through either a new way of allocating rights, such as through contract, or where there are changes in the nature of the strategic behaviors themselves, semicommon ownership may no longer be efficient.¹⁴⁸ In the open fields example used by Smith, the ability of private users to efficiently allocate ownership and use rights through contracts led to enclosure.¹⁴⁹

In summary, a semicommon property regime will come into existence where the benefits resulting from the dynamic interaction of private and common uses outweighs the costs imposed by the need to prevent strategic behavior on the part of the private users. Other forms of property ownership are likely to emerge where changes in the nature of the strategic behavior raise enforcement costs or where the methods of allocating rights lead to lower costs for these other forms of property ownership. We now address how this scheme plays out in the information realm.

B. The Information Semicommons

Moving from the semicommons as outlined by Smith to the information semicommons is a logical step. Once we accept that information ownership law creates a property regime in information, the application of the semicommons theory to that property regime is readily achieved. That achievement, however, entails a more detailed exposition of the dynamic nature of the private and common uses of information than was required for Smith's account of the semicommons in the open fields.

Smith was concerned, in significant part, with showing that scattering in the open fields was in fact an economically efficient response to the

147. *Id.*

148. *Id.* at 160.

149. Whether a similar fate is in store for the information semicommons is a question left aside for the moment.

need to control strategic behavior.¹⁵⁰ On its face, scattering appeared to be inefficient, as it required farmers to farm in varied locations. But, as Smith pointed out, this misses the overall economic picture.¹⁵¹ There were no economies of scale for farming during this time period, but there were economies of scale to herding. In order to take advantage of scale in herding, while maintaining the incentives for private individuals to farm, strategic behavior needed to be prevented. In addition, the costs of avoiding the strategic behavior needed to be lower than the benefit that resulted from this form of semicommons cooperation. Scattering of private plots, and thus private uses, was part of the overall scheme used to prevent strategic behavior, and viewed in this light, it was a rational and economically efficient method of ownership.¹⁵²

As noted above, these burdens, benefits, and strategic behaviors can be readily described and discussed. Smith's attention to the economic justification for scattering, rather than on developing categories for strategic behaviors is thus easy to understand. Making the case for the existence of the semicommons, as opposed to explaining the particulars of its functioning, was the comparatively easy part.¹⁵³

The information semicommons is more complicated. For this conception of information as property to be plausible, this Article must establish that the relationship between common and private uses is in fact dynamic, with interdependencies between the uses analogous to those found in the open field semicommons. If the benefits flowing from the combined private and common uses do not flow to both types of uses, a semicommons approach is inappropriate.

1. Private and Common: Information Use Dynamics

Private and common information uses interact. This interaction is not just the "push and pull" variety: it is not simply a conflict of uses where one is always antagonistic with the other, where each vies for the right to use information in a particular way. Rather, the interaction is dynamic: private use affects common use and common use affects private use, and both types of uses are better off for the existence of the other. This is not a claim that every single type of use, even that which cannot be considered

150. Smith, *supra* note 7, at 134.

151. *Id.*

152. *Id.*

153. This is not to say that making the case for the existence of the semicommons was easy per se, but rather that describing the implications in terms of dynamic relationships between private and common use for the open fields required comparatively greater effort.

strategic behavior, is beneficial on a micro level to other uses. Not *all* common use is beneficial to every private use, and not *all* private use is beneficial to every common use. Rather, the claims made here about the value, benefit, and dynamic interaction of information uses are made at the macro level, and must be addressed in that framework.

Even at the macro level, identifying the precise ways in which the interactions occur, ways that show the dynamic nature of the interdependence of the private and common uses, is not a simple task.¹⁵⁴ For purposes of this Article we will refer to two levels of interaction: content level and distribution level. Content level interactions are those that are involved in acting on or because of the substance of the information or its meaning. An example of such an interaction is the discussion that ensues when a new scientific theory is presented at an academic conference. The presentation of the theory, along with the reasoning that supports it, will undoubtedly lead to a discussion of the principle by the participants and, perhaps, an improvement upon the idea. This is content level interaction: the information is presented, other information and processes act on it, and the information is in some way changed.

A change resulting from interaction occurring at the content level can be an improvement in the information from a normative perspective, that is to say that the information might be better than it was before. On the other hand, the importance of the information may be diminished, resulting in the rejection of the information based on its content. Interaction at the content level is a critical aspect of the information dynamic of the type being described here in the information semicommons.

Information interaction also occurs at the distribution level. Distribution level interactions are those that occur when an interaction has an effect—market-based or otherwise—on the distribution of information.

Sometimes content level and distribution level interactions appear to overlap or conflate, but in fact the interactions can generally be teased out

154. It might be argued that authors or creators are not the proper focus of creative activities. The rejection of “authorship” has been in vogue in recent years, with a steady stream of scholarship that rejects the notion that works are created by individuals laboring alone and away from the rest of the culture. Such theories hold that the author is thus not the correct focus of information protection, and that information protection has been and is being overextended in terms of protections provided to authors. Whether one accepts or rejects these arguments is not at issue here, as this piece is concerned primarily with arguing that information rights regimes *as they currently exist* create a semicommons in information. It is clear that information rights are based on the acceptance of the author/creator paradigm, and as such attacks on it would, if successful, alter the nature of the information ownership regime in ways that are beyond the scope of this Article. See *infra* note 155 and accompanying text.

separately. For example, if a book review is published about a new novel, the book review will likely have dynamic effects on the original novel at both levels of interaction. Assume the book review is positive, encouraging readers to take the time to read the novel, while also discussing the novel's theme. At the content level, the review's treatment of the novel's theme might cause another novelist, or a critic or discussant, to take up or consider alternative points that otherwise might not have been raised in relation to that theme. That is interaction at the content level.

At the distribution level, the positive review will likely lead to greater sales than would have occurred without the review. The strength of the distribution level interaction effect will vary depending on the place in which it occurs. For example, a book review in a small circulation newspaper will likely give less of a boost to sales than inclusion on the reading list for the now defunct "Oprah's Book Club." Note that the effect at this level does not extend to the content: we are unconcerned with the substance of the book when considering the distribution level interaction. Instead, we see the common use interacting with the private use—in this example, the discussion of the book and its theme interacting with the initial creation of the novel.

The dynamic relationship between the private and common uses of information is thus encapsulated within these two levels. Within the content level, information interaction leads to the evolution of existing information and to our understanding of that which we do or do not "know." At the distribution level, the effects are primarily, although not exclusively, economic; the discussion or passing on of information has some effect on the market for the original and subsequent uses. Both levels may be involved in any particular interaction or use, but the effects of each are distinct. Even where both levels are present it is useful to tease out these distinct effects.

2. *Information "Creation" and Information Dynamics*

One of the easiest ways to see how private and common uses of information interact dynamically is to look at the information creation process. While the conception of a solitary author creating meaning entirely from whole cloth is well established in western culture, there are a variety of scholars who argue that this notion is a false one that is not born out by critical inquiry.¹⁵⁵ Instead, the people who modern society considers authors create based on the whole of their experience, which includes previ-

155. See MARK ROSE, *AUTHORS AND OWNERS: THE INVENTION OF COPYRIGHT* (1993); see also JAMES BOYLE, *SHAMANS, SOFTWARE AND SPLEENS: LAW AND THE CONSTRUCTION OF THE INFORMATION SOCIETY* (1996).

ously created works and personal experience, understanding, learning, and knowledge.¹⁵⁶ “We stand on the shoulders of giants,” is one metaphor used to describe creative efforts.¹⁵⁷ We cannot discount entirely everything that has come before. Even when things seem entirely new, they borrow, beg, or steal from things which came before. This line of reasoning accentuates the idea that information interacts with information. Even if information is owned, it is likely to be recast and included (many times legally) in subsequent information creation.

Thus, for those who argue that information creation does not occur anew, it is easy to acknowledge that both private and common uses of information interact dynamically. In this conception of creation, existing information forms the basis of new information and understanding. That what is perceived as the “new” information, or at least, new expression of information, is privately owned does not alter this picture. New creation cannot be separated from old, and the dynamic interaction of the alleged private use with the common is apparent from the face of the positive claims about information creation.

This view of information creation is, within the framework of interaction laid out above, a primarily content level interaction. It is one in which existing information leads to and is enlivened by new information. It shows that private use, while limited within this conception of authorless creation, interacts with common use. New ideas spring from old ones, reusing or reframing concepts. Without the existing ideas, new ideas would not spring from the well of the commons, and would not then contribute again to it. Common use adds value to the private use, and private use adds value back to the common use.

As we have seen, however, this conception of authorship is not currently implemented in the structures of information ownership that are currently relied upon.¹⁵⁸ Instead, the current system appears to accept at face value the idea of authorship and the concomitant need to provide economic incentives for information creation.¹⁵⁹ Within the more accepted

156. See BOYLE, *supra* note 155; ROSE, *supra* note 155.

157. See Michael D. Birnhack, *The Idea of Progress in Copyright Law*, 1 BUFF. INTEL. PROP. L.J. 3, 42-45 (2001) (analyzing the standing on the shoulders of giants metaphor).

158. *But see* L. Ray Patterson, *Copyright Overextended: A Preliminary Inquiry into the Need for a Federal Statute of Unfair Competition*, 17 DAYTON L. REV. 385, 386 (1992) (arguing that the conception of authorship to works of “low authorship” was inappropriate).

159. Recall that claims such as these are made here in a positive, not normative, fashion. This Article should not be read to argue either that a property regime in information is necessary to provide sufficient incentives for the creation and distribution of informa-

notions of information creation, then, the question of whether private and common information uses interact dynamically remains.

Even if we accept that the basis of current systems of information ownership rely on a conception of authorship as one in which the creator has rights because he created some information good that is new or unique, we can still identify a number of dynamic interactions between private and common uses of information at both the content and the distribution levels. Initially, let us consider the example of a biography of an actual person in printed book format that is written by an author and distributed by a publishing house. When the book is released, information about it must also be released. Without information describing the book, or at least its availability, no one will know it is available and thus no one will buy it. The information made public about and surrounding the book's release provides us with information about dynamic interactions between private and common uses of information.

The book is a private use of information, protected by the law of copyright. In this case, it is built upon information that was available to the author about a certain person, the times in which she lived, and the facts surrounding her life. It may also contain conclusions about the events and facts in which the biography's subject was involved, and these conclusions may appear in explicit form for the first time in this biography. To that extent, the biography is a mix of information, some novel, some not.

When the book is released, it is in the best interest of the author and distributor to make the general public aware of it. In this regard, advertisements might be placed in newspapers, magazines, or other publications. As these uses are all by and for the author and distributor, they are still private uses, and are under the control and within the legal structures established for such uses. But advertisements can only have so much effect on the public's buying preferences. Advertisements are paid for by the author and distributor, and the public knows this. Thus, claims by the author about the book's relevance, importance, style, or value are perceived as biased. The private use of information in this way is not likely to meet all of the author's and distributor's goals, which are to distribute the book as widely as possible for profit.

The author and publisher thus might welcome additional publicity in the form of book reviews. These reviews, written by independent third parties, will almost always include quotations, sometimes rather lengthy ones, in which the words of the author are used verbatim to emphasize a

tion, nor that the author model of information creation embodied in the current regime is accurate in terms of the manner in which information is in fact created.

point being made in the review. In a world of purely private rights in information, the book reviewer would need to obtain permission from the author, and perhaps even pay a fee, to use her words in his review. However, under copyright law, no permission is necessary where the use being made is fair.¹⁶⁰ Criticism in a book review, and inclusion of some of the author's words, is likely to be viewed as fair use in this context.¹⁶¹ This then becomes a common use of the information that is otherwise subject to the private ownership of the author.

This common use of the text in the review allows a reader of the review to see the actual writing of the author, to examine the style and method used in determining whether it is something she would enjoy reading. A reader's likelihood of buying the book will depend on whether or not she likes what she reads. This is a distribution effect of the interaction of the private and common uses. Of course, strategic behavior is possible on the part of the private use author. For example, the author might try to forbid uses of her text in negative reviews¹⁶² But putting such strategic behavior aside for now (it will be addressed in greater detail later), the dynamic interaction at the distribution level between the common use and the private use is clear.

At the content level, the dynamics of the interaction can be even more pronounced, even if we only consider the effects of the uses on each other and not on outside markets or on considerations such as free speech. Within these more limited constraints, let us take a new example of an information good: this Article. It explicitly attempts to build upon the ideas that are contained within, among others, the writings of Henry Smith. His semicommons article forms the basis for both the conception of information as property identified here and its application. In this manner, his article has formed the basis for a new work, a new work that arises out of his private use.

This new information work, however, also has the potential to have an effect on the work from which it draws. If the ideas within this article are

160. 17 U.S.C. § 107 (2000) ("Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work . . . for purposes such as criticism, comment, news reporting, teaching . . . scholarship, or research, is not an infringement of copyright.").

161. *See id.*

162. Francesco Parisi and Ben Depoorter detail another potential private ownership outcome that negatively affects the information semicommons, namely, the fracturing of ownership interests. Ben Depoorter & Francesco Parisi, *Fair Use and Copyright Protection: A Price Theory Explanation*, 21 INT'L REV. LAW & ECON. 453 (2002). This "tragedy of the anticommons" poses dangers that differ from those that come from viewing information as a commodity in search of perfect privatization. *See id.*

accepted and adopted, or even discussed beyond the publication of the article itself, the original article has the potential to become more valuable in relation to the purpose for which it was originally written. It may be cited more frequently, and interest in it may also increase. This would represent both a content level and distribution level interaction. It is dynamic interaction at the level of information development, with significant potential for distribution effects. Readers who agree or disagree with the current article may well return to the work upon which it draws heavily, yielding an increase in the distribution of that work. This example shows that the dynamic effect need not be limited to one level, but that the effects can be evaluated in each case so as to more fully understand the nature of the dynamic interaction.

Finally, moving beyond the artificial limitation we placed on content level interactions, it is clear within the scheme we have described¹⁶³ that private and common uses interact to the benefit of the development of knowledge and ideas of society as a whole. That is, when an author creates a new work, that work has the potential to enter the public debate. Where it does, parts or portions of the work may be used by common users in evaluating, discussing, criticizing, or advancing the ideas of the work. Some of these parts or portions may be subject to the protection of copyright, and their use may or may not be authorized within that scheme.

A new example illustrates the point. Assume that Copernicus has published his great work *De Revolutionibus*, in which he asserts that the world is round and orbits the sun.¹⁶⁴ His work enters the public consciousness and becomes the subject of discussion. In terms of the search for knowledge, his work will be read, quoted, discussed, and critiqued. The common uses, such as discussion and critique, add to the contribution that the “original” work has made to the human store of knowledge. As such, if we were to provide information owners with the authority to exclude all common uses, the original work could only be advanced upon permission of the author, and would be limited by that author’s own vision and understanding.¹⁶⁵

163. Recall that the scheme as described accepts *arguendo* that the information property scheme’s basis of creating economic incentives for information creation is valid. This is not an acceptance of that core ideal, but rather an acknowledgment that it has been implemented within the current legal structure that leads to property rights in information.

164. Nicholas Copernicus, *De Revolutionibus* (1520-1541), available at http://bj.uj.edu.pl/bjmanus/revol/titlepg_e.html.

165. The dangers of such an approach should be apparent, but an example will help to illustrate the point: Albert Einstein, famed for his pathbreaking and novel theories of

Taken together, these examples make the case that private and common uses of information interact dynamically. When common uses are restricted, something is lost, not only to common use, but also to private use, something that is otherwise unrecoverable. In addition, loss of private use rights would yield an overall loss to both common and new private uses, assuming *arguendo* that private economic incentives are themselves necessary to spur information creation. This dynamic interaction of uses shows that information ownership in the current regime is not purely a private property regime, nor is it purely a commons. It is a semicommons.

3. *Strategic Behavior in the Information Semicommons*

a) Identifying Strategic Behavior in the Information Semicommons

As was true of the dynamic relationship between the private and common uses, the nature of possible strategic behavior in the open fields was readily identifiable and largely limited in its scope.¹⁶⁶ A private user would be expected to bribe the shepherd to graze the flock on other private users' plots during the day and to pen up the flock on his property at night so as to avoid the detrimental effects of grazing while capturing the benefit of fertilization.¹⁶⁷ Strategic behavior involves capturing as many of the benefits of the dynamic relationship as possible, while avoiding as many of the costs as possible.

There are many possibilities for strategic behavior in the information realm. One example is the unauthorized copying and distribution of information goods. Licenses and contracts are particularly relevant in this context. Because the owner of intellectual property has the explicit right to license or authorize its use by others (and thereby to exclude users as well),¹⁶⁸ creators of digital goods may restrict the end user's right to make duplicates of the goods they have purchased.¹⁶⁹ For our purposes, strategic

relativity and special relativity, is said to have rejected the idea of chaos playing a role in the laws of physics. If Einstein had been able to stop those who also believed in the role of chaos in physics from applying his theories, the progress of science would have been held back many years.

166. Smith, *supra* note 7, at 144.

167. *Id.*

168. 17 U.S.C. § 106 (2000) (“[T]he owner of copyright under this title has the exclusive rights to do *and to authorize* any of the following.”) (emphasis added).

169. Owners of information goods tend to use language that indicates the consumer has not bought the good itself, but rather has purchased only the right to use the good in accordance with the license. While licenses are often not available to purchasers prior to purchase, and thus questions of bargaining, consent, and related contract theory questions are certain to arise, judicial decisions seem on a trend to recognize such licenses and their

behavior may either take the form of imposing restrictive licensing terms, especially where imposed by a monopolist, or it may entail a violation of those terms. Keep in mind that a violation of the terms of the license should not automatically be considered strategic behavior, just as imposing terms, in general, should not be considered strategic behavior on the part of the producer. It is when the actor seeks to impose costs or distribute costs across other common or private users, as outlined in Smith's theory, that strategic behavior should be of concern within the semicommons model.

Information counterfeiting is an example of a strategic behavior that either imposes a cost on or is distributed across other common or private users. Such counterfeiting is a multi-million dollar affair in the United States and the world today.¹⁷⁰ It includes unauthorized copying and distribution of computer software, music and video compact discs, digital video disks, video cassette recordings, and other musical, visual, electronic print, and multimedia goods. Counterfeiting involves attempting to pass off an unauthorized product as the legitimate, authorized original good as produced by the owner. The copyright industry has dubbed this activity "piracy," and it includes any "unauthorized" copying of protected works.¹⁷¹ Counterfeiting is primarily a distribution level interaction: the content of the work is not being advanced, considered, or even evaluated. It is simply being distributed. Making a copy of a digital work is not inherently illegal, nor does such behavior by its very nature reflect a strategic choice on the part of the copyist. Problems arise only when the creation of the copy violates the rights protected by the information property scheme. Thus, it is not the falsity of the copies with which we are concerned, but rather, the

restrictions, and as such, the use of such language may be an accurate description of the transaction that occurs between the buyer and seller of information goods. *See* ProCD, Inc. v. Zeidenberg, 86 F.3d 1447 (7th Cir. 1996) (holding shrinkwrap license valid even though user did not have access to the license until after the purchase of the software package).

170. *See, e.g.*, Motion Picture Association of America ("MPAA") Anti-piracy Web Site, at <http://www.mpaa.org/anti-piracy/index.htm> (last visited Nov. 4, 2003). The MPAA states that

The [MPAA] and its international counterpart, the Motion Picture Association (MPA), estimate that the U.S. motion picture industry loses in excess of \$3 billion annually in potential worldwide revenue due to piracy. Due to the difficulty in calculating Internet piracy losses, these figures are NOT currently included in the overall loss estimates. However, it is safe to assume Internet losses cause untold additional damages to the industry.

Id.

171. *Id.*

distribution of the work in a commercial manner and in violation of the license provided by the manufacturer of the information good.

In addition to considering mass copying and distribution of “fakes,” there are other strategic behaviors that involve copying and distribution. These differ from those mentioned above, and are not considered counterfeiting. Part of the distinction lies in the fact that their existence is not an attempt to profit monetarily from the original, but rather is intended to engage in a system of copying and distribution outside of the terms of the license for other purposes. Such other purposes might include the belief that the owner of the information good is too powerful, or that information should be “free” and unowned.

This non-commercial or non-competitive copying and distribution, generally considered a violation of the rights of the owners absent some valid claim of fair use (in the copyright arena), exists at various quantitative levels. Some non-commercial distribution occurs through organized peer-to-peer networks discussed throughout this paper, such as Kazaa or the now defunct Napster. These systems allow users to connect directly to each other’s computers, and while they often facilitate the transfer of files through such services as generating lists of currently available files,¹⁷² in general they were designed to facilitate the sharing of any files, and are not specifically directed to the sharing of copyrighted works. Not all large scale sharing of digital goods takes place through peer-to-peer networks, however. Webpages facilitating the exchange of files, often more directly aimed at exchanging goods that are known to be covered by copyright, also exist,¹⁷³ as do other methods for large scale sharing, such e-mail lists or special FTP servers.

In addition to the mass sharing of files, which appears to be a significant development in strategic behavior in relation to digital information goods, individuals often share information goods on a smaller scale, from the showing of videos in dormitory common rooms¹⁷⁴ to the copying of video tapes or DVDs for friends to the “sharing” of software applications by providing installation disks to friends in situations where the original

172. For a file to be currently available, at least one user with the file on her computer must be signed on to the service in question at the time the file is sought.

173. These file sharing communities often utilize Internet storage that is not designed for the purpose of exchanging files, and depending upon the size of the community, can cause problems with websites that provide such services, such as overloading them and preventing them—at least for a time—from providing services.

174. This is technically a public display of the video, and is outside the license of most videos, whether owned or rented.

installation is still in place. Such behavior implicates notions of strategic behavior.

Examples of non-strategic copying might be the making of back-up copies of software or musical works to allow a user to make use of the information which he or she has purchased if something were to happen to the original. In addition, altering the format of music files—for example, converting CD audio files into the smaller and more efficient MP3 format, and storing the files on one's computer or on one's MP3 player—may also be outside the definition of strategic behavior.¹⁷⁵ The idea that the creator of the information good which one has purchased can dictate not only the number of copies that one may make of the work, but also how the person can listen to it on various media seems to be stretching the concept of ownership.¹⁷⁶

As with counterfeiting or faking, small scale copying and distributing is primarily distribution level interaction. These activities are not generally concerned with discussing or advancing understanding on a content level, but are instead are only concerned with providing additional copies. This does not have to be the case, as the creation of playlists of songs, for example, is a creative exercise, and reflects a content level interaction with the information used—in this case, the songs that make up the list. The distribution may also lead to content level interactions by people who otherwise would not have been exposed to the information in question.

The common use strategic behaviors identified so far have been primarily distribution level interactions; content level strategic behavior is more difficult to categorize. There are some cases in which a person takes a portion of an information good, such as an article or a book, and claims it as his or her own.¹⁷⁷ While at first this might seem to resemble counterfeiting, it does so only at the most basic level. However, in cases of plagia-

175. See, e.g., *Recording Indus. Ass'n of Am. v. Diamond Multimedia Sys., Inc.*, 180 F.3d 1072 (9th Cir. 1999) (bringing against manufacturer of MP3 player for facilitating copyright infringement).

176. Information owners, of course, will not necessarily agree with this analysis. If the tracks of an audio CD are converted into MP3 format on a computer, they remain on the CD as well. If someone is listening to the CD at the same time that the MP3 tracks are being played elsewhere, information owners would argue copyright law, both in spirit and in explicit terms, is being violated. See *Diamond Multimedia Systems, Inc.*, 180 F.3d 1072 (9th Cir. 1999), in which the Ninth Circuit rejected these claims, bringing MP3 player use, even the “ripping” of music tracks, within the protection of the Audio Home Recording Act, 17 U.S.C. § 1001 (2000). 180 F.3d at 1081.

177. This is often referred to as plagiarism, especially in the academic world. See Lisa Lerman, *Misattribution in Legal Scholarship: Plagiarism, Ghostwriting, and Authorship*, 42 S. TEX. L. REV. 467 (2001).

ism, the plagiarist does not acknowledge the existence of the owner, and instead personally claims the rights to the work, whereas the counterfeiter acknowledges someone else's ownership of the good, but pretends that the good is authorized or produced by that owner. The plagiarist's act is no less strategic behavior, however, because of this difference. In both cases, the common user attempts to claim the work, including the rights to the original, as his own property. This situation presents us with strategic behavior at a mixed distribution and content level, as the plagiarist is likely to change at least some parts of the stolen text. In so doing, the plagiarist introduces content level effects (recall that content level effects need not be positive).

Strategic behavior, however, does not just exist on the part of the common users in terms of copying and distributing information goods outside of the terms of a license. Private owners of information can act strategically as well. These strategic behaviors arise out of the method by which information owners grant access to their information. Owners will be selective in deciding who may receive rights and will likely base their decision on the proposed use to be made of the information. Owners may also base this decision on the economic effects of the manner in which they license the work. In addition, if the information protection scheme is designed to provide incentives for creation and innovation, then actions that keep information from use would seem to run counter to that notion, and the use of intellectual property law for these purposes is likely to fall within our definition of strategic behavior.

The types of behavior that would be strategic in this latter respect include requiring the purchase of additional information or non-information goods so as to be granted legal access to the desired information.¹⁷⁸ Such agreements attempt to take advantage of the desirability of particular information to advance the economic interests of the owner. As such, the interaction is primarily at the distribution level, for it is concerned not with the content of the information, but with tying other, less desirable products with it. Tying is one type of distribution level strategic behavior in which private owners of information engage.¹⁷⁹

178. This is an anti-competitive behavior known as tying. See David Gilo, *Retail Competition Percolating Through to Suppliers and the Use of Vertical Integration, Tying, and Vertical Restraints to Stop It*, 20 YALE J. ON REG. 25 (2003).

179. Not all tying should be considered negatively in terms of strategic behavior. Economic literature is primarily concerned with the negative effects of tying only when it involves restrictive circumstances, such as a regulated good being tied to an unregulated good (and thus raising questions of regulated prices subsidizing unregulated goods). Tying of goods benefiting from economies of scale may raise similar issues. See Daniel J.

A second type of private use strategic behavior is a mix between the content and distribution levels. It involves situations in which the owner licenses use of an information good based on the potential use proposed by or expected of the user. Where the use is viewed as being critical of or otherwise negative toward the information owner, the use is withheld. Where the use is viewed as positive, it will be allowed. Recall the earlier book review example where the book's author or publisher is more likely to permit a direct quote from the book if the proposed review will be a positive one.

The resulting interaction occurs at both the content and the distribution levels. A negative review may lower demand for the book. This is the effect that the information owner wants to prevent. But it will also have the effect of precluding criticism, or content level interactions, surrounding the book. Where only positive discussions of a new information product can occur, content level interactions will suffer significantly. Of course, the primary reason for the decision to preclude the quote need not be economic; it can also be based on author pride, or a desire to suppress dissent about the ideas contained in the information of the work. Such reasons constitute a purely non-economic, indeed, an economically irrational, reason for withholding permission.¹⁸⁰ Of course, discrimination need not be explicitly based on the granting of or refusal to grant a license. Information owners may also use price discrimination to mimic these effects, charging a prohibitively high price for access they wish to discourage or prevent.

Finally, the limiting of fair use, which is a constitutionally required¹⁸¹ and statutorily provided¹⁸² common use of information, through technological methods should be considered a private use strategic behavior.¹⁸³ This is another act that affects interactions at both the content and the distribution levels.

Gifford, *The Antitrust/Intellectual Property Debate: An Emerging Solution to an Intractable Problem*, 31 HOFSTRA L. REV. 363 (2003); James B. Speta, *Tying, Essential Facilities, and Network Externalities: A Comment on Piraino*, 93 NW. U. L. REV. 1277 (1999).

180. Howard Hughes, for example, attempted to purchase the copyright to articles that were being used in an unauthorized biography of him, with the eventual aim of stopping the biography. He was unsuccessful. See *Rosemont Enters., Inc. v. Random House, Inc.*, 294 N.Y.S.2d 122 (N.Y. Sup. Ct. 1968), *aff'd*, 301 N.Y.S.2d 948 (N.Y. App. Div. 1969).

181. See *Feist Publ'n, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 349-50 (1991).

182. See 17 U.S.C. § 107 (2000).

183. See Georgia K. Harper, *Copyright Endurance & Change*, 35 EDUCAUSE REV. 20, 24 (2000).

Strategic behavior in the semicommons of the open fields involved the physical elements of farming and grazing. Strategic behavior in information ownership is a more complex matter. It involves layers of content and distribution level interactions, not simple physical burdens and benefits. But the strategic behaviors used to maximize gain in the information semicommons are still identifiable, and while other strategies may exist,¹⁸⁴ those identified here will serve sufficiently to advance the core argument of this Article: that information ownership is a semicommons, complete with a dynamic relationship between private and common uses, and strategic behaviors engaged in by both private and common users. We now turn to a discussion of how the semicommons addresses the strategic behaviors identified above.

b) Avoiding Strategic Behavior

Strategic behavior in the open fields was avoided through a combination of norms, physical controls (technology and architecture), and law.¹⁸⁵ Strategic behavior in the information semicommons is dealt with in a similar manner: law, norms, and controls work together to combat strategic behavior in information.¹⁸⁶

One of the simplest methods for controlling strategic behavior is the requirement that property rights in information not be automatically perpetual. The four legal schemes of information ownership—copyright, patent, trademark, and trade secret—provide for the time and manner in which rights end. Copyright and patent law provide express terms of protection,¹⁸⁷ whereas the protection provided by trademark and trade secret law is of indefinite but terminable length.¹⁸⁸ While strategic behavior on the part of the private user is still a concern during the term of protection, the

184. For example, as to trademarks, one might use another company's or product's trademark so as to try to fool consumers into believing that the trademarked product is associated with his product. Another is "over enforcement," the use of legal action or the threat of legal action to dissuade common uses that are, in reality, non-strategic and permitted.

185. Smith, *supra* note 7, at 168-69.

186. Neal Kumar Katyal, *Architecture as Crime Control*, 111 YALE L.J. 2261, 2262 (2002).

187. The life of the author plus seventy years in the case of copyright for a natural individual; in the case of works for hire the period is either ninety-five years from publication or one-hundred and twenty years from creation (whichever is shorter). 17 U.S.C. § 302. In the case of patent, it is twenty years from the date of filing for the patent. 35 U.S.C. § 154 (2000).

188. Trade secrets are lost when the holder fails to take adequate measures to protect them. Trademark protection is lost if the mark becomes generic, or if it is abandoned (no longer used in commerce).

prospect of eventual termination of an owner's rights may serve to moderate the effect as owners who demand too high a price for their information goods may find users waiting until they are no longer protected.

Another method used to dissuade strategic behavior is the use of enforcement actions by information owners. Where information owners may bring lawsuits for the unauthorized use of information, common users may reduce their strategic behavior.

Yet another powerful method of combating strategic behavior is the use of definitions that set the boundaries of information protection. For example, copyright law protects particular expressions of information fixed in a tangible medium; ideas may not be copyrighted.¹⁸⁹ This powerful method of defining the applicability of the statutes, and in so doing defining the contours of the property, limits the ability of private users of information to behave strategically. As noted above, the idea/expression dichotomy has come under attack in recent years as unworkable, but its underlying theory remains doctrinally sound: while it may be useful to provide incentives to owners to create and invent, it does too much violence to constitutional notions of free speech to provide for ownership of ideas.¹⁹⁰ Protection of particular expressions of ideas is one thing; protecting the ideas themselves is another. Were ownership to be extended to the ideas themselves, speech would be unquestionably diminished.¹⁹¹

The fair use provision in copyright law is one of the best examples of an explicit statutory provision that helps to define common use rights of information.¹⁹² Fair use serves to exclude from infringement claims uses relating to "criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research"¹⁹³ The fair use provision demonstrates that the ownership of information exists in a semicommons.¹⁹⁴ It protects common users' rights to use information to

189. See *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 547 (1985).

190. See *Feist Publ'n, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 349-50 (1991); see also Cohen, *supra* note 101.

191. Note that strategic behavior by information owners approximates this result even where ideas cannot be protected. Howard Hughes, for example, in an attempt to stop the publication of an unauthorized biography, purchased and then refused to license for purposes of the biography the historical sources the biographer had used to develop his work. See *supra* note 180.

192. 17 U.S.C. § 107.

193. *Id.*

194. That fair use shows the inherent nature of information as a semicommons does not contradict earlier statements that fair use is viewed as fair use of someone else's property. See *supra* note 2 and accompanying text. Rather, it is an acknowledgement that fair use plays an integral role in the scheme we perceive as creating information owner-

criticize information owners, even in the face of the information owners' explicit objection. It thus combats the strategic behavior certain to arise when private and common uses must coexist.¹⁹⁵

In the open fields, technology and architecture were also used as methods of avoiding strategic behavior.¹⁹⁶ As the architecture for information is almost purely technological, these two methods of avoiding strategic behavior are blended for purposes of description. This should not be read to indicate that these methods are somehow less powerful in the information context. In fact, they may be more powerful than was the scattering of privately owned plots in the open fields. For example, digital rights management ("DRM") technologies, such as the DVD CSS encoding system,¹⁹⁷ provide potentially absolute bars against actions that private users consider strategic on the part of common users.¹⁹⁸ Where technology manufacturers and information producers cooperate to assure technological protective measures are implemented, architecture and technology combine to provide what is a generally strong, effective, and inexpensive method of combating strategic behavior.¹⁹⁹

Given the ubiquitous networked environment in which we live today, and the existence of peer-to-peer networks, private information owners have had to combat the allegedly strategic behaviors of common users in the trading of copyrighted files, such as music and movie files. One technical method adopted to combat such trading is the release of "fake" files onto the networks.²⁰⁰ The files are identified as "real" files, but they are in fact either incomplete or contain an audio recording that condemns the

ship, but that it is wrongly viewed as a tolerated infringement. A more accurate description of fair use is that offered here: fair use is part of what shows us that we have not gone far enough in our thinking about information as property; we have not asked what kind of property our structures create. When we do, we no longer see fair use as tolerable or tolerated infringement, but as a crucial element in the nature of private and common uses of information.

195. Note that fair use is generally perceived as a constitutional requirement. Too much change in the landscape of fair use, and the Supreme Court would reject the changes based on the needs of the First Amendment. This is an additional reason why all information could not be owned. In other words, the constitutional fair use requirement is another reason why information ownership *must be* a semicommons.

196. Smith, *supra* note 7, at 161.

197. *See supra* note 23.

198. *See Universal City Studios, Inc. v. Corley*, 273 F.3d 429 (2d Cir. 2001) (holding that the DMCA anti-circumvention provisions prohibit circumvention of the DVD CSS)).

199. As with other methods of avoiding strategic behavior, technology can also be used on the private side as a method of facilitating private strategic behaviors.

200. *See, e.g., Doug Bedell, 'Spoofs' Plague Music Swappers*, THE DALLAS MORNING NEWS, July 11, 2002, at 3D.

trading of copyrighted works. Users who wish to act strategically by trading copyrighted files are already implementing methods to avoid fake files.²⁰¹ The release of and reaction to fake files are clear examples of how private and common users adjust their actions to changes in technology and adopt technological methods to avoid strategic behavior in the information semicommons.

c) The Digital Revolution and Strategic Behavior

To develop an understanding of how strategic behavior occurs in the information semicommons, it is necessary to consider how the digitalization of information affects information goods and information related actions. Strategic behaviors in information related goods are not new. Detecting counterfeits in the fine art world, for example, is a field unto itself, an important component of the market for rare art.²⁰² What has changed is the cost of engaging in these strategic behaviors. Costs have changed because of technology, and all three aspects of the property rights paradigm—possession and use, alienation and distribution, and exclusion—are affected by this change.²⁰³ Technology now allows or has the imminent potential to allow increased ease of copying of digital goods, increased ease of distribution of digital goods, and, although apparently in contradiction to the first two changes,²⁰⁴ increased potential for technological exclusion of common users from use.²⁰⁵

Each of these changes has altered the way in which strategic behaviors affect the balance of uses in the information semicommons, and the changes have occurred both in terms of common uses and private uses. Digital works can be easily and exactly copied. This decreases the marginal cost of production of information goods to essentially zero, but creates new difficulties for manufacturers of digital goods.

201. One response has been the development of software that allows peer-to-peer users to listen to the files they are downloading as the download is occurring, which allows them to abort downloads of fake files.

202. See Raul Jauregui, Comment, *Rembrandt Portraits: Economic Negligence in Art Attribution*, 44 UCLA L. REV. 1947, 1960 (1997).

203. JORGE REINA SCHEMENT & TERRY CURTIS, TENDENCIES AND TENSIONS OF THE INFORMATION AGE: THE PRODUCTION AND DISTRIBUTION OF INFORMATION IN THE UNITED STATES (1995).

204. The Internet is often the example of nearly costless distribution, but even the existence of recordable CDs, each costing only a few pennies when purchased in quantity, has lowered distribution costs.

205. Exclusion is accomplished through the use of DRM technologies, which allow a creator or distribution to “lock” or control uses engaged in by users. Supplemented by the DMCA, which criminalizes circumvention of such technologies, exclusion rights in information appear at this time to be particularly strong. See 17 U.S.C. § 1201 (2000).

In addition to enabling the creation of perfect copies, digital technology has lowered the cost of distribution of digital goods. In place of a production plant with systems and workers producing, packaging, and shipping each good, digital goods are distributed via networks such as the Internet with virtually no overhead. This means that the marginal cost of distribution is also zero or close to zero. It also means that nonexcluded (and thus nonpaying) users can themselves distribute such goods inexpensively, essentially at the cost of their time used to obtain or distribute the goods.

Finally, digital technology has changed the way that users may gain access to digital goods. In the past, if a publishing house wanted to prevent a consumer from reading a book more than, for example, five times, there was little it could do to put its desire into practice. In today's digital markets, however, if the book is distributed in digital (often referred to as "electronic" or "e") format, the code that underlies the display of the book on the viewer or computer might limit the number of times the pages can be viewed.²⁰⁶ Producers, therefore, at least theoretically, may have the technological ability to control access to information goods in a way that was not possible before. This can, of course, lead to strategic behavior on the part of both producers, who may attempt to be overly restrictive in the rights they provide, as well as users, who may attempt to circumvent the protections put in place by the producer and thus obtain an unrestricted or less restrictive version of the good at the restricted price.

The effect of these changes brought about by the digital revolution is not yet fully cognizable. What we can say about them is limited to current ad hoc observation and predictions about the future. The technology itself is not settled. The "technological arms race"²⁰⁷ in digital works, where owners continually develop new protection that is eventually circumvented, is ongoing, and the pace of technological development seems to be slowing little, if at all. The point in discussing the effects of digital production of goods here is simply to highlight the ways in which digital goods are important to our understanding of strategic behaviors. The digi-

206. See *United States v. Elcomsoft*, 203 F. Supp. 2d 1111 (N.D. Cal. 2002) (describing the eBook software developed by Adobe Systems for this purpose, in the context of a constitutional challenge to the applicability of the DMCA to software developers).

207. See Trotter Hardy, *Property (and Copyright) in Cyberspace*, 1996 U. CHI. LEGAL F. 217, 251 (1996) ("The problem with this scenario is that it constitutes a kind of wasteful 'arms race' of technological-protection schemes, with each side increasing its spending to outperform the other's technology."). Note that the "property" in Hardy's title is in reference to property rules versus liability rules, and not property as a conception of ownership per se.

tal nature of information goods in the modern world overlays nearly all strategic behavior questions today, and thus must be kept in mind when discussing modern information goods.

C. Conclusion

By statute and judicial interpretation, information can be owned. The structure of the property scheme that describes the ownership of information, however, is not the dichotomy of private versus common use. It is a semicommons, complete with a strong dynamic relationship between the private and common uses, and a laundry list of both strategic behavior and methods for avoiding strategic behavior. Now that we have examined the positive claims concerning the nature of information ownership, we will discuss the normative implications of this new perspective.

VI. THE SEMICOMMONS EFFECT

A. Introduction

Considering information ownership as a semicommons provides us with an additional analytical tool with which to understand information and information use.²⁰⁸ It provides additional explanatory power with which to consider the changes occurring in the information age. For example, what is the purpose of fair use in a world of information goods. Wendy Gordon²⁰⁹ has offered a justification for fair use based on market failure. Where market failures in information exchanges could be worked out, such as through priceless or nearly priceless transactions in the “point and click” world of the web, fair use seems harder to justify.²¹⁰

Semicommons theory has the potential to put this debate to rest. Fair use exists as part of the recognition of the nature of information ownership. Attempting to privatize all information uses is destined to fail due to the nature of information and its uses and ownership. All information cannot be owned. Private uses must interact with common uses. Semicommons theory best recognizes this implication and is best used to tease out the implications of this conclusion.

208. This section of the Article owes a great deal to Professor Timothy Lytton of Albany Law School, who asked the following question while an earlier version of this Article was being presented at a Faculty Workshop: “So what? Aren’t there other ways of achieving the same ends without relying on a new and unique application of property theory?” The answer is no, but I am indebted to Professor Lytton for asking the question.

209. See Wendy J. Gordon, *Fair Use as Market Failure: A Structural and Economic Analysis of the Betamax Case and Its Predecessors*, 82 COLUM. L. REV. 1600 (1982).

210. See *id.*

While it might be possible to obtain similar results within each of the four intellectual property regimes, namely copyright, patent, trademark, and trade secret, absent the semicommons, the justification for laws that authorize “infringement” of private users’ property rights required within each area would necessarily be different. For example, in copyright, a re-reading of fair use that includes overall welfare calculations rather than private owner gain might be used to save peer-to-peer networks. But fair use would play no role in considering extensions to the length of protection offered for patents. In such a case, new arguments would be needed, but without the semicommons, such arguments would most likely seek to maximize the benefits of either the private or the common use.

In contrast, semicommons theory allows us to perceive the structure of information ownership more clearly and to acknowledge the dynamic relationship that *must* exist due to the nature of information. It is nearly impossible to conceive of a world in which *all* use of information is private, necessitating that any use by any person be authorized by an owner. This is one of the primary differences between land and information. It is possible to conceive of a world in which all land is subject to private use rights, and any use of any land must be authorized by its owner. In other words, it is possible to conceive of a world in which there are no common uses of land. That does not work in the information arena; as we have seen above, all information cannot be owned, regardless of how we try to categorize its various aspects.²¹¹

The semicommons effect, then, helps us to more accurately describe what information ownership looks like, how private and common uses of information are not only currently interacting, but must continue to interact due to the very nature of information itself. The normative effect of this positive state of affairs has the potential to be quite profound.

B. The Normative Effect of the Semicommons Approach: Acknowledging Reality

The primary normative implications that flow from understanding information ownership as a semicommons derive from the distinctions between the nature of land as property and the nature of information as property. Where information as a whole cannot be subject to precise ownership, where certain individual attributes of information cannot be owned in any sense of the word, it is inappropriate to continue along a path that is based on the premise that increased private use will yield increased wel-

211. This is not to argue that very *important* information cannot be owned or controlled, or even that certain categories of information cannot be owned.

fare. This is the argument against the commons in land; left as a pure commons, the resources of the land will be overutilized to the point of depletion.²¹² Placed in private use, so the argument goes, the resources will be used as efficiently as possible, thereby increasing the overall welfare.

In the context of information, however, this argument falls short of the mark. It is based on the false premise that all information can be privatized, or that certain information can be made completely and efficiently private, while at the same time other information will remain (ostensibly by choice) in the public domain. As a positive matter this is simply not the state of affairs as to information. Instead, information ownership exists in a semicommons, with a dynamic relationship between private and common use of the *same information*. Judicial interpretations that rely on the old private/common use dichotomy to interpret current and future statutes are destined to twist efficiency rather than enhance it as the law struggles to address information ownership.

The normative claim is thus that judicial decisionmakers, administrative agencies, and legislative policymakers must take into account not simply the desires of and benefit to private users when making decisions and adopting laws, but also the overall effect of those decisions and laws on the *dynamic relationship* between private and common uses. These interests must be balanced with an aim toward increasing *overall* welfare, not simply the welfare that is assumed to result from increasing private use. The implications of this change in perspective are made clear as we return to our peer-to-peer example.

C. Peer-to-Peer Revisited

After considering the primarily positivist claims above regarding the existence and operation of the information semicommons, it is useful to clarify the application of the theory with an example of how it applies to the issues in a current and ongoing debate regarding strategic behavior in the information semicommons. We need not revisit detailed theories of peer-to-peer sharing at this point, nor deeply investigate the factual scenarios involved in current realworld file trading to ascertain that semicommons theory provides us with background to more fully understand and analyze the interaction between the private and common uses of information.

212. Garrett Hardin, *The Tragedy of the Commons*, 162 *SCI.* 1243, 1243-48 (1968); see also ELINOR OSTROM, *GOVERNING THE COMMONS: THE EVOLUTION OF INSTITUTIONS FOR COLLECTIVE ACTION* (1990).

Copyright owners consider file trading a damaging kind of strategic behavior on the part of common users. Under a traditional property conception of information ownership, the files are “owned by” their private owners. The only uses allowed are those authorized by the owner, who is provided with authority in law to exclude others from use. This matches with our understanding of property. Although limitations on ownership, such as fair use, may apply, copyright owners are otherwise correct in stating that file traders are behaving illegally.

Within this traditional conception, common users of copyrighted files have offered myriad defenses for their activities. Some of these fit neatly into the private/common paradigm set out in copyright law—fair use in particular. Large scale trading, however, would seem to be outside of fair use, and courts have found this to be the case in litigation involving the peer-to-peer software applications.²¹³ From the semicommons perspective, a more interesting defense offered by users is that file sharing is justified because it increases demand for information by providing users, not with a substitute for the original, but with an opportunity to sample the original and determine whether it is a good worth purchasing. Restated in the terms of the information semicommons, the claim is that peer-to-peer file sharing on the whole creates positive distribution and content level interactions in information. It is a claim that outside of a semicommons understanding would be irrelevant—that by somehow “stealing” property, welfare is increased. Within the semicommons notion, the claim takes on new meaning and importance, and gains legitimacy based on the nature of information ownership.

This, of course, does not mean that common users who make these claims about peer-to-peer file sharing are *correct* about the effect their activities have on the market for information. It does mean that if we acknowledge that information ownership is a semicommons and look to maximize the dynamic relationship between private and common uses, we would seek empirical or theoretical evidence to support or deny the users’ claims.²¹⁴ Absent a semicommons understanding of information ownership, empirical evidence of the effect of peer-to-peer file sharing on the

213. *See supra* Part II.

214. While the Ninth Circuit’s first *Napster* decision discusses the proof offered by Napster of this very effect, recall that the court explicitly rejected the notion that even this allegedly positive effect could preclude a finding of infringement, as Napster displaces the record companies’ abilities to enter the digital market. *Napster II*, 239 F.3d 1004, 1081-19 (9th Cir. 2001). Even if Napster had this effect, it’s an effect that—according to the *Napster* courts—should be controlled by the copyright owners, emphasizing the private nature of the use as controlling. *See id.*

purchase of information goods is irrelevant. Within the semicommons, that information is critical to making correct policy and legal choices to maximize the efficiency and welfare generated by the system.

D. Using Content and Distribution Level Interactions in Decisionmaking

Concepts of content level and distribution level effects should influence the decisionmaking for information usage. In terms of functional implementation, it would be most convenient to devise a sliding scale whereby common uses receive less protection at the distribution level than at the content level in the semicommons. Such a scheme, however, would oversimplify the complex interaction between common and private uses.

Instead, we require a model more clearly approximating the interaction of common and private uses at the content and distribution levels. As previously discussed, uses at the content levels interact with the distribution levels. Returning again to our example of a book review, a reviewer may add to or subtract from the theories presented in the book. This is a content level effect that may in turn influence the distribution level of the book. For example, book sales may decline because the reviewer criticizes the content of the book. This resulting loss in sales should not be viewed as somehow negative in and of itself. In line with our understanding of robust speech protections, it seems appropriate that imperfect information, or information that can be improved upon, should experience lower distribution. In this case, the new information becomes the superior good and should, all things being equal, dominate the information market for that particular content.

Even common use of information exclusively on the distribution level should likewise be encouraged and protected. For example, a review simply criticizing the author's writing style is a critique that does not go to a content level interaction because the reviewer is not attempting to interact with the content by advancing or commenting on it per se. Rather, the review goes to the form of the information. Such common uses with pure distribution level effects are also desirable in the semicommons.

But allowing all common uses without restraint would irrevocably erode the underpinnings of the information copyright system. There is a point at which common use must yield to private use.²¹⁵ The semicommons theory provides us with a functional guide to such situations. When an original information good is substituted with an identical or nearly identical information good, the distribution of the original information

215. Recall that this Article assumes, neutrally, that private incentives are necessary.

good is impaired purely at the distribution level. According to the semicommons theory, we must not ask whether the information's private creator is losing opportunities to benefit from her information, but rather must consider whether the private information creators still have sufficient incentives to continue creating information. If they do, the common use should be allowed as a reflection of the understanding of the nature of private and common use interactions.

The presumption in favor of allowing the use is stronger still where content level interactions also take place. Even if many information producers would lose significant incentives to continue producing information due to common use at the content level, the common use should be allowed so long as there is sufficient incentive for private information users to continue to produce some information and so long as the quality of the perceived information good is perceived as increasing due to the content level interaction (for example, because false information is displaced with true information or because higher standards are used in information production).

In the Napster case study, this means that we would need to investigate claims that audio sales increased because of peer-to-peer file sharing. If theoretical or empirical evidence were to support this assertion, the common use should be allowed. Common users could share their files on the network without fear of displacing the incentives necessary to encourage that work to begin with. Such a result would be far different from the restrictions that were imposed on Napster, which precluded the trading of copyrighted works. Under a semicommons regime, the trading itself would be permitted through an understanding of the content and distribution level effects the trading has on the semicommons, and not simply what effect the common use has on the private user's rights.

VII. CONCLUSION

Henry Smith's explanation of the semicommons in the open fields provides a powerful theory that can be applied to information ownership. The semicommons form not only acknowledges this dynamic interaction, but also articulates the methods used to discourage the strategic behavior that is likely to arise in a semicommons. As sources of property rights, the statutory and judicial doctrines surrounding copyright, patent, trademark and trade secret play an integral role in the information semicommons. With methods in place to avoid strategic behavior—methods that adapt to changes in law and technology—semicommons theory allows us to under-

stand more fully the nature of an overall property system and to view the interactions in the system as related, as opposed to purely antagonistic.

As shown in the peer-to-peer example addressed in this Article, many arguments and discussions that are now rejected under the rubric of fair use should be given careful consideration when the information semicommons model is properly understood. Some of these are already included in the fair use doctrine, but others are not. As shown by the peer-to-peer example, the paradigm shift would produce valuable insights for decision making regarding information ownership.

The effect of acknowledging that information ownership forms a semicommons is perhaps radical. It requires that we move our notions of information ownership away from the land and asset-based model toward a model that recognizes and enhances the dynamic interaction between information uses. Such a step would, however, allow us to come to terms with the changing nature of information production in the digital age.

The antagonistic duality between private and common use is not reflective of the reality of information ownership. The semicommons is.

A RETURN TO LILLIPUT: THE *LICRA v. YAHOO!* CASE AND THE REGULATION OF ONLINE CONTENT IN THE WORLD MARKET

By Marc H. Greenberg[†]

ABSTRACT

Over the past three years, a see saw battle has raged in Paris, France and in the heart of Silicon Valley in Santa Clara County, California, over the regulation of content on the Internet. The arena for this battle is the case of *LICRA v. Yahoo!*, which pits two non-profit human rights groups in France against giant Internet search engine and information portal Yahoo!, Inc. (“Yahoo”). The issues are (1) whether Yahoo may be prosecuted in France under French law for maintaining both auction sites that sell Nazi-related items and information sites promoting Nazi doctrine and (2) whether U.S. courts should enforce the resulting judgment.

The first section of this Article presents the laws governing Internet content providers and the jurisdictional regime that gave rise to this see saw battle. The second section examines a series of court proceedings. The first two proceedings in France in 2000 resulted in a French court order directing Yahoo to add geo-location filtering software to its servers in Santa Clara. The subsequent California district court litigation filed in 2001 resulted in summary judgment for Yahoo. This judgment is on appeal. The third and final section explores the global implications of the French and U.S. proceedings. The section concludes that the international community should restructure certain principles governing international jurisdiction in Internet cases and adopt shared guidelines on online content available to the world market. These changes would promote the principle of international comity while allowing the Internet to retain most of its unique, borderless nature. Without such changes we may, like the people of Lilliput and Blefuscu in Jonathan Swift’s

© 2002 Marc H. Greenberg

† Marc H. Greenberg is an Associate Professor of Law and the Director of the LL.M. Program in Intellectual Property Law at Golden Gate University School of Law in San Francisco, California. The author thanks his colleagues on the faculty, Clifford Rechtschaffen, Helen Hartnell, Helen Chang, Bernard Segal, Christian Okeke, and Sompong Sucharitkul for their help in shaping this Article and the concepts expressed. The author also thanks student research assistants Jonathan Bruce, Marie Galanti, and Juna Kim for their invaluable assistance. This Article is dedicated to Kim, without whose patience it would never have been completed.

Gulliver's Travels, be locked in senseless conflict for years over which end of the egg we should break, instead of developing the tremendous potential of the Internet as a means for truly global communication.

TABLE OF CONTENTS

I.	BACKGROUND AND INTRODUCTION.....	1192
A.	Yahoo and Its Auction Pages	1194
B.	European Laws Making the Promotion of Nazism a Hate Crime	1196
C.	Whose Content Prohibitions Govern? Which End of the Egg is the Right End to Break?.....	1197
D.	The Legal Dilemma: Jurisdiction and Enforceability on the Worldwide Internet	1199
II.	CROSSING THE POND—THE LAWSUIT IN FRANCE, THE FRENCH COURT'S INTERIM ORDERS, AND THE ENSUING U.S. LITIGATION.....	1205
A.	The Opening Salvo: <i>LICRA et UEJF v. Yahoo and Yahoo France</i>	1206
1.	<i>The Source of Conflict: The Yahoo U.S. Auction Sites</i>	1206
2.	<i>The May 22, 2000 Interim Decision of the French Court</i>	1207
a)	The French Court's Application of the "Effects Test" to the Jurisdiction Issue	1208
b)	The "Technical Measures" Ordered by the French Court	1209
3.	<i>The International Response to the May 22 Order</i>	1210
4.	<i>The French Court's Interim Order of November 20, 2000—The Verdict of the Experts</i>	1213
5.	<i>The International Response to the November Hearing and Order</i>	1218
B.	Yahoo v. Licra—The United States Litigation and the Question of Enforceability	1227
1.	<i>Yahoo Fights Back—The U.S. District Court Action</i>	1227
2.	<i>The Ruling Denying LICRA's Jurisdictional Motion to Dismiss</i>	1230
a)	Case or Controversy	1231
b)	Jurisdiction	1235
3.	<i>The Ruling Granting Yahoo's Motion For Summary Judgment</i>	1241
C.	The Real Issue: It's Their Law Against Ours	1247
III.	A MODEST PROPOSAL FOR INTERNATIONAL GUIDELINES, AND A FORMULA FOR NOTICE AND A REDUCTION OF UNCERTAINTY	1250
A.	The Movement Towards an International Treaty Regulating Jurisdiction and Enforcement of Judgments	1250
B.	Towards an Interim Solution: Effects Plus Targeting Plus Increased Foreseeability—Breaking the Egg in the Middle.....	1253
IV.	CONCLUSION	1257

I. BACKGROUND AND INTRODUCTION

The battle began when the League Against Racism and Anti-Semitism (Ligue Contre La Racisme Et L'Antisémitisme, or "LICRA") filed a lawsuit against Yahoo in France. LICRA, based in Paris, sought enforcement of French laws making the offering of Nazi memorabilia a hate crime, claiming Yahoo violated French law by allowing Yahoo users

to offer Nazi-related items for sale. Yahoo contested the jurisdiction of the Paris court, but the court rejected this, finding jurisdiction on the theory that Yahoo's conduct caused harm in France, thereby justifying the exercise of jurisdiction.

After losing a preliminary decision on the merits in the French court, Yahoo responded by filing an action for declaratory relief in U.S. District Court in California.¹ Yahoo ultimately obtained a ruling that the Paris court's order directing Yahoo to install geo-location filtering software in its California-based servers violated Yahoo's First Amendment rights and was therefore unenforceable.²

The French and U.S. cases raise a number of significant issues. Jurisdictional issues raised include the extent to which the "effects test," typically applied to establish jurisdiction in tort cases, applies in web content cases. If courts use this test,³ should they modify it to require evidence of targeting or other grounds (the so-called "effects-plus test") before finding jurisdiction? Finally, should the filing of a judgment obtained in one country by a plaintiff, without subsequent enforcement efforts, be a sufficient contact with the forum of the defendant's country to confer jurisdiction over that plaintiff in a declaratory relief action filed in the defendant's country?

Technological issues raised include the potential role of geo-location filtering. Is geo-location filtering a flawed technical measure offering no substantive assistance to countries seeking to enforce their laws over the "borderless" Internet, or can it afford a means for those countries to maintain their cultural values and mores without seeking to impose them on the rest of the world?

A broader question these cases pose is whether litigation in multiple fora all over the world is the best way to resolve the international disputes that are arising with increasing frequency over the clash between web content and local laws. This Article suggests that the litigation route is fruitless and endlessly draining of valuable resources, and proposes an alternative approach: provide notice to website hosts of the possible liability their content may expose them to under other countries' laws and develop international guidelines, or possibly treaties, addressing the regulation of online content.

1. See *Yahoo!, Inc. v. La Ligue Contre Le Racisme et L'Antisemitisme*, 145 F. Supp. 2d 1168, 1171 (N.D. Cal. 2001) [hereinafter *Yahoo! I*].

2. *Yahoo!, Inc. v. La Ligue Contre Le Racisme et L'Antisemitisme*, 169 F. Supp. 2d 1181, 1194 (N.D. Cal. 2001) [hereinafter *Yahoo! II*].

3. See, e.g., *Yahoo! I*, 145 F. Supp. 2d at 1173-74 (utilizing the "effects test").

A. Yahoo and Its Auction Pages

Yahoo is a Santa Clara, California based Internet communications, commerce, and media company.⁴ A leading search engine for the Internet, Yahoo reports reaching more than 232 million web surfers per month.⁵ The company hosts more than twenty-five websites all over the world, with an extensive database of content ranging from stock information, to reproductions of political cartoons, to listings for local movie theaters. The French site (“Yahoo France”) is located at the Universal Resource Locator (“URL”) <http://fr.Yahoo.com>.

Yahoo’s regional sites target local users, but users from other parts of the world can access these sites through links on other Yahoo pages or through the URL. Presented in the language of their host countries, the regional sites generally observe the laws of those countries for content under their control. Users of the U.S. sites and the regional sites can access any Yahoo site by clicking on a link on each site that allows them to view instantly the contents of the desired site. Each site contains ways for people to interact online, including chat rooms, auction pages, shopping pages, e-mail services, and clubs users can join.

The auction pages on each of Yahoo’s twenty-one regional sites are accessible to all net surfers. Anyone over the age of eighteen may list an item for sale under a wide range of categories. Yahoo records the posting and sends an e-mail to the seller detailing the highest bid and the buyer’s contact information. The parties then complete their transaction without further involvement by Yahoo.⁶

Although Yahoo does not regulate sales terms for transactions on its auction sites, Yahoo monitors the U.S. Yahoo auction sites for compliance with U.S. copyright laws⁷ and warns auction sellers that they may not offer for sale goods or services that violate U.S. laws, such as the sale of stolen goods, body parts, prescription and illegal drugs, weapons, or goods

4. Yahoo! Inc. was founded in 1994 by two Ph.D. candidates in electrical engineering at Stanford University as a way to keep track of their favorite websites on the Internet. Originally named “Jerry’s Guide to the World Wide Web” after co-founder Jerry Yang, the founders shortly thereafter changed the name to Yahoo!. The Yahoo! name is an acronym for “Yet Another Hierarchical Official Oracle,” but the co-founders insist they chose the name because they liked its dictionary definition as meaning a person who is “rude, unsophisticated, uncouth.” Yahoo! Media Relations, *The History of Yahoo—How It All Started*, at <http://docs.yahoo.com/info/misc/history.html> (last visited Oct. 14, 2002).

5. *Id.*

6. Yahoo! Shopping Auctions, *Shopping Auctions Guidelines*, at <http://user.auctions.shopping.yahoo.com/html/guidelines.html> (last visited Oct. 14, 2002).

7. *Id.*

that would violate the Iranian or Cuban embargoes.⁸ Yahoo also advises sellers that they may not offer items to buyers in jurisdictions that prohibit the sale of those items.⁹ Yahoo polices only for copyright law and places the burden of complying with other laws on users—specifically sellers.

Proprietary content sites such as AOL, Prodigy, and Yahoo have struggled with the issue of how much control to exercise over the content they provide to their members and subscribers. Prodigy, for instance, initially represented that it would police the appropriateness of content on its site to attract family use, in effect guaranteeing that the content would be safe for all ages to see. This policy earned Prodigy liability for damages in the much-publicized case of *Stratton Oakmont, Inc. v. Prodigy Services Co.*¹⁰ In the post-*Stratton* era, content-rich sites have steered away from attempting to control all material viewers see, opting instead for the safe-harbor protections afforded to them as passive Internet Service Providers (“ISPs”) under the terms of the Digital Millennium Copyright Act.¹¹

According to *Stratton*, Internet providers that offer services to subscribers and members de facto exercise some control over their activities, since they have the power to do so.¹² Yahoo exercises this control through the Terms of Service it imposes on its members.¹³ These terms prohibit members from using the Yahoo service to “intentionally or unintentionally violate any applicable local, state, national or international law.”¹⁴ This provision shifts the burden of determining whether content

8. *Id.*

9. *Id.*

10. No. 31063/94, 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995).

11. 17 U.S.C. § 512 (2000).

12. *Stratton*, 1995 WL 323710, at *3.

13. *Yahoo! Terms of Service*, at <http://docs.yahoo.com/info/terms> (last visited Oct. 14, 2002). As noted previously, Yahoo offers membership free of charge. Members receive certain benefits, including personalized content on their home page. In exchange, Yahoo receives demographic information on its members, advertising revenues, and other benefits.

14. *Id.* ¶ 6. The use of the term “national law” may be confusing. Does this mean U.S. law, or the law of other nations? This paragraph also raises a central question in this entire dispute: to what extent has a member violated these rules by posting content lawful in the United States, but not lawful in another country? The Terms of Service are ambiguous on this point. The next term provides the following sentences, which neither shed much light on the issue, nor provide much guidance for the members: “Recognizing the global nature of the Internet, you agree to comply with all local rules regarding online conduct and acceptable Content. Specifically, you agree to comply with all applicable laws regarding the transmission of technical data exported from the United States or the country in which you reside.” *Id.* ¶ 7. The first sentence’s meaning is unclear, because the agreement neglects to define “local rules.” The second sentence is unrelated to the first—

violates “any applicable local, state, national or international law” to the members who post it—a seemingly Herculean task, difficult even for Yahoo itself to comply with, as we shall see. Through its Terms of Service, Yahoo restricts what members can post or upload onto its service overall (not just the auction sites). Restrictions include prohibitions against defamatory statements and material that breaches fiduciary duties or violates intellectual property law.¹⁵

B. European Laws Making the Promotion of Nazism a Hate Crime

The ease with which information can be disseminated via the Internet inexpensively to a large audience has, in the view of some commentators, sparked a tremendous growth in hate group advocacy for racial supremacy, religious discrimination, and other fringe group views.¹⁶ While in the United States the First Amendment protects a fairly broad spectrum of hate speech,¹⁷ a different set of limits applies to certain kinds of speech in other countries, including France—and it is this difference that triggered the dispute in the *LICRA v. Yahoo!* cases.

France, Germany, the Netherlands, and other countries have laws that make it a crime to exhibit for sale objects relating to Nazism and the Third Reich.¹⁸ Judge Jean-Jacques Gomez, the judge who presided over the *LICRA v. Yahoo!* case in Paris, France, explained the rationale behind the French statute:

Whereas the exhibition of Nazi objects for purposes of sale constitutes a violation of French law (Section R.645-2 of the Criminal Code), and even more an affront to the collective memory of a country profoundly traumatised by the atrocities

what does the global nature of the Internet have to do with the applicable laws relating to transmission of a specific kind (technical) of data?

15. *Id.* ¶ 7.

16. Lisa Guernsey, *Mainstream Sites Serve as Portals to Hate*, N.Y. TIMES, Nov. 30, 2000, available at <http://www.nytimes.com/2000/11/30/technology/30HATE.html>. According to the Simon Wiesenthal Center, there are more than 2000 groups with sites on the Internet promoting anti-Semitism or white supremacy. *Id.*

17. *See, e.g.*, *Collin v. Smith*, 578 F.2d 1197 (7th Cir. 1978).

18. *See, e.g.*, CODE PENAL [C. PÉN.] art. R.645-1 (Fr.), translation available at <http://www.lex2k.org/yahoo/art645.pdf> (last visited Dec. 4, 2003); § 130(1)(3) STRAFGESETZBUCH [StGB] (F.R.G.). Section 130(3) of the Federal Criminal Code in Germany provides that “Imprisonment, not exceeding five years, or a fine, will be the punishment for whoever, in public or in an assembly, approves, denies or minimizes {the Holocaust} committed under National Socialism, in a manner which is liable to disturb the public peace.” § 130(1)(3) StGB (F.R.G.).

committed by and in the name of the criminal Nazi regime against its citizens and above all against its citizens of the Jewish faith¹⁹

There has been no challenge presented in any proceedings claiming that France, Germany, the Netherlands, or any other countries with similar hate crime laws lack either the political or moral right to enact those statutes. However, the unique nature of the Internet has given birth to a controversy unforeseen when these laws were enacted: what is the responsibility of national legal systems when online content is posted in one country where it is legal but can be viewed in another country where the same content is illegal?

C. Whose Content Prohibitions Govern? Which End of the Egg is the Right End to Break?

The pointless struggle of nations over their cultural differences, and the effort to impose their laws on each other in an effort to protect those differences rather than seeking means to preserve them without conflict, is an old and regrettably venerated tradition. Literature is replete with examples of the destructive nature of such endless fights.

In Jonathan Swift's brilliant eighteenth century satire *Gulliver's Travels*,²⁰ Captain Lemuel Gulliver is shipwrecked and washes ashore on the island of Lilliput. The people of Lilliput are physically tiny, and to them, Gulliver appears to be a giant. After assuaging the initial fears the Lilliputians have towards him, the King of Lilliput enlists Gulliver's help to resolve a bitter war that has raged for "six and thirty moons"²¹ against the people of a neighboring island, Blefescu.

The conflict's source, the King explains, is that the two countries, once on friendly terms, became bitter enemies over a dispute regarding which end of an egg should be broken before it is consumed. The Lilliputian view is that the small end of the egg should be broken, and the Blefescu view is that the large end is the one to break. The conflict's consequences are summarized as follows:

19. Pl.'s Compl. for Decl. Relief, ex. A, at 5, *Yahoo!, Inc. v. La Ligue Contre Le Racisme et L'Antisémitisme*, 169 F. Supp. 2d 1181 (N.D. Cal. 2001) (No. 00-21275), available at <http://www.cdt.org/speech/international/001221yahoo.complaint.pdf> (last visited Dec. 2, 2003).

20. JONATHAN SWIFT, *GULLIVER'S TRAVELS AND OTHER WRITINGS* (Louis A. Landa ed., The Riverside Press 1960) (1726).

21. *Id.* at 39.

It is computed, that eleven thousand persons have, at several times, suffered death, rather than submit to break their eggs at the smaller end. Many hundred large volumes have been published upon this controversy: but the books of the Big-Endians have been long forbidden, and the whole party rendered incapable by law of holding employments.²²

Swift used satire to point out the absurdity and terrible consequences of the political, religious, and economic conflicts between Ireland and England, in terms of trade and agriculture, all of which he saw as being unfairly controlled by England. As a clergyman loyal both to his native home of Ireland and to the Church of England, he sought to reform and bring an end to this pointless and damaging conflict.²³ Swift hoped that by illustrating that the two peoples of Lilliput and Blefuscu were foolish to engage in a terrible and costly war over a matter of insignificance (which end of the egg to break) when their common goal was in fact the same (to eat the egg), that he could encourage the English and the Irish to put aside their differences and strive for the common good (enough eggs in famine-stricken Ireland to feed them all).

The goal of this Article, with apologies to Swift, is similarly to urge the combatants in the Internet content regulation disputes to refrain from endless litigation in disjunctive legal systems, which results in stalemates and unenforceable judgments, and instead to look to means by which to achieve the common goal: an Internet that allows for the free exchange of information in the global marketplace. The United States, France, and perhaps other countries have a common goal, which may extend to a more general interest in international comity. France and the United States share a long history of support for a free press and democratic principles in governance.²⁴ Both countries recognize that the Internet's capacity for the free flow of information and opinions represents a quantum leap forward in advancing the goal of an expanded range of exchange in contemporary society.²⁵ Both countries also recognize that free speech needs to be limited in some respects to preserve domestic and international peace, and to respect and preserve the cultural values of each country.²⁶ Where the

22. *Id.*

23. *Id.* at xviii-xxvi.

24. See Julian Mailland, *Freedom of Speech, The Internet, and The Costs of Control: The French Example*, 33 N.Y.U. J. INT'L L. & POL. 1179, 1184 (2001).

25. See Lyombe Eko, *Many Spiders, One Worldwide Web: Towards a Typology of Internet Regulation*, 6 COMM. L. & POL. 445, 477-81 (2001).

26. See generally *id.* at 468-70; Mailland, *supra* note 24; Angela E. Wu, *Spinning a Tighter Web: The First Amendment and Internet Regulation*, 17 N. ILL. U. L. REV. 263 (1997).

countries occasionally differ is the emphasis each places in that regulatory scheme—differences likely resulting from distinct sets of historical experiences and societal pressures. The challenge for countries like these is to find a way to allow the maximum amount of freedom in communication and exchange, while at the same time not permitting that exchange to violate those local cultural values and laws and to find that balance in a way that encourages the expansion of the doctrine of comity among nations.

D. The Legal Dilemma: Jurisdiction and Enforceability on the Worldwide Internet

One of the challenges of the Internet is determining whether we can solve a particular problem by applying existing legal principals to the digital realm, or whether only the creation of new legal doctrine will suffice. The interaction between the Internet and the traditional legal doctrines of jurisdiction, national and international, has been the subject of considerable debate, both in case law and in scholarly works. The *LICRA v. Yahoo!* cases show how far we have to go in resolving how jurisdictional principles do or should work on the international Internet and illustrate the importance of addressing these issues as a global community.

Traditional jurisdiction in the United States was articulated definitively in the seminal 1877 decision of *Pennoyer v. Neff*.²⁷ The U.S. Supreme Court ruled that for a court to assert personal jurisdiction over an individual, that person must be present physically in that court's state when served.²⁸ By the middle of the twentieth century, changes in technology mandated a less stringent standard. It had now become possible to do business by catalog sales, by telephone or telegraph, and commercial aviation and the development of the national highway system meant that interstate travel was no longer an onerous burden.

The Supreme Court recognized the changes wrought by increased interstate contact in *International Shoe v. Washington*.²⁹ Here, the Court extended states' jurisdictional reach with a two-part test to determine whether jurisdiction could be asserted against an out-of-state defendant: (1) whether the defendant had sufficient "minimum contacts" with the forum state to justify the exercise of jurisdiction,³⁰ and (2) whether allowing the defendant to be sued in the forum state would offend

27. 95 U.S. 714 (1877).

28. *Id.* at 722.

29. 326 U.S. 310, 316 (1945).

30. *Id.* at 319.

“traditional notions of fair play and substantial justice.”³¹ Both parts of the test must be satisfied before a case can proceed.

Typically, courts look to the long-arm statutes of a particular state to determine what the state views as sufficient “minimum contacts” to invoke jurisdiction. This inquiry has led to the development of the doctrine of “purposeful availment.” In applying the doctrine, the court focuses on whether a defendant has initiated contacts and expects protection by a given forum’s laws for the defendant’s commercial or personal benefit.³² If a showing of minimum contacts can be established, courts next look to the reasonableness of the exercise of jurisdiction against a nonresident defendant, considering the interests of the state in protecting the rights granted to its citizens, the burden on the plaintiff if denied the choice of venue, and the burden on the defendant in coming to a foreign jurisdiction to defend against the action.³³ Mere inconvenience or added expense will not be sufficient to overcome a court’s decision to exercise jurisdiction; the inconvenience must be severe and the expense significant to warrant denial of jurisdiction on the fairness ground.³⁴

The question then becomes what constitutes a reasonable exercise of jurisdiction over the global Internet. Professor Michael Geist has argued that a “foreseeability metric lies at the heart of the reasonableness standard,” which he defines as meaning that “a party should only be haled into a foreign court where it was foreseeable that such an eventuality might occur.”³⁵ He asserts that the “borderless Internet” and the worldwide availability of the Internet makes foreseeability very difficult, and creates instead an “all or nothing environment in which either every jurisdiction is foreseeable or none is foreseeable.”³⁶

Within the United States, online transactions via the Internet and websites established to provide information and/or goods have challenged the “purposeful availment” and “minimum contacts” tests of *International Shoe* and its progeny, since these transactions and sites do not actually “enter” into a specific jurisdiction. A California-based website advertising goods for sale across the nation would not, absent any other evidence of activity in a given state, meet the test of minimum contacts and purposeful availment, and jurisdiction would be denied. As cases dealing with these

31. *Id.* at 320.

32. *See, e.g.*, *Burger King Corp. v. Rudzewicz*, 471 U.S. 462, 476-78 (1985).

33. *Id.*

34. *Id.*

35. Michael A. Geist, *Is There a There There? Toward Greater Certainty for Internet Jurisdiction*, 16 BERKELEY TECH. L.J. 1345, 1356 (2001).

36. *Id.* at 1356-57.

issues began to emerge, it became apparent that courts needed a different jurisdictional model for Internet-based activity.

The model evolved into a “passive versus active” test based on the nature of the website involved. If the site required a high degree of interaction between site host and visitor, the visitor who became a plaintiff would be able to assert jurisdiction against the host/defendant.³⁷ A purely passive, information-only site, by contrast, would generally be outside of the jurisdictional reach of the plaintiff’s home state.³⁸

Initially, courts applied the passive versus active test as a refinement of traditional jurisdiction that recognized the unique characteristics of Internet commerce. One of the earliest decisions applying the test was *Bensusan Restaurant Corp. v. King*.³⁹ Using the *Bensusan* case as a guide, a federal district court in Pennsylvania elaborated on the passive/active test in the decision in *Zippo Manufacturing Co. v. Zippo Dot Com, Inc.*⁴⁰ This case presented a claim of trademark infringement and dilution based on the use of a trademarked name on a website. The plaintiff, Zippo Manufacturing, based in Pennsylvania, manufactured the famous “Zippo” cigarette lighters. The defendant, Zippo Dot Com, was an Internet news service located in California that had registered the domain name “Zippo.com” for its service.⁴¹ Through its online presence, Zippo.com attracted subscribers to its news service from all over the country. At least 3000 of those subscribers came from Pennsylvania, all of whom found

37. See *Bensusan Rest. Corp. v. King*, 126 F.3d 25, 27-29 (2d Cir. 1997); *Zippo Mfg. Co. v. Zippo Dot Com, Inc.*, 952 F. Supp. 1119, 1121 (W.D. Pa. 1997).

38. *Zippo*, 952 F. Supp. at 1121.

39. 126 F.3d. 25 (2d Cir. 1997). The Blue Note was a nightclub in Columbia, Missouri. Club owner Richard King promoted the musical performances at his club through a Web site which posted the club’s calendar, and gave ticket purchasing information. To avoid confusion between his club and the historic New York club of the same name, the Missouri club’s Web site contained the following disclaimer: “The Blue Note’s CyberSpot should not be confused with one of the world’s finest jazz club[s], [the] Blue Note, located in the heart of New York’s Greenwich Village. If you should ever find yourself in the big apple [sic] give them a visit.” *Id.* at 27. The proprietors of the Blue Note in New York, the Bensusan Restaurant Corporation, were not mollified by this disclaimer. As holders of a federal trademark registration in the Blue Note name, they brought a trademark infringement action in New York federal court. *Id.* The case presented the court with the dilemma of imposing jurisdiction of the federal court in New York over a Missouri club owner whose only presence in the New York district was a virtual one, only manifesting itself through the fact that a Web surfer in New York could access the Missouri club’s site online. *Id.* at 29. The court noted that King’s Web site was passive in nature and his business operations were “of a local character.” *Id.* The court found that jurisdiction would not lie for this level of passivity. *Id.*

40. 952 F. Supp. 1119 (W.D. Pa. 1997).

41. *Id.* at 1121.

their way to the site despite the fact that Zippo.com had no physical presence in the state—no employees or agents, and no office.⁴²

In analyzing the jurisdiction issue, the court opined that there are three different types of web presence, and found that the degree of interactivity on a given site would determine whether it was reasonable to impose jurisdiction. At one extreme was the situation in *CompuServe, Inc. v. Patterson*,⁴³ where the court found jurisdiction in a trademark infringement case brought by plaintiff CompuServe. The court found that defendant Patterson had availed himself of the benefits of the laws of the state of Ohio, where CompuServe was based, by entering into a shareware agreement with a party based in Ohio, by sending shareware electronically to CompuServe, and by advertising the availability of that shareware on CompuServe.⁴⁴ At the other extreme for the court was the pure information site, illustrated in *Bensusan* by a jazz nightclub's website that did not sell tickets or merchandise online, by phone, or by mail. The court in *Bensusan* declined to extend jurisdiction to the website's owner. The *Zippo* court followed this application of the test and reaffirmed that the mere fact that a website contained information about a business's activities would not confer jurisdiction.⁴⁵

The court in *Zippo* then expanded the definition of these passive and active sites by noting that there was a third group to consider between the two ends of the spectrum: interactive websites in which, even though there is no solicitation from the host computer, there is interaction in the form of an exchange of information. Depending on the extent of the exchange and the commercial nature of the information exchanged, the court held that jurisdiction might or might not be imposed.⁴⁶ The court found that the extent of interaction between Zippo.com and its customers in Pennsylvania was closer in its conduct to the defendant in *CompuServe* than the defendant in *Bensusan*, and applying a similar analysis of the contacts, concluded similarly that jurisdiction was proper against Zippo.com.⁴⁷

42. *Id.*

43. 89 F.3d 1257 (6th Cir. 1996).

44. *Id.* at 1265-66.

45. *Zippo*, 952 F. Supp. at 1124 (citing *Bensusan Rest. Corp. v. King*, 937 F. Supp. 295 (S.D.N.Y. 1996), *aff'd*, 126 F.3d 25 (2d Cir. 1997)).

46. *Id.*

47. *Id.* at 1125-27.

A number of decisions in U.S. courts have followed the passive/active approach over the past few years.⁴⁸ However, courts and scholars have begun to note some limitations of this approach.⁴⁹ One area of difficulty for courts trying to apply the *Zippo* test was for cases dealing with defamation on the web. A purely passive site published in California that, for example, defames former New York City Mayor Rudolph Guiliani would not, under the *Zippo* test, allow a New York court to impose jurisdiction. The fact that the site could be seen in New York, and could harm Guiliani's reputation in New York, would not be enough for the court to find jurisdiction if the host site had no business activity in New York, no offices, employees or agents, and did not solicit any contact with New York citizens. The worldwide availability of Internet access, coupled with the anonymity afforded to persons posting material online, seemed like an open invitation to spread lies and defamatory content with no fear of liability for damages.

To remedy this oversight, courts looked for guidance in earlier decisions, and found it in *Calder v. Jones*, a 1984 decision by the U.S. Supreme Court employing the so-called "effects test" to establish jurisdiction for a web-based tort claim.⁵⁰ In *Calder*, an entertainer resident in California brought a libel case against a publisher in Florida.⁵¹ The case was filed in a California superior court, and the defendants objected to the jurisdiction of the court.⁵² Instead of focusing on the defendants' contacts with the forum in order to apply the *International Shoe* criteria, the Court, however, concentrated on the effects of their actions.⁵³ The Court ruled that in such cases, the "effects doctrine" required the Court to find personal jurisdiction properly imposed when the following four elements are present: a) the defendant's tortious actions, b) expressly aimed at the forum state, c) cause harm to the plaintiff in the forum state, d) which the defendant knows is likely to be suffered.⁵⁴ Applying this test, the Court found that plaintiff Jones suffered injury to her professional reputation in California, that she lived in California, and that she suffered emotional

48. See, e.g., *Cybersell, Inc. v. Cybersell, Inc.*, 130 F.3d 414 (9th Cir. 1997); *Decker v. Circus Circus Hotel*, 49 F. Supp. 2d 743 (D.N.J. 1999); *Blumenthal v. Drudge*, 992 F. Supp. 44 (D.D.C. 1998).

49. See, e.g., *Panavision Int'l, L.P. v. Toeppen*, 141 F.3d 1316, 1320 (9th Cir. 1998); *Bochan v. La Fontaine*, 68 F. Supp. 2d 692, 701-02 (E.D. Va. 1999).

50. 465 U.S. 783 (1984).

51. *Id.* at 784, 785.

52. *Id.* at 784.

53. *Id.* at 789.

54. *Id.*

distress arising from the defamatory conduct in California.⁵⁵ Based on those reasons, the Court found that the defendant had intentionally targeted a California resident for harm and was therefore subject to the jurisdiction of the California courts.⁵⁶

Courts have applied the effects test doctrine, focusing on where the harm occurs, in a number of Internet cases.⁵⁷ Not all of the cases have been based on tort claims; they have also encompassed other forms of business activity, usually in the context of unfair competition or trademark or copyright claims.⁵⁸ Notably, in all these cases the conduct at issue (i.e., tortious actions, trademark, or copyright infringement) was unlawful in both the plaintiff's and the defendant's home states or countries. It is not surprising, consequently, that the question of the validity of the doctrine was never raised in the context of attempting to apply it in a case where the law in one forum would allow the conduct at issue, whereas the law in another forum would make that same conduct illegal.

Internet interaction makes the jurisdictional question of whose law to apply pressing, as the conduct at issue may be simultaneously lawful and unlawful depending on where in the world the parties post to and/or access the web. Traditionally, the effects test made jurisdiction and hence the choice of law clear: usually some item containing the offending conduct physically made its way across a border where it caused harm according to local law—for instance, an obscene film. In this case, the effects test established jurisdiction at the item's physical location, leaving it to the local court to decide which law to apply. Material posted on the Internet, however, is simultaneously available all over the world, making it harder to prove intent to do harm in a particular jurisdiction, yet easier to show an effect in all jurisdictions around the world. Efforts to regulate content in this context spark claims, such as what Yahoo advanced in this instance, that the only means available to comply with all countries' laws is to remove the content entirely, a course that then gives rise to claims of

55. *Id.* at 789-90.

56. *Id.* at 791.

57. *See generally* Geist, *supra* note 35, at 1360–80 (discussing the “effects test” on Internet-related cases).

58. *See, e.g., id.* at 1373-76 (citing court decisions finding jurisdiction under the effects test in *Nissan Motor Co. v. Nissan Computer Corp.*, 89 F. Supp. 2d 1154 (C.D. Cal. 2000), and *Euromarket Designs Inc. v. Crate & Barrel Ltd.*, 96 F. Supp. 2d 824 (N.D. Ill. 2000), and court cases finding insufficient evidence to support application of the test in *People Solutions, Inc. v. People Solutions, Inc.*, No. 3:99-CV-2339-L, 2000 WL 1030619 (N.D. Tex. Jul. 25, 2000), and *Winfield Collection, Ltd. v. McCauley*, 105 F. Supp. 2d 746, 751 (E.D. Mich. 2000)).

ensorship and that the defendant is being forced to accede to the lowest denominator in the regulation of information it distributes.

Like the United States, France has adopted a jurisdictional test similar in nature to the “effects test,” allowing its courts to exercise jurisdiction over defendants who cause harm within the borders of France regardless of the defendant’s place of residence.⁵⁹ The Internet makes this principle necessary to preclude defendants from locating outside a country’s borders for the purpose of avoiding jurisdiction. United States courts have been fighting this battle on a number of fronts in recent years, in efforts to prevent offshore website owners from providing online gambling opportunities in violation of U.S. federal or state laws,⁶⁰ or from posting software systems that allow parties to circumvent copy-protection systems in violation of Section 1201 of the Digital Millennium Copyright Act.⁶¹

A significant issue not addressed in these jurisdictional battles is whether, after the dust settles on these conflicts, any lasting change is actually accomplished due to the remaining problem of enforceability. Little will be accomplished if the boundaries of jurisdiction are expanded without concurrently expanding the ability to enforce judgments obtained in those foreign courts.

II. CROSSING THE POND—THE LAWSUIT IN FRANCE, THE FRENCH COURT’S INTERIM ORDERS, AND THE ENSUING U.S. LITIGATION

The following history of the dispute between the French plaintiffs and Yahoo serves as a sad illustration of the inability of the litigation process, either in France or in the United States, to deal with the complex cultural and legal issues that arise when material posted lawfully on servers in one country violates the law when viewed by web surfers in another country. The courts in each country attempt to walk the fine line between preserving their sovereignty and preserving the principle of international comity. The results are less than satisfying on all sides. Perhaps the most disappointing element of this dispute is that after more than three years of litigation, the parties are no better off than when they started, and the

59. See Pl.’s Compl. for Decl. Relief, ex. A, Yahoo!, Inc. v. La Ligue Contre Le Racisme et L’Antisémitisme, 169 F. Supp. 2d 1181 (N.D. Cal. 2001) (No. 00-21275), available at <http://www.cdt.org/speech/international/001221yahoo.complaint.pdf> (last visited Dec. 2, 2003).

60. See *infra* note 177.

61. See *United States v. Elcom Ltd.*, 203 F. Supp. 2d 1111 (N.D. Cal. May 8, 2002); see also discussion *infra* note 264.

issues they attempted to address in the litigation are still unresolved. Like the endless battles in Lilliput, the combatants seem unable to recognize that the larger goals they both share are disserved by their unrelenting positions.

A. The Opening Salvo: *LICRA et UEJF v. Yahoo and Yahoo France*

The battle began with LICRA's contention that the ease with which French web surfers could access auction pages on the Yahoo U.S. website where they could find Nazi memorabilia for sale was a clear violation of French criminal law. LICRA's unsuccessful efforts to convince Yahoo to remove the offending materials led to the initiation of litigation in a Paris court, raising questions of first impression over jurisdiction, technical measures, and the impact of the U.S. Constitution's First Amendment on the dispute.⁶² Finding jurisdiction, the French court ruled against Yahoo and issued orders for Yahoo to implement in France and California.⁶³ International flak soon followed.

1. The Source of Conflict: The Yahoo U.S. Auction Sites

In September 2000, visitors could find as many as 1500 Nazi and Third Reich related objects offered for sale on the Yahoo U.S. site.⁶⁴ These goods included Nazi uniforms, medals, and photographs, as well as literary works, such as Adolf Hitler's *Mein Kampf*, and *The Protocols of the Elders of Zion*, a notorious report produced by the Russian secret police in the early 1900's alleging a world wide Jewish conspiracy to foment corruption and sedition.

On April 5, 2000, LICRA sent a cease and desist letter to Yahoo in Santa Clara, California, advising the company that both the sale of the Nazi objects on the auction site and the promotion of pro-Nazi books and other written materials violated French law.⁶⁵ LICRA threatened to take

62. See Pl.'s Compl. for Decl. Relief, ex. A, at 4, *Yahoo!, Inc. v. La Ligue Contre Le Racisme et L'Antisémitisme*, 169 F. Supp. 2d. 1181 (N.D. Cal. 2001) (No. 00-21275), available at <http://www.cdt.org/speech/international/001221yahoo.complaint.pdf> (last visited Dec. 2, 2003).

63. See *id.* exs. A, B.

64. Yahoo! Shopping Auctions, at <http://search.auctions.yahoo.com/search/auc?p=mazo&alocale=1us&acc=us> (last viewed Sept. 18, 2000) (on file with author). Downloaded U.S. auction pages showed 1589 items total. Descriptions of some of the items included the following: "Nazi 25 yrs Service Medal LOOK!!," "Nazi War Songs & Marches 6 Tapes NEW!," "Nazis in Front of a Beautiful House," "Nazi Officers with Sweetie," "Hitler 20 Mint NH Stamps Nazi Germany."

65. *Yahoo! II*, 169 F. Supp. 2d. 1181, 1184 (N.D. Cal. 2001); Pl.'s Compl. for Decl. Relief ¶ 14, *Yahoo! II* (No. 00-21275).

legal action if Yahoo failed to take steps to prevent the sales and remove pro-Nazi content within eight days.⁶⁶ On April 10th, LICRA filed a complaint in the Superior Court of Paris against Yahoo U.S. and Yahoo France.⁶⁷ On April 20th, LICRA was joined by the Union of Jewish Students of France (Union Des Etudiants Juif De France or “UEJF”), which filed a similar claim in the same action.⁶⁸ Yahoo U.S. was served with process by a United States Marshal and retained counsel to appear especially on its behalf in the French court.⁶⁹

2. *The May 22, 2000 Interim Decision of the French Court*

On May 15, 2000, the Paris court heard arguments on an expedited emergency basis before a single judge, Judge Jean-Jacques Gomez, First Deputy Chief Justice of the Superior Court of Paris. All parties were represented by counsel. Yahoo objected to the expedited emergency review status of the proceeding and requested a trial before a full judicial panel.⁷⁰ The court rejected this objection and proceeded to take evidence in response to the request of the plaintiffs for a preliminary injunction.⁷¹

The French court received evidence from the parties and the Sheriff of Paris regarding the contents of the Yahoo U.S. auction site, and the ease with which French citizens could access the site by using the link on the Yahoo France site. While not disputing that the offer of these goods for sale violated French law, Yahoo argued that the French court had no jurisdiction because the goods were offered for sale within the United States, on the Yahoo site targeted for U.S. users, and further that any prohibition of the sales or barring of the posting of the written works (*Mein Kampf* and *The Protocols of the Elders of Zion*) would violate the First Amendment of the U.S. Constitution.⁷² Yahoo France also argued that it had no liability because the Yahoo France regional site did not contain any offers to buy or sell Nazi or Third Reich items.⁷³ Finally, Yahoo argued that it was technologically impossible to prevent French web surfers from accessing the auction sites at issue.⁷⁴

66. *Yahoo! II*, 169 F. Supp. 2d at 1184; Pl.’s Compl. ¶ 17, *Yahoo! II* (No. 00-21275).

67. Pl.’s Compl. for Decl. Relief ¶ 17, *Yahoo! II* (No. 00-21275).

68. *Id.* ¶ 18.

69. *Yahoo! II*, 169 F. Supp. 2d at 1184.

70. Pl.’s Compl. for Decl. Relief ¶ 20, *Yahoo! II* (No. 00-21275).

71. *Id.*

72. *Id.*

73. *Id.*

74. *Id.*

After considering this evidence Judge Gomez, on May 22, 2000, issued his first preliminary order (“May 22 Order”). The order addressed the various objections raised by Yahoo, and found that none of them precluded the court from issuing a preliminary injunction.⁷⁵ Because the reasoning Judge Gomez applied in this first order would come to frame the issues in the subsequent proceedings, both in France and in the United States, it therefore warrants close examination.

a) The French Court’s Application of the “Effects Test” to the Jurisdiction Issue

The threshold issue for Judge Gomez to decide was whether the French court had jurisdiction over Yahoo. Applying French law, which embodies an approach similar to the “effects test” used in the United States, Judge Gomez produced a jurisdictional analysis virtually identical to that in the *Calder v. Jones* decision.⁷⁶ Noting that Yahoo had objected to the jurisdiction of the French court on the grounds that the alleged unlawful conduct was “committed on the territory of the United States,”⁷⁷ the court next considered the evidence of harmful effects in France. Judge Gomez asserted that there is no dispute that “surfers” (presumably web surfers) who link to Yahoo.com (the U.S. site) from French territory may see on their screens the pages that contain auction offerings of Nazi objects.⁷⁸ The court then reiterated the policy consideration behind the passage of French law Article R.645-2 of the Criminal Code by noting that the “exhibition of Nazi objects for purposes of sale,” in addition to being a violation of the Code, was also “an affront to the collective memory of a country profoundly traumatised by the atrocities committed [by the Nazis]”⁷⁹

Judge Gomez concluded that jurisdiction may be imposed based on the harm done in France, noting also his view that Yahoo’s actions were probably unintentional:

Whereas while permitting these objects to be viewed in France and allowing surfers located in France to participate in such a display of items for sale, the Company YAHOO! Inc. is therefore committing a wrong in the territory of France, a wrong whose unintentional character is averred but which has caused damage to be suffered by LICRA and UEJF, both of whom are

75. Pl.’s Compl. for Decl. Relief, ex. A, at 5-6, *Yahoo! II* (No. 00-21275).

76. *See supra* note 50 and accompanying text.

77. Pl.’s Compl. for Decl. Relief, ex. A, at 4, *Yahoo! II* (No. 00-21275).

78. *Id.* ex. A, at 5.

79. *Id.*

dedicated to combating all forms of promotion of Nazism in France, however insignificant the residual character of the disputed activity may be regarded in the context of the overall running of the auctions service offered in its Yahoo.com site;

Whereas, the damage being suffered in France, our jurisdiction is therefore competent to rule on the present dispute under Section 46 of the New Code of Civil Procedure⁸⁰

Having determined that the court had jurisdiction and that the evidence established that Yahoo had violated Article R.645-2 of the Criminal Code, Judge Gomez turned to the issue of remedies. These included orders directed both to Yahoo's French website located in France and ultimately to the home site of Yahoo in Santa Clara County, California.⁸¹

b) The "Technical Measures" Ordered by the French Court

In the May 22 Order, the French court's solution to the problem of unfettered access to content prohibited in France was a technological one: use code to forbid access to pages containing Nazi goods for all French users. Judge Gomez noted that Yahoo asserted in its defense that it was impossible for it to determine the national identity of people visiting its auction sites, which meant that such an order would have the effect of forcing Yahoo to remove the offending material from its site entirely.⁸² Asserting that Yahoo was in a position, in most cases, to verify the geographical origin of surfers visiting its auction pages based on the IP address of the Internet service provider,⁸³ the court acknowledged that for those surfers who accessed the web through portals that guaranteed their anonymity, Yahoo would have more difficulty in exercising control over what pages they could access, but also noted that control could still be exerted by limiting page access only to surfers who disclosed their geographical origin.⁸⁴ The court went on to assert that since Yahoo! Inc. was in a position, in most cases, to verify the geographical origin of surfers visiting its auction pages based on the IP address of the caller, it was in a position to prohibit surfers from France from viewing those pages

80. *Id.*

81. *Id.* ex. A, at 6-7.

82. *Id.* ex. A, at 4.

83. *Id.* ex. A, at 5.

84. *Id.* ex. A, at 6. This part of the Order apparently reflects the understanding held by the Court at that point in the proceedings that changes in the programming code for the site would be an effective way to regulate what content a given group of surfers could see. As will be shown, the Court later heard from experts that the use of code alone would yield only limited success in regulating access. *See* note 116 and accompanying text.

that violated Criminal Code Article R.645-2 by offering Nazi items for sale.⁸⁵ Judge Gomez concluded from this analysis that “the real difficulties encountered by Yahoo do not constitute insurmountable obstacles.”⁸⁶ He therefore ordered that Yahoo “[t]ake any and all measures of such kind as to dissuade and make impossible any consultations by surfers calling from France to its sites and services in dispute, . . . , especially the site selling Nazi objects.”⁸⁷

The court retained jurisdiction over these technical measures, gave Yahoo two months to “formulate any proposals in respect of technical measures capable of facilitating settlement of the present dispute,”⁸⁸ and set a second hearing date for July 24, 2000 for Yahoo to return and advise the court of the measures it was prepared to adopt.⁸⁹

At the second hearing, Yahoo advised the court that it “would be technically impossible for Yahoo! to comply with the May 22 Order,”⁹⁰ supporting this claim with a technical report prepared by its French Internet expert Jean-Denis Gorin, and also submitting a supplemental declaration from Yahoo engineer Geoff Ralston.⁹¹ Judge Gomez took the matter under submission and on August 11, 2000, ordered the formation of an expert panel to study the feasibility of compliance with the “technical measures” portion of the May 22 Order.⁹² He followed with another order on September 18, 2000, establishing the expert panel and requiring that the panel report back to the court with its findings by November 5, 2000.⁹³

3. *The International Response to the May 22 Order*

Judge Gomez’s May 22 Order stirred up a firestorm of controversy throughout the European Union and the United States, as commentators questioned the wisdom of such a solution, the French court’s authority, and the future of the Internet.

85. *Id.*

86. *Id.*

87. *Id.*, ex. A at 7. Judge Gomez also ordered Yahoo! France to post a warning on the French language site advising any surfers who used the site’s link to the U.S. Yahoo site that if they encountered material on the U.S. site pages that violated French law, they were to cease their review of those pages on penalty of the sanctions French law authorized for such violations. *Id.*

88. *Id.*, ex. A at 6.

89. *Id.*, ex. A at 7.

90. *Id.* ¶ 22.

91. *Id.*

92. *Id.* ¶ 23.

93. *Id.* ¶ 24. Judge Gomez allowed each party to designate an expert and chose a third expert himself. The panel included Francois Wallon from France, Ben Laurie from the United Kingdom, and Vinton Cerf from the United States. *Id.*

Philippe Guillanton, director general of Yahoo France, warned that the decision was very dangerous:

The whole question goes above Yahoo The point is whether we want to condemn the Internet to be closed in the same way that the media have traditionally been closed by frontiers This case could set a potentially dangerous precedent. . . . It is the first case where a judge in one country feels he is competent to decide over what actions he thinks an actor (in another country) should be taking.⁹⁴

Marc Knobel, spokesman for the plaintiffs, retorted, "If Yahoo doesn't respect the court's ruling, we'll press criminal charges against individuals such as Jerry Yang and perhaps the president of Yahoo France."⁹⁵ Spokespeople for the auction site eBay and online bookseller Amazon.com offered the alternative view that content regulation of the Internet was best left up to the community of users, in other words, the marketplace, leading one reporter to conclude, "The lack of international consensus on how to deal with conflicting legal issues might for the moment mean that the market ultimately decides the fate of these Internet companies."⁹⁶

Henry H. Perritt Jr.⁹⁷ offered the following summary of the problem in asserting that the French court was wrong to find jurisdiction:

The Yahoo case points up a dilemma in the law of jurisdiction If a web site is accessible to all, and is subject to jurisdiction by every nation on earth, then the laws of the lowest common denominator nation [will govern the Internet]. On the other hand, if we say that the only important law is the one where the content provider resides, then local values of foreign nations will not be enforced. We also run the risk of creating havens for shyster practices.⁹⁸

94. Reuters, *Yahoo Says French Ruling May Set Precedent*, CNET NEWS.COM, at <http://news.cnet.com/news/0-1007-200-1930850.html> (last viewed May 23, 2000).

95. Jenny Oh, *French Court Gives Yahoo More Time*, THESTANDARD.COM, at <http://www.codoh.com/newsdesk/2000/000724std.html> (last viewed July 24, 2000).

96. *Id.*

97. Mr. Perritt is the Dean of the Chicago-Kent College of Law and an expert in Internet law.

98. Carl S. Kaplan, *French Nazi Memorabilia Case Presents Jurisdiction Dilemma*, N.Y. TIMES, Aug. 11, 2000, available at <http://www.moglen.law.columbia.edu/CPC/archive/hatespeech/11law.html> [hereinafter Kaplan, *French Nazi Memorabilia*].

Dean Perritt expressed his hope that an appellate court in France would overturn the May 22 Order based on a finding that Yahoo had not targeted French citizens. He cited the existence of Yahoo France and its exclusion of Nazi items from its site pages as evidence of the absence of such targeting and characterized Judge Gomez's decision as "an exorbitant exercise of jurisdiction that is inconsistent with emerging best practices."⁹⁹ The "emerging best practices" to which Dean Perritt is alluding are presumably the use of a "plus" element in addition to the effects test, in which a showing of harm in a forum is insufficient to confer jurisdiction; some additional factor must also be present, with evidence of targeting of the jurisdiction being the most favored "plus."

Professor Jack Goldsmith¹⁰⁰ proposed that it was proper for France to exercise jurisdiction over Yahoo because "Yahoo has something on its web site that is being accessed by French citizens that violates French law."¹⁰¹ He noted that the United States could likewise enforce its own laws against content posted in France and concluded, "The harmful effects are running in both directions."¹⁰² Professor Goldsmith felt that an answer to the dilemma may be found in the deployment of filtering technology that, although not necessarily perfect in its ability to identify and block all French surfers, would have a significant enough effect on the ability of those surfers to access illegal content.¹⁰³ Such a solution could potentially resolve many Internet-related jurisdictional disputes.¹⁰⁴

The opinions of Professor Goldsmith and Dean Perritt focus, alternatively, on legal and technological solutions to the jurisdictional problems posed by *LICRA v. Yahoo!*. For reasons discussed below, the best solution to this problem will require a mixture of these elements, also referred to as "modalities," as well as a marketplace that appeals to our cultural values and mores. The effect of these elements/modalities on the regulation of the Internet is discussed in detail by Lawrence Lessig in his seminal work, *Code and Other Laws of Cyberspace*.¹⁰⁵ As Part III argues, it is only through the adoption of a restructured approach to international jurisdiction, and a set of guidelines or international standards

99. *Id.*

100. Professor Goldsmith is an expert in Internet jurisdiction from the University of Chicago.

101. *Id.*

102. *Id.*

103. *Id.*

104. *Id.*

105. LAWRENCE LESSIG, *CODE AND OTHER LAWS OF CYBERSPACE* 88 (1999) [hereinafter LESSIG, CODE].

incorporating aspects of these four elements, that we are likely to reach a workable solution to this jurisdictional dilemma.¹⁰⁶

4. *The French Court's Interim Order of November 20, 2000—The Verdict of the Experts*

On November 6, 2000, the French court conducted the hearing in Paris to receive the report of the expert panel on the technical measures, and to hear one last round of general arguments on jurisdiction and on the technical issues before finalizing its decision. The decision, rendered by the Court on November 20th (“November 20 Order”), although characterized as an Interim Court Order, was a final order on the application for relief by the plaintiffs, and as such was appealable under French law.¹⁰⁷

The November 20 Order appeared to add a targeting test to the effects test the court first used in finding jurisdiction over Yahoo. Judge Gomez began with a reiteration of the jurisdictional rationales supporting his May 22 Order.¹⁰⁸ He then added an additional piece of evidence that Yahoo was targeting French citizens in its advertising: “Whereas YAHOO [sic] is aware that it is addressing French parties because upon making a connection to its auctions site from a terminal located in France it responds by transmitting advertising banners written in the French language”¹⁰⁹

By adding the further ground for the imposition of jurisdiction, Judge Gomez had now employed two different approaches to justify his decision. The effects test previously examined remained the principal

106. The merits of this approach were addressed by Thomas P. Vartanian, a Washington, D.C., attorney who at that time chaired a committee on cyberspace law for the American Bar Association and said that he expected to see similar cases in the coming years “unless the world can agree on what the standards for jurisdiction should be.” Kaplan, *French Nazi Memorabilia*, *supra* note 98. Mr. Vartanian and an international group of lawyers had just completed a two-year study that called for the creation of an international body to develop uniform global principles of Internet jurisdiction. *Id.* The failure to act on this problem, Mr. Vartanian warned, could result in the smothering of the emerging e-commerce golden goose. *Id.*

107. Pl.’s Compl. for Decl. Relief, ex. B at 20, *Yahoo!, Inc. v. La Ligue Contre Le Racisme et L’Antisémitisme*, 169 F. Supp. 2d 1181(N.D. Cal. 2001) (No. 00-21275), available at <http://www.cdt.org/speech/international/001221yahoocomplaint.pdf> (last visited Dec. 2, 2003).

108. *Id.* ex. B, at 3-4. Judge Gomez appeared irritated that counsel for Yahoo raised these jurisdictional arguments at this November hearing, writing, “Whereas in support of its incompetence plea, *reiterated for the third time*, the company Yahoo [sic]” *Id.* ex. B, at 3 (emphasis added).

109. *Id.* ex. B, at 4.

basis for jurisdiction. His second ground added a targeting analysis, which is the approach Professor Geist and other scholars are encouraging the United States to embrace as a more reasonable and less all-encompassing basis for jurisdiction.¹¹⁰

Judge Gomez also rejected Yahoo's argument that an order of the French court directing Yahoo to take actions in the United States would violate the First Amendment and would therefore be unenforceable. In his view, the enforceability of the order was irrelevant to the decision to impose jurisdiction:

Whereas any possible difficulties in executing our decision in the territory of the United States, as argued by YAHOO [sic] Inc., cannot by themselves justify a plea of incompetence;
Whereas this plea will therefore be rejected.¹¹¹

Having dispatched the jurisdiction argument for the third time,¹¹² the court next reviewed the report of the expert panel on the issue of whether a technical solution existed that would allow Yahoo to comply with the May 22 Order.¹¹³ The experts' report begins with a disclaimer:

The undersigned consultants are at pains to point out that their brief is limited to answering the technical questions put by the Court. In no circumstances may their answers be construed as constituting a technical or moral backing of the decisions of the court or, on the contrary, a criticism of these decisions.¹¹⁴

The report went on to detail the various options available for blocking the Yahoo U.S. auction pages from access by French web surfers. They concluded that approximately 70% of French surfers could be identified by their IP addresses,¹¹⁵ and that this group could be effectively blocked

110. Michael Geist, *supra* note 35, at 1384–1404. Professor Geist suggests replacing the effects test with a new standard focusing on three factors: contracts, technology, and actual or implied knowledge. *Id.*

111. Pl.'s Compl. for Decl. Relief, ex. B at 4, *Yahoo! II* (No. 00-21275).

112. *See id.* Counsel for Yahoo! Inc., raised the jurisdiction argument at the May 22 hearing, the July 24 hearing, and again at the November 6th hearing.

113. The court included the report of the experts, entitled "Opinion of the Consultants," as part of the Interim Court Order. *See id.* ex. B, at 4–15.

114. *Id.* ex. B, at 4. As will be illustrated later, the experts were very uncomfortable with the directions they received from the French court, and in at least one case, an expert published an online apology for his contribution to the report. *See infra* note 141 and accompanying text.

115. *Id.* ex. B, at 8.

from accessing those pages via the use of geo-location filtering software, which could be loaded onto Yahoo's servers in California.¹¹⁶

The viability of geo-location filtering has been the subject of some debate.¹¹⁷ Geo-location filtering software works through algorithms that can identify the geographical source of a web surfer by 1) cross-comparing the results of a tracer analysis of the route of the Internet transmission, following the transmission nodes the message went through and their physical locations, and 2) by mapping the Internet Protocol addresses in the surfer's header with IP address databases.¹¹⁸

Neither of these two tracking systems guarantees 100% accuracy in identifying all surfers' geographic origins. Cross-referencing the results of both tracking systems, however, generates results ranging from a low figure of 80% to a high figure of 99% accuracy.¹¹⁹ Several companies are actively marketing geo-location filtering software to allow Internet providers to comply with local laws and engage in target marketing by routing surfers from particular locations to products tailored to their regional needs.¹²⁰ For several reasons, however, geo-location filtering is not always an optimal solution to online content regulation. The software is expensive, and as noted, not 100% accurate. It is, as the experts found, also circumscribable with relative ease.¹²¹ A further objection to this kind of software involves concerns about the preservation of the privacy rights of surfers to be free of software that identifies them as they surf the net, and maintaining the end-to-end Internet without technological roadblocks.

The experts next considered whether the number of surfers who could be blocked would increase if all surfers were asked to verify their national origin before accessing the Yahoo site. They concluded that verification would increase the number of surfers blocked on the basis of national access point by about 20%.¹²² In sum, their opinion was that the use of geo-location filtering software coupled with a requirement for a

116. *Id.* ex. B, at 9-10.

117. For an overview of this debate, see Jack L. Goldsmith & Alan O. Sykes, *The Internet and the Dormant Commerce Clause*, 110 YALE L.J. 785, 809-12 (2001).

118. *Id.* at 810-11.

119. *Id.*

120. Examples of this software include the NetLocator™ software developed by Infosplit, available at http://www.infosplit.com/prod/ip_address_country.htm (last visited Oct. 20, 2003), and the EdgeScape technology produced by Akamai, at <http://www.akamai.com> (Oct. 20, 2003). Infosplit was one of the advisors to the UEJF in the Yahoo! case. See Sayer & Deveaux, *Jurisdiction in Cyberspace*, IDG News Service, July 28, 2000, at <http://www.pcworld.com/news/article/0,aid,17868,00.asp>.

121. Pl.'s Compl. for Decl. Relief, ex. B at 12, *Yahoo! II* (No. 00-21275).

122. Pl.'s Compl. for Decl. Relief, ex. B at 8, *Yahoo! II* (No. 00-21275).

declaration of nationality would produce a filtering success rate approaching 90%.¹²³

The experts' opinions were not unanimous. American based expert Vinton Cerf wrote a "minority report" criticizing the efficacy of the geo-location filtering solution on several grounds. He began by pointing out a common problem with all self-generated online authentication measures: the user can lie about her identity characteristics.¹²⁴ For example, how effective can a website identification page that requires a person to verify that she is over the age of eighteen really be in barring entry by underage surfers? Without any means for verifying the truth of the response, this kind of measure is of limited value.

Cerf went on to note that requiring users to provide national identity information might violate the laws of privacy of other countries, and that enforcing compliance with this term in those other countries might be difficult.¹²⁵ This would seem to be less of an obstacle than Cerf might think, since it would not be difficult for Yahoo to require a statement of national identity as a prerequisite to entry into any Yahoo site other than one based in the user's own country. What would be lost in such a system is the ease with which surfers all over the world can now reach the Yahoo U.S. website. Thus, this objection points to the crux of the distasteful dilemma that enforcing local content laws on the Internet presents to websites that seek to cultivate an international audience: either create sites in each country you wish to reach, tailored to that country's social and cultural values and laws, and protected with technological measures that threaten privacy and the free end-to-end Internet, or create a single "dumbed down" version of the site that attempts to cater to all of the limitations that are or will be imposed by each nation.

Neither of these choices is likely to appeal to web providers. One clear byproduct of the first choice is that less well-funded sites would be shut down because they are unable to create and maintain local versions of their sites. Creating a site that is palatable to all countries and reflects all the limits each nation would propose may prove equally daunting and would doubtless reduce Internet sites to a level of blandness that would eventually sap all interest in the Internet as an effective means of communication between nations. Cerf's criticism again highlights the need for a better alternative than these two choices—each of which is a way of

123. *Id. ex. B*, at 14.

124. *Id. ex. B*, at 15.

125. *Id.*

opening the egg that is likely to lead to many more years of international squabbling and uncertainty.

The court discounted Vinton Cerf's objections and noted that he did not disagree with the findings of the other consultants. Judge Gomez felt that Cerf's concerns were targeted to the degree of success geo-location filtering would achieve, rather than whether there was any merit to attempting it.¹²⁶ The judge went on to take the experts' estimate that 90% of French surfers could be blocked by the combination of geo-location filtering and national identity confirmation as sufficient to accomplish the remedy sought by the plaintiffs.¹²⁷ He rejected Yahoo's arguments that the technical means to achieve these results were not available or would be too burdensome, pointing out that Yahoo already refused to accept offers for the sale of human organs, drugs, works, or objects connected with pedophilia, cigarettes, or live animals on its auction pages, and noting that these limitations did not seem to be deemed a violation of the First Amendment.¹²⁸

In making this argument, Judge Gomez kicked a hole in the fence erected by protectors of the First Amendment. He correctly noted that in the United State, the First Amendment is not an absolute bar to restrictions on speech; instead, it is a conditional bar that legislatures and courts have the power to restrict or expand depending on legal and cultural values. The significance of this argument in the context of online content regulation is that it moves the debate off the moral high ground of absolute protection for all speech to the murkier swamp where the question is who has the power to limit what speech—comparing, for example, the U.S. ban on child pornography with the French ban on Nazi propaganda.

To Judge Gomez, the answer is clear. He concludes this portion of his opinion by exhorting Yahoo to consider the value of banning Nazi-related items: "Whereas it would most certainly cost the company very little to extend its ban to symbols of Nazism, and such an initiative would also have the merit of satisfying an ethical and moral imperative shared by all democratic societies"¹²⁹

Concluding that Yahoo had the ability to install geo-location filtering software and to demand nationality verification from its users, the court ordered Yahoo to put these technical measures in place within 90 days of the date of the order, on penalty of 100,000 Francs (about \$13,000 U.S.

126. Pl.'s Compl. for Decl. Relief, ex. B at 16, *Yahoo! II* (No. 00-21275).

127. *Id.* ex. B, at 14.

128. *Id.* ex. B, at 18.

129. *Id.*

dollars as of the date of the order) per day for each day thereafter that it remained out of compliance with the order.¹³⁰

5. *The International Response to the November Hearing and Order*

Reaction to the French Court's November ruling was swift and vigorous. *The Economist* characterized the ruling as one that "threaten[ed] the freedom of the web," and criticized the filtering order:

But such filtering would not be foolproof. Like most things on the Internet, attempts to control or censor content are soon circumvented. . . . The French ruling could result in many more such restrictions by governments. This is an alarming prospect for many of the Internet's pioneers, or anyone else who opposes government censorship.¹³¹

The New York Times characterized the decision as "a shot heard 'round the world,'"¹³² summarizing the conflict in views arising from the ruling thusly:

Some lawyers say the decision earlier this week, rooted in a French anti-Nazi statute, is an alarming example of a foreign court's willingness to impose its national law on the activities of a United States-based web site.

Even worse, they say, is the ruling's implication. Under the Paris court's logic, any Web site [sic] with global reach could be subject to the jurisdiction of every nation on earth. Forced to comply with a patchwork of local laws, global e-commerce could grind to a halt.¹³³

Yahoo counsel Michael Traynor viewed the decision as an alarming effort of a foreign court to dictate standards for the worldwide Internet: "One country is purporting to exercise and impose its standards on a worldwide conversation. It's fundamentally an interference with freedom

130. *Id.* The Court also found that Yahoo! France had complied with the requirements imposed on it in the May 22 Order, except as to displaying a banner warning French surfers about the legal risks of connecting to the U.S. Yahoo! site, and so the Court reiterated this portion of the May 22 Order and directed immediate compliance with it. *Id.* ex. B, at 19-20.

131. *Yahoo!'s French Connection*, THE ECONOMIST, Nov. 20, 2000, available at http://www.economist.com/agenda/PrinterFriendly.cfm?Story_ID=431328.

132. Carl S. Kaplan, *Ruling on Nazi Memorabilia Sparks Legal Debate*, N.Y. TIMES, Nov. 24, 2000, available at <http://www.nytimes.com/2000/11/24/technology/24CYBERLAW.html> [hereinafter Kaplan, *Ruling on Nazi Memorabilia*].

133. *Id.*

of speech and expression.”¹³⁴ David J. Loundy¹³⁵ also worried about the long-term consequences of the decision:

The next thing you know a court in the Middle East will order another U.S. Internet company to block Middle Eastern consumers from seeing soft-core pornography, which is legal here but illegal there. You can pick your country and pick your problem. Will every Internet company in the future have to put on 42 geographical filters to make everybody happy? Or 420 filters?

There are other questions, Loundy said with a sigh: Does an Internet company have an affirmative duty to figure out the laws of every nation in the world and put on the appropriate geographical filters, or do they just have to put on filters following a court order? And as the technology gets better, does a company have a duty to slap on the newest filtering gizmo?¹³⁶

Others interviewed by *The New York Times*, however, hailed the decision as correctly decided and disagreed with critics who predicted dire consequences for e-commerce:

Rubbish—the sky is not falling, other lawyers say with equal fervor. The Paris court’s decision was perfectly reasonable under the circumstances, they claim. Indeed it is a welcome harbinger of things to come.¹³⁷

Goldsmith said it was “too simplistic” for some lawyers to complain that the ruling threatens global e-commerce.

“IBM, McDonald’s and other international companies sell stuff into every country in the world and they have to comply with local laws,” he said. The fact that real-space companies have to obey a patchwork of laws “hasn’t brought real-space commerce to a halt”

. . . .

“This decision is significant because it shows that geographical filters, though not perfect, are feasible and that nations can take reasonable steps to keep content out I have

134. *Id.*

135. David J. Loundy a Chicago-based specialist on Internet law.

136. Kaplan, *Ruling on Nazi Memorabilia*, *supra* note 132.

137. *Id.*

no doubt that the Internet will become more geographically filtered. This ruling will enhance the trend.”¹³⁸

These sharply conflicting viewpoints illustrate why the ruling in the *LICRA v. Yahoo!* case has attracted so much attention. Professor Goldsmith is right to point out that companies doing business in other jurisdictions have always had to deal with adapting to local cultural values to be successful. But Michael Traynor and David Loundy are also right in noting that in the case of website hosts, the burden of having to try to determine and satisfy all of the different laws of every country may be too onerous for many businesses to handle.

At the core of this problem, as the Introduction noted, is the unique simultaneity and global reach that mark the Internet. Yes, companies doing business abroad have always had to deal with the burden of conforming to local country standards to do business—yet Yahoo tried to do just that by creating local websites such as Yahoo France that strictly complied with the law. The problem was that by nature of the Internet the site they hosted in the United States was viewable with equal ease as the local site. This suggests yet another solution to the problem: block all visitors from any site but local sites. In essence, this would reduce the Internet to the model of television. Each country would have the opportunity to strictly control everything that came into the country online, through a combination of technical measures, including filtering, geographic and age identifiers, regulation of ISP licensing, and other measures.¹³⁹ But following this course of ultimate national sovereignty would end all hope for the Internet to serve as a contemporary commons—a meeting place for the free exchange of ideas.¹⁴⁰ By deciding that each country has the absolute right to dictate which end of the egg must be broken, we would discard a singular opportunity to enhance worldwide communication, literacy, and awareness of our common and different ways of life.

In a dramatic development, critics of the French court’s decision were joined by one of the experts on whose opinion the court relied. On November 21, 2000, European expert Ben Laurie wrote and posted online

138. *Id.*

139. This is the outcome feared by Stanford Law School Professor Lawrence Lessig, which he outlines in detail in his two works on Internet regulation, *Code and Other Laws of Cyberspace* and *The Future of Ideas*. See LESSIG, *CODE*, *supra* note 105; LAWRENCE LESSIG, *THE FUTURE OF IDEAS: THE FATE OF THE COMMONS IN A CONNECTED WORLD* 17–100 (2001) [hereinafter LESSIG, *FUTURE OF IDEAS*].

140. See LESSIG, *FUTURE OF IDEAS*, *supra* note 139, at 17-100.

an extraordinary document, entitled *An Expert's Apology*.¹⁴¹ Critical of the press coverage of the decision, which he characterized as marked by a "remarkable lack of deep thought on this matter," Laurie felt compelled to explain his contribution to the report and his disagreement with its use.¹⁴² Laurie pointed out that the panel of experts answered the first question put to them, whether it was technically possible for Yahoo to comply with the judgment against them, with a "no" answer—full compliance was not possible.¹⁴³ He then noted, "I was not allowed to leave it at that; remember that if it was not possible to comply completely, I was asked to say to what extent compliance is possible."¹⁴⁴ It was in response to this second question that he and the other experts noted that the seventy percent compliance could be achieved by filtering, and an additional amount by asking for geographic identity confirmation. Laurie characterized the percentage figures he and the rest of the panel agreed on as "a rather flaky guess."¹⁴⁵

Laurie asserted that both the filtering technology and the requirement of geographical identification "can be trivially circumvented,"¹⁴⁶ and went on to tell readers exactly how to do it, giving a website URL for anonymizer software that can be loaded onto a user's computer for the purpose of masking their geographic identity, thereby precluding filtering programs from preventing access.¹⁴⁷ As to the confirmation request issue, he pointed out that this can be avoided "simply by lying."¹⁴⁸

Despite his critiques, in answering the question "so what does it all mean?," Laurie agreed that the French court properly asserted jurisdiction and suggested that if Yahoo wanted to be beyond France's reach, it could withdraw its operations in France.¹⁴⁹ He characterized the November 20 Order as ineffective, but noted that this is a natural result of the legal process: "Yes, the solution is half-assed and trivially avoidable. We know that. But it is still the natural outcome of applying the law. Law-abiding citizens are aided in obeying the law, and law-breakers are able to do so, just as they can slash tires, or mug people in the street."¹⁵⁰

141. Ben Laurie, *An Expert's Apology*, Nov. 21, 2000, at <http://www.apache-ssl.org/apology.html>.

142. *Id.*

143. *Id.*

144. *Id.*

145. *Id.*

146. *Id.*

147. *Id.*

148. *Id.*

149. *Id.*

150. *Id.*

Laurie concluded by asserting that the case simply reaffirms that governments will continue to try, and fail, to regulate the Internet.¹⁵¹

No one will argue with Laurie's assertion that governments will continue to try to regulate the Internet. However, there is little evidence to suggest that they will fail completely in those efforts. Both of Professor Lessig's books bear eloquent and often chilling testimony to the many successes achieved by government in regulating aspects of the Internet in the United States and abroad.¹⁵² It would also be an error to conclude that the universal viewpoint in the United States supported the free speech position advocated by Yahoo in the French court.

Another article in *The New York Times* summarized the contrary views on this issue¹⁵³—particularly those of Rabbi Abraham Cooper, an associate Dean at the Wiesenthal Center, which has launched an effort to convince online services to eradicate sites that promote anti-Semitism or white supremacy. The Wiesenthal Center became involved in this effort as it realized that such sites were proliferating, with the number reaching more than 2000 by the spring of 1999.¹⁵⁴

Rabbi Cooper rejected the First Amendment defense and instead urged, as Judge Gomez did, a return to moral values.

“It's good to try to wrap yourself around free speech . . . but in this case it doesn't wash. Television stations, newspapers and magazines refuse to accept some advertisements in an effort to marginalize viewpoints and products that the vast majority of Americans think are disrespectful or even potentially dangerous. Internet companies . . . should just do what American companies have been doing for half a century: reserve the right not to peddle bigotry.”¹⁵⁵

Rabbi Cooper's argument is the same voiced by Judge Gomez: First Amendment protection in the United States is not, and never has been, protection of an absolute standard. Instead, the standard has been subject to exceptions, particularly in the area of commercial speech, which are often driven by our social and cultural values. Cigarette ads were widespread on television during the 1950s and 1960s, but were later banned for health reasons. To the cigarette companies, this ban was an

151. *Id.*

152. *See generally* LESSIG, *CODE*, *supra* note 105; LESSIG, *FUTURE OF IDEAS*, *supra* note 139.

153. Guernsey, *supra* note 16.

154. *Id.*

155. *Id.*

infringement of their right to freely speak out in promotion of their goods, yet the ban withstood challenge and remains in effect.¹⁵⁶

Rabbi Cooper's argument proved effective with Yahoo. A month after meeting with him and other representatives of the Wiesenthal Center, Yahoo voluntarily announced that it would more actively enforce its terms of service agreement by pre-screening and eliminating hateful and racist material, such as Nazi memorabilia and Ku Klux Klan artifacts, from its auction sites. This new enforcement approach did not apply to non-commercial content on Yahoo's sites, appearing to be limited to commercial speech material.¹⁵⁷ Yahoo representatives claimed that the change in policy had nothing to do with Judge Gomez's ruling, but was instead part of a general housekeeping of their auction sites, and in response to their ongoing discussions with Jewish groups in the United States such as the Wiesenthal Center.¹⁵⁸

Judge Gomez, for his part, remained confident that his decision properly followed French law. In an interview,¹⁵⁹ he expanded on his reasoning in applying the "effects test" as follows:

For me, the issue was never whether this was an American site, whether Yahoo had a subsidiary in France, the only issue was whether the image was accessible in France. It is true that the Internet creates virtual images, but to the extent that the images are available in France, a French judge has jurisdiction for harm caused in France or violations of French law.

But in the case of my decision, it was extremely simple: the Nazi collectibles were visible in France, this is a violation of French law, and therefore I had no choice but to decide on the

156. See Mark R. Ludwokowski, *Proposed Government Regulation of Tobacco Advertising Uses Teens To Disguise First Amendment Violations*, 4 COMM. LAW CONSPICUOUS 105, 106 (1996).

157. Carl S. Kaplan, *Experts See Online Speech Case as Bellweather*, N.Y. TIMES, Jan. 5, 2001, available at <http://www.nytimes.com/2001/01/05/technology/05CYBERLAW.html> [hereinafter Kaplan, *Experts See Online Speech*].

158. *Id.*

159. Although much has been written about the decision of the French court in this case, none of the articles published to date solicited direct comment from Judge Gomez about the debate and controversy flowing from his Order. With the assistance of Marie Galanti, a highly capable bilingual (French/English) law student and research assistant, the author requested that Judge Gomez consent to an interview regarding the reasoning behind his Orders. Noting that this was the first time a source from the U.S. had sought to discuss the case with him directly, Judge Gomez agreed to the interview. Ms. Galanti met with the Judge on June 5, 2001, and posed questions written by the author. The result provides us with a rare glimpse into the judicial reasoning that led to the May 22 Order and the November 20 Order in this case.

face of the issue. Whether the site is all in English or not makes no difference. The issue of visibility in a given country is the only relevant issue. Even if the image is virtual, if it is accessible on a computer monitor in France, a French judge has jurisdiction to intervene if the image violates French law. The issue is strictly a geographic one. Even an American in France violates French law by consulting a site prohibited by French law.¹⁶⁰

As to Yahoo's decision to pull the Nazi objects off its auction pages, Judge Gomez reinforced the view expressed in his orders that free speech was not really at issue here. Rather it was simply a business decision for Yahoo to make: Should it add Nazi items to those it already regulated on its site? He noted that American judges, in his view, were also becoming skeptical of First Amendment defenses raised in purely commercial sales contexts:

Yahoo's decision confirms the fact that it had the ability to control its auction site. We knew this because there were precedents showing us that it could exercise this control: for example, Yahoo chose not to sell cigarettes in certain countries where the sale was prohibited. Banner ads were specifically tailored for and directed to audiences in certain countries. Therefore, it was a business decision to put these items on its site and a business decision to remove them. Behind the philosophical discussion, there are basic technical questions and financial interests.

Now Yahoo wants an American judge to say that a foreign judge may not interfere with its freedom of speech. But, even in the U.S., this freedom of speech is not an absolute right. Some of my American colleagues are having second thoughts about using the First Amendment to promote endeavors that are strictly financial, strictly business. For my part, I am thrilled that capitalistic enterprises are getting involved in this debate.¹⁶¹

In its earliest incarnations, the Internet was viewed as a newly-minted anarchic format—a frontier where all forms of communication were acceptable and which was not, and could not, be bound by any rules or regulation. As Ben Laurie said, “the Internet does not adapt well to the control of subject matter—people have been trying to do that [intervene

160. Interview by Marie Galanti with J. Jean-Jacques Gomez, First Deputy C.J. of the Super. Ct. of Paris, in Paris, Fr. at 2-3 (Marie Galanti trans., June 5, 2001) (on file with author) [hereinafter Interview, Gomez].

161. *Id.* at 3-4.

and censor content] since it started, and they've never got anywhere. This case is no exception."¹⁶²

In Lawrence Lessig's *Code and Other Laws of Cyberspace*, he takes the view that whether the Internet will be regulated or free is a choice to be made—and the question is by whom.¹⁶³ In a dark view of that choice, he suggests that while an Internet largely free of regulation is a possibility, the forces of law and code and the inertia of society militate more in the direction of a highly regulated Internet.¹⁶⁴

Indeed, many would agree that the vision that the Internet would be a place for the untrammelled exchange of ideas, in a purely anarchic form, was always just a vision and a dream. If you consider that the initial architecture of the Internet, before even getting to the point of software and applications and operating systems, is the telecommunications network, you come crashing back to reality with the recognition that someone has to pay to create that physical network, and has to continue to pay to maintain it. Initially, the U.S. Government paid the bill through the Department of Defense.¹⁶⁵ As commercial activity erupted online with the development of hypertext mark-up language ("HTML") and the web, the cost was assumed by the commercial sectors of many countries. In some countries, such as China, government still retains control of and pays for the infrastructure of telecommunications.¹⁶⁶ It is a sweet but naïve view to think that these commercial entities and/or governments would willingly pay for a system over which they could exercise no control. The Internet never was, and never will be, "free."

In the debates that followed his orders, Judge Gomez shared the view that a free Internet is not possible, but for a different reason. He explained:

Philosophically, the issue of a totally free Internet is utopia. Clearly if we are dealing with what we call in France the "bon père de famille" or what you call in the United States the "reasonable person" standard, then a free Internet would be possible. If users did not attempt to use the Web for anything illegal, immoral or otherwise reprehensible, we could dispense

162. Laurie, *supra* note 141.

163. LESSIG, *CODE*, *supra* note 105, at 213-30.

164. *Id.*

165. Vinct Cerf, *A Brief History of the Internet and Related Networks*, at <http://www.isoc.org/internet/history/cerf.shtml> (last visited Dec. 7, 2003).

166. *See, e.g.*, Hanna Beech, *Living It Up in the Illicit Internet Underground*, *TIME*, July 22, 2002, at 4 ("Although the Public Security Bureau has deployed a corps of Internet police to block surfers from offending websites, there's no way a few hundred officers can filter the whole Web and maintain blocks that stymie users for long.").

with regulations. But, this is not always the case. Just look at the newspapers. Specialty publications give all kinds of tricks and tips to bypass regulations on the Net.¹⁶⁷

The Internet, Judge Gomez has maintained, is simply a tool, a machine, and as such can no more be left unregulated than we would leave a car to drive itself, or a drill to run without a hand to guide it. When asked whether he considered the Internet and the web to be American creations, Judge Gomez commented:

The origin is of little importance. I look at their creation as I would look upon the discovery of a vaccine for a disease. No matter where it comes from, it brings a contribution to humanity. The origin of the Internet was military and surely the U.S. was the first country to develop it. This does not disturb me. On the contrary, I am pleased that this development has taken place.

The question of its origin is not the most important. The fact that it has become a tool is the issue. A tool cannot be left to its own devices, without guidance or human involvement.

When people say that the Internet cannot or should not be controlled or regulated, I cannot agree. A machine has no sense of responsibility: man has a sense of responsibility. The machine cannot be left to control man. To say that the Internet has not been conceived to be regulated is a monumental mistake. Man's illegal acts cannot hide behind a machine. Whether we are dealing with the activities of pedophiles or with racist acts, we cannot allow men to hide behind the machine. We cannot forget that behind the Internet there are people and there must be some form of regulation. Responsibility cannot hide behind technical difficulties.¹⁶⁸

Judge Gomez reiterates the argument made by Rabbi Cooper, that the ability to spread hate on the web with relative anonymity cannot be justified by reliance on laws that allow that message to be spread under the protection of freedom of expression, or because the architecture of the web makes regulating any kind of speech more difficult. Instead, both men argue that the third of Professor Lessig's modalities, the regulation by cultural and social norms, must enter the picture and become a guiding principal in the quest for standards for the regulation of the Internet.

The courtroom is not the best forum to debate the importance of using cultural norms in the regulation of online content, and as the *LICRA v.*

167. Interview, Gomez, *supra* note 160, at 5.

168. *Id.* at 6-7.

Yahoo! case moved from France to the United States, this discussion took a back seat to issues better suited to the litigation process. The next phase of the case would concentrate on the extent to which the French court's orders were enforceable in the United States.

B. Yahoo v. Licra—The United States Litigation and the Question of Enforceability

As the case crossed the Atlantic, the focus of the litigation moved from the examination of the technical measures and the question of whether the U.S. Yahoo site violated French law, to the issue of the impact an American court's heightened scrutiny, from a First Amendment perspective, would have on the likelihood of a court enforcing Judge Gomez's order. An additional issue posed was whether there was a justiciable controversy as of the time when the litigation shifted to the United States.

When asked what he thought the likelihood of enforcement of his decision might be in the United States, Judge Gomez replied,

[k]eep in mind that my decision was only the first step in the process. . . . Yahoo had many options. It had the option to appeal in France since my courtroom is only the first level. There was also the option to appeal in the US. To be enforceable, my decision required the "exequatur" or execution by an American court. An American court was free to grant this or not. If the court executed my judgment, then Yahoo was free to appeal that decision. Or the French associations are also free to appeal it had it been contrary to them. My decision left a host of legal options available to all parties. . . . My responsibility was not determining how the order might be enforced internationally. In fact, I denied the French associations' request to allow for attorneys' fees to defend in the United States.¹⁶⁹

While Judge Gomez declined to speculate on the likelihood of enforcement, he instead emphasized that Yahoo had many options as well as a fair forum in the French courts.

1. Yahoo Fights Back—The U.S. District Court Action

On December 21, 2000, several weeks before it removed the Nazi memorabilia from its auction pages, Yahoo formally responded to the November Order of Judge Gomez by filing a complaint for Declaratory

169. *Id.* at 5-6.

Relief in the United States District Court for the Northern District of California, San Jose Division, naming LICRA and UEJF as defendants.¹⁷⁰

As a threshold issue, Yahoo dealt with how to assert jurisdiction over the two defendants, who admittedly were citizens of France, with no offices, assets or agents in the United States. Yahoo, following Judge Gomez's analysis in the November hearing, began its jurisdiction claims in the complaint by using a "targeting" approach. Yahoo asserted that the defendants had committed acts targeted against Yahoo by 1) sending a cease and desist letter to Yahoo in Santa Clara demanding removal from the U.S. auction site of items constitutionally protected in the United States; 2) filing a complaint in France relating to the material they viewed on the U.S. site; 3) "repeatedly" using the U.S. Marshal's Office to serve complaints and orders in the Paris lawsuit on Yahoo in Santa Clara; and, 4) seeking an injunction forcing Yahoo to suppress and restrain constitutionally protected speech and to spend significant funds to modify and reengineer Yahoo's U.S. servers to suppress and restrain speech.¹⁷¹

Seeking to bolster its jurisdiction argument, Yahoo next asserted that LICRA, which had established an e-mail account with the U.S. site, had therefore agreed to Yahoo's Terms of Service, which included a jurisdiction clause requiring users to agree to the personal and exclusive jurisdiction of the courts of California.¹⁷² Yahoo similarly asserted that UEJF agreed to its Terms of Service by viewing pages on the U.S. Auction site and directory listings on Yahoo's search engine.¹⁷³

170. Pl.'s Compl. for Decl. Relief, ¶¶ 1-3, *Yahoo!, Inc. v. La Ligue Contre Le Racisme et L'Antisémitisme*, 169 F. Supp. 2d. 1181(N.D. Cal. 2001) (No. 00-21275), available at <http://www.cdt.org/speech/international/001221yahoocomplaint.pdf> (last visited Dec. 2, 2003).

171. *Id.* ¶ 6. This jurisdiction argument adopts the approach recommended by Professor Michael Geist, *supra* note 35, at 1384-1406. This targeting approach provides greater certainty of result than the broader based effects doctrine. Both doctrines, however, carry the risk that a party who initiates an action in their home forum against a content provider whose content violates local laws, must be prepared to find themselves subject to the jurisdiction of the hosting party's home country courts as a result of obtaining a favorable judgment in their own country, and serving either notice, or attempting to enforce that judgment in the hosting party's country.

172. Pl.'s Compl. for Decl. Relief, ¶ 8, *Yahoo! II* (No. 00-21275).

173. *Id.* ¶ 9. This allegation creates an interesting dilemma. Suppose a party in Germany posts a listing on the Yahoo U.S. auction site, and a party in France purchases the goods. If a dispute arises in which one of the parties names Yahoo, the Terms of Service require all parties to come to Santa Clara County, rather than using more convenient forums such as the World Trade Organization or the World Intellectual Property Organization's dispute resolution systems. This functions as a disincentive to sue Yahoo. Additionally, this could provide a defense to the seller of the goods. The

In the balance of its complaint, Yahoo argued that its U.S. site targeted an audience of U.S. users and that the articles on the U.S. site displaying Nazi symbols were constitutionally protected speech in the United States. In addition to arguing that the French court's orders were technically impossible to implement, Yahoo also asserted that the orders constituted an unconstitutional prior restraint on freedom of expression guaranteed by the First Amendment.¹⁷⁴

Finally, Yahoo argued the district court should not recognize the French court orders for several other reasons, including:

U.S. law immunizes Internet Service Providers ("ISPs") such as Yahoo! from responsibility and liability for the content of postings by third parties, as provided by Section 230 of the Communications Decency Act, 47 U.S.C. Section 230(a). If permitted to stand, the French judgment would give foreign nationals a cause of action against U.S.-based ISPs that U.S. citizens do not have.¹⁷⁵

The orders violate sections of three international treaties to which France is a signatory: Article 19 of the International Covenant on Civil and Political Rights; Article 10 of the Convention for the Protection of Human Rights and Fundamental Freedom; and Article 19 of the Universal Declaration of Human Rights.¹⁷⁶

The orders exercise an unreasonable, extraterritorial jurisdiction over the operations and content of a U.S.-based web service belonging to a U.S. citizen. The complaint alleges that the French court instead should have enforced the French Penal

seller could assert that Yahoo was an indispensable party and that the action should not proceed without their participation, forcing the plaintiff to prosecute the action in the U.S. Courts or drop the action.

174. *Id.* ¶¶ 12, 15, 34.

175. *Id.* ¶ 42(a). Section 230 of the Communications Decency Act is often referred to as the "Safe Harbor" clause of the Act. 47 U.S.C. § 230 (2000). This characterization as a safe harbor is not quite accurate. Subpart (c)(1) provides: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." *Id.* § 230(c)(1). Congress intended this subpart to protect the ISP who serves as a mere conduit of information. *See, e.g.,* Stratton Oakmont, Inc. v. Prodigy Servs. Co., No. 31063/94, 1995 WL 323710, at *1 (N.Y. Sup. Ct. May 24, 1995); *see supra* notes 10-12. However, since Yahoo, through its Terms of Service, exercises control over the material posted on its auction sites, arguably Section 230 (c)(1) would fail to provide Yahoo any shelter.

176. Pl.'s Compl. for Decl. Relief ¶ 42(b), *Yahoo! II* (No. 00-21275). The Complaint does not provide details as to the basis for the claim that the Orders violate these sections of the enumerated treaties.

Code directly against French citizens who break French law by accessing the auction sites at issue.¹⁷⁷

The orders are contrary to California's public policy that discourages granting specific or preventive relief to enforce penal law and that discourages implying a private right of action from penal law.¹⁷⁸

France would not give reciprocal recognition to, or enforce a judgment of the State of California in the reverse situation.¹⁷⁹

Based on these arguments, Yahoo alleged that a declaratory judgment was necessary to resolve whether the Paris court's orders were enforceable.¹⁸⁰

LICRA and UEJF retained counsel and filed a Motion to Dismiss the complaint for lack of personal jurisdiction. The parties briefed the issue. On June 7, 2001, United States District Court Judge Jeremy Fogel denied the motion.¹⁸¹ His reasoning remains controversial and on appeal.

2. *The Ruling Denying LICRA's Jurisdictional Motion to Dismiss*

Judge Fogel's ruling denying LICRA's Motion to Dismiss for lack of jurisdiction relied on questionable assumptions that may provide a basis for reversal. The Court of Appeal presently retains the case under submission, a year after the parties presented oral argument in November 2002. Whether the U.S. court had jurisdiction was ultimately irrelevant, however, since Yahoo clearly indicated in the French court that it had no intention of complying with Judge Gomez's order. Just like the residents of Lilliput, Yahoo, convinced of the validity of its position, was prepared

177. *Id.* ¶ 42(c). Here again, the drafters of the Complaint tread on dangerous ground. The U.S. Government has shown little reluctance to assert extraterritorial jurisdiction over websites that host material that violates U.S. law. *See, e.g.*, Joel Michael Schwarz, *The Internet Gambling Fallacy Craps Out*, 14 BERKELEY TECH. L.J. 1021, 1042-49 (1999) (detailing the view that the U.S. Government and State governments may exercise extraterritorial jurisdiction over offshore gambling sites because the resultant gambling activity is in violation of U.S. law). The argument that the French government should instead pursue the Web surfers is equally applicable in the gambling context. Those parties hosting those sites could just as easily argue that the U. S. Government should leave them alone, and limit their policing activities to pursuing the surfers who log on to those sites to gamble in violation of their local or federal law.

178. Pl.'s Compl. for Decl. Relief ¶ 42(d), *Yahoo! II* (No. 00-21275).

179. *Id.* ¶ 42(e). This section of the complaint further alleges, in subparagraphs (f) and (g), that the Paris court lacked personal jurisdiction over Yahoo, the Paris court denied Yahoo a full and fair opportunity to present its case, and the proceedings failed to comport with U.S. notions of due process. Yahoo also presented these arguments in the Paris proceedings, but Judge Gomez rejected them. *Id.* ¶ 42(f), (g).

180. *Id.* ¶ 44.

181. *Yahoo! I*, 145 F. Supp. 2d 1168, 1180 (N.D. Cal. 2001).

to fight on endlessly, rather than seek a less contentious resolution of the dispute.

To what extent countries will attempt to regulate online content depends on whether local courts will extend jurisdiction. In view of the importance of the jurisdictional argument, Judge Fogel's decision warrants careful examination. He first examines whether Yahoo inappropriately attempted to obtain an advisory opinion or whether Yahoo required a ruling to avoid the enforcement of a foreign court's impermissible orders.

a) Case or Controversy

A threshold issue in the dispute over whether the U.S. court had jurisdiction over the French defendants is whether the case presented a "case or controversy" that was ripe for adjudication.¹⁸² Judge Fogel acknowledged that neither LICRA nor UEJF had taken steps to enforce the judgment of the French court.¹⁸³ However, he asserted that this failure to enforce the judgment created a chilling effect on Yahoo's constitutional rights, because the penalties imposed by the French Court order would continue to mount.¹⁸⁴ Judge Fogel also voiced concern that declining to rule on the matter until the French defendants sought to enforce the judgment could set a precedent encouraging parties to create legal uncertainty as a means to coerce website owners to engage in self-censorship. He stated, that

[d]efendants' approach would force the provider to wait indefinitely for a determination of its legal rights, effectively causing many to accept potentially unconstitutional restrictions on their content rather than face prolonged legal uncertainty. California's interest in adjudicating this dispute thus weighs strongly in favor of the exercise of personal jurisdiction.¹⁸⁵

LICRA and UEJF, in their Opening Brief on appeal, argued that what Judge Fogel was asserting in this section of his decision is that the expectation of future harm in an enforcement proceeding gives the court a basis for issuing an advisory opinion, in essence reassuring Yahoo that even if a subsequent order by a French court authorized collection of damages, that order would be unenforceable in the United States.¹⁸⁶ The

182. *Id.* at 1172 n.2 (citing U.S. CONST. art. III). Judge Fogel dealt with this issue initially in a brief footnote, and later in the text of the opinion.

183. *Id.*

184. *Id.*

185. *Id.* at 1179.

186. Opening Brief of Appellants at 31-34, *Yahoo!, Inc. v. La Ligue Contre Le Racisme et L'Antisemitisme*, 169 F. Supp. 2d 1181 (9th Cir. 2002) (No. 01-17424).

appellants cited *Garcia v. Brownell* for the proposition that “the mere possibility, or even probability, that a person may in the future be affected by acts not now threatened does not create an “actual controversy.”¹⁸⁷

Further, the appellants cited authority that holds no case or controversy exists when enforcement depends on potential actions and the evidence fails to show that the other party will take these actions in the near term.¹⁸⁸ Finally, appellants turned to the Supreme Court and cited the decision in *Poe v. Ullman* for the proposition that “the best teaching of this Court’s experience admonishes us not to entertain constitutional questions in advance of the strictest necessity.”¹⁸⁹

In response, Yahoo’s defenders repeatedly argue that declining to rule on the issue until the French groups seek to enforce their order will have a chilling effect on Yahoo’s freedom of speech.¹⁹⁰ Various parties filed amicus briefs in support of Yahoo. They cautioned that unless the Court of Appeals affirms the finding of ripeness, website hosts will be faced with uncertainty which will, in turn, cause them to engage in self-censorship, allowing their attackers to win without even commencing litigation. They view Yahoo’s decision to remove the Nazi memorabilia from its auction sites as proof of this prior restraint.¹⁹¹

The difficulty with this argument is that it asks the court to carve out an exception for web-based businesses that does not exist for comparable brick-and-mortar enterprises. This exception could allow some traditional enterprises to avoid responsibility for compliance with local laws. Advertising created for those markets aims to be sensitive to local laws and cultural values. If we accept the argument that web-based marketing

187. *Id.* at 32 (citing *Garcia v. Brownell*, 236 F.2d 356, 358 (9th Cir. 1956)).

188. *Id.* (citing *Int’l Soc. for Krishna Consciousness of Cal., Inc. v. City of Los Angeles*, 611 F. Supp. 315, 319-20 (C.D. Cal. 1984) (finding no case or controversy in a situation where a resolution barring First Amendment activities at an airport had not yet been enforced, nor could it be without further action by the City Council) and *Japan Gas Lighter Ass’n v. Ronson Corp.*, 257 F. Supp. 219 (D.N.J. 1966) (finding no “Damoclean threat” absent evidence that the defendant was going to “act affirmatively to enforce the protection which he claims”)).

189. *Id.* at 33 (citing *Poe v. Ullman*, 367 U.S. 497, 503 (1961)).

190. *See, e.g.*, Brief of Amici Curiae Center for Democracy and Technology, American Civil Liberties Union et al. at 31-32, *Yahoo!, Inc. v. La Ligue Contre Le Racisme et L’Antisemitisme*, 169 F. Supp. 2d 1181 (9th Cir. 2002) (No. 01-17424); Brief of Amici Curiae Chamber of Commerce of the U.S. et al. at 23-26, *Yahoo!, Inc. v. La Ligue Contre Le Racisme et L’Antisemitisme*, 169 F. Supp. 2d 1181 (9th Cir. 2002) (No. 01-17424).

191. *See, e.g.*, Brief of Amici Curiae Center for Democracy and Technology, American Civil Liberties Union et al. at 31-32, *Yahoo! II* (No. 01-17424); Brief of Amici Curiae Chamber of Commerce of the U.S et al. at 27-28, *Yahoo! II* (No. 01-17424).

originating in the United States need not be sensitive to those laws or values, we provide a means for companies to avoid complying with those restrictions. By withdrawing from actual presence in these foreign countries and maintaining only a virtual presence, U.S. businesses can avoid enforcement of any local judgments rendered against them. A further complication arises from this scenario: web-based businesses in other countries will try to navigate a similar course. Offshore gambling websites have already taken this route,¹⁹² and other businesses are likely to follow.

Yahoo drew support for a finding of ripeness from several Ninth Circuit cases that held that where First Amendment issues are involved, pre-enforcement cases may present a current case or controversy.¹⁹³ In one of these decisions, *LSO, Ltd. v. Stroh*,¹⁹⁴ the Ninth Circuit faced a case similar in many ways to the *Yahoo! v. LICRA* case. On appeal, the court may decide that *LSO*'s holding does not apply to this case.

In *LSO*, appellant Life-styles Organization, Ltd. ("LSO"), sought relief from an effort by the California Department of Alcoholic Beverage Control ("ABC") to prevent it from hosting an annual Erotic Art Exhibition and Trade Show at the Palm Springs Convention Center in Southern California.¹⁹⁵ California's Administrative Code prohibits liquor license owners from displaying graphic sexual images on their licensed premises.¹⁹⁶ ABC took the position that this Code section applied to situations where a liquor license holder, such as the Convention Center, allowed the display of these images, even if no liquor was sold or provided at the showing of the images.¹⁹⁷ ABC advised the Conference Center that if it allowed LSO's show to be presented, it could lose its liquor license.¹⁹⁸ The Center, intimidated by this threat, cancelled the show.¹⁹⁹ LSO filed an action in federal district court seeking declaratory and injunctive relief, both as to the present actions of ABC, and to prevent similar actions in the years to come, due to its plans to return to Palm Springs for subsequent conventions. The district court granted injunctive relief, and LSO

192. See, e.g., Schwarz, *supra* note 177 at 1047-49.

193. See, e.g., Brief of Amici Curiae Center for Democracy and Technology at 31, *Yahoo! II* (No. 01-17424) (citing *LSO, Ltd. v. Stroh*, 205 F.3d 1146, 1155 (9th Cir. 2000), *Bland v. Fessler*, 88 F.3d 729, 736-37 & n.11 (9th Cir. 1996), and *American-Arab Anti-Discrimination Comm. v. Reno*, 70 F.3d 1045, 1057-58 (9th Cir. 1995)).

194. 205 F.3d 1146 (9th Cir. 2000).

195. *Id.* at 1150.

196. *Id.* at 1151 (citing CAL. CODE REGS. tit. 4, § 143.4 (2002)).

197. *Id.*

198. *Id.*

199. *Id.* at 1152.

presented the art show. LSO continued on with its action as to future shows, giving rise to the issue of whether there was a present case or controversy.²⁰⁰

The Ninth Circuit began its analysis by referring to two U.S. Supreme Court decisions and a Ninth Circuit decision that required a party seeking to challenge a pre-enforcement decision to demonstrate “a realistic danger of sustaining a direct injury as a result of the statute’s operation or enforcement.”²⁰¹ An additional significant factor was whether there were past instances of enforcement.²⁰² Finally, the court noted,

[W]hen the threatened enforcement effort implicates First Amendment rights, the inquiry tilts dramatically towards a finding of standing. . . . Accordingly, we have noted that the tendency to find standing absent actual, impending enforcement against the plaintiff is stronger “in First Amendment cases, for free expression—of transcendent value to all society, and not merely to those exercising their rights—might be the loser.” *Bland*, 88 F. 3d at 736-37 (quoting *Dombrowski v. Pfister*, 380 U.S. 479, 486 (1965)). Accord *Navegar, Inc. v. United States*, 103 F. 3d 994, 999 (D.C. Cir. 1997) (“Federal courts most frequently find preenforcement [sic] challenges justiciable when the challenged statutes allegedly ‘chill’ conduct protected by the First Amendment.”).²⁰³

The weight of this authority suggests that whether Yahoo was likely to face enforcement may prove irrelevant to the outcome of the pending appeal. The mere fact that its First Amendment rights might end up being the subject of an attempt at enforcement seems enough under this line of authority to give the Court of Appeals a basis for affirming Judge Fogel’s determination of standing.

However, in this instance, Yahoo was unlikely to face enforcement. By voluntarily removing the auction postings shortly after receiving the orders of the French court, Yahoo dramatically reduced the likelihood of enforcement. Further, no evidence existed that indicated that the French parties were going to seek enforcement of the orders.

Despite all of these facts that distinguish this case from *LSO*, possibly the strong support for finding a case or controversy in matters where First

200. *Id.*

201. *Id.* at 1154 (citing *Babbitt v. United Farm Workers Nat’l Union*, 442 U.S. 289, 298 (1979), *Blanchette v. Conn. Gen. Ins. Corps.*, 419 U.S. 102, 143 n.29 (1974), and *Bland v. Fessler*, 88 F.3d 729, 736-37 (9th Cir. 1996)).

202. *Id.* at 1155.

203. *Id.* at 1155-56.

Amendment claims are involved may lead the Court of Appeals to affirm the finding of ripeness.

b) Jurisdiction

A finding of a ripe case or controversy shifts the jurisdictional analysis to the issue of whether LICRA and UEJF, by their conduct in the United States, meet the other requirements for imposing jurisdiction over them. Given the undisputed fact that neither LICRA nor UEJF regularly conducted business in the United States, the court has no basis for asserting general jurisdiction over them. The question remains whether the Court of Appeals will affirm Judge Fogel's determination that the French defendants' actions meet the criteria for application of specific jurisdiction.

Judge Fogel first noted that the Ninth Circuit applies a three-part test to determine whether a court has grounds to exercise specific jurisdiction: 1) the non-resident defendant must do some act or consummate some transaction within the forum by which the defendant purposely avails itself of the privilege of conducting activities in the forum, thus invoking the benefits and protection of its laws; 2) the claim must be one which arises out of or results from the defendant's forum related activities; and, 3) the exercise of the court's jurisdiction must be reasonable.²⁰⁴

The first element of this test, referred to as the "purposeful availment" requirement,²⁰⁵ is in many ways the most important. Judge Fogel began by noting that the purposeful availment requirement gives "notice to a nonresident that it is subject to suit in the forum state, thereby protecting it from being haled into local courts solely as the result of 'random, fortuitous or attenuated' contacts over which it had no control."²⁰⁶ Noting that Yahoo properly argued that purposeful availment under the "effects test" was the applicable standard,²⁰⁷ Judge Fogel concluded that Yahoo "has made a sufficient prima facie showing of purposeful availment under the effects test."²⁰⁸ Three examples of conduct by the French defendants supported this decision: 1) the "cease and desist" letter sent by the defendants to Yahoo, in Santa Clara, California; 2) the defendants' request

204. *Yahoo! I*, 145 F. Supp. 2d 1168, 1173 (N.D. Cal. 2001) (citing *Bancroft & Masters, Inc. v. Augusta Nat'l, Inc.*, 223 F.3d 1082 (9th Cir. 2000); *Panavision Int'l. L.P. v. Toepfen*, 141 F.3d 1316, 1320 (9th Cir. 1998); *Cybersell v. Cybersell*, 130 F.3d 414, 416 (9th Cir. 1997)).

205. *Id.* (citing *Burger King v. Rudzewicz*, 471 U.S. 462, 476 (1985); *World-Wide Volkswagen Corp. v. Woodson*, 444 U.S. 286, 297 (1980)).

206. *Id.*

207. *Calder v. Jones*, 465 U.S. 783, 789 (1984).

208. *Yahoo! I*, 145 F. Supp. 2d at 1174.

that the French Court order Yahoo to place the geo-location filtering software on its Santa Clara based servers; and 3) the defendants' "utilization of United States Marshals to effect service of process on Yahoo in California."²⁰⁹

A significant problem confronting the District Court in its decision to apply the "effects test" to the jurisdictional issue in this case is that previously the Ninth Circuit had applied the test only in cases arising out of a tort or alleged tortious conduct.²¹⁰ Judge Fogel acknowledged this history. He then justified his decision to extend the test to this case, despite the absence of tortious conduct by the French defendants, by noting that in several of those prior cases, the Ninth Circuit had focused "less on the characterization of the plaintiff's cause of action than on whether the defendant's forum-related acts evidenced intentional, or at the very least, knowing, targeting of a forum resident(s)."²¹¹

Judge Fogel asserted that proper application of the effects test requires consideration both of the nature of the defendant's conduct, as well as an evaluation of whether the evidence reflects that the defendant intentionally or knowingly targeted the forum's resident.²¹² Bypassing the issue of whether the French defendants' conduct was tortious, Judge Fogel found that the defendants targeted California residents, and on that basis he found that they had met purposeful availment criteria.²¹³

Judge Fogel's emphasis on the "targeting" element may be misplaced here. As counsel for LICRA and UEJF argued in their appellate brief, the Ninth Circuit and other jurisdictions have required targeting as an

209. *Id.* The Court fails to cite any authority to support its determination that these three actions are sufficient to establish that the French defendants purposefully availed themselves to the benefits of California law. The defendants argued in their opening brief on appeal that this determination is contrary to case law establishing that a defendant should not be compelled to risk having to submit to jurisdiction in a distant forum to exercise their rights in their local jurisdiction. *Id.* (citing *Bancroft & Masters, Inc. v. Augusta Nat'l, Inc.*, 232 F.3d 1082, 1089 (9th Cir. 2000); *Douglas Furniture Co., Inc. v. Wood Dimensions, Inc.*, 963 F. Supp. 899, 902 & n.1 (C.D. Cal. 1997)).

210. *Yahoo! I*, 145 F. Supp. 2d at 1175 (citing *Meyers v. Bennett Law Offices*, 238 F.3d 1068, 1074 (9th Cir. 2001); *Bancroft & Masters, Inc. v. Augusta Nat'l, Inc.*, 232 F.3d 1082, 1089 (9th Cir. 2000); *Panavision Int'l, L.P. v. Toepfen*, 141 F.3d 1316, 1321 (9th Cir. 1998); *Cybersell, Inc., v. Cybersell, Inc.*, 130 F.3d 414, 429 (9th Cir. 1997); *Ziegler v. Indian River County*, 64 F.3d 470, 473-74 (9th Cir. 1995); *Caruth v. Int'l Psychoanalytical Ass'n*, 59 F.3d 126, 128 n.1 (9th Cir. 1995)).

211. *Id.* (citing *Bancroft & Masters, Inc. v. Augusta Nat'l, Inc.*, 232 F.3d 1082, 1087 (9th Cir. 2000); *Asahi Metal Indus. Co. v. Super. Ct. of Cal.*, 480 U.S. 102, 109-13 (1987); *Burger King v. Rudzewicz*, 471 U.S. 462, 472-77 (1985)).

212. *Id.*

213. *Id.* at 1175-76.

additional element needed to warrant the imposition of the effects test, never as an element that can *substitute* for the requisite tortious conduct.²¹⁴ The district court's decision, now on appeal, presents the Ninth Circuit an opportunity to clarify whether the effects test is limited to tortious or otherwise wrongful conduct cases and whether, in an international context, the definition of wrongful conduct will encompass conduct lawful in another country, but which violates First Amendment rights in the United States.

Having determined based on his targeting analysis that the French defendants purposely availed themselves of the benefits of California law, Judge Fogel had little difficulty finding that Yahoo's claims arose out of the forum-related conduct of the French defendants.²¹⁵ Applying a "but for" test,²¹⁶ the court found that but for the French defendants having filed and prosecuted a lawsuit in France and having used the U.S. Marshal's office to serve process, Yahoo would have had no need to file an action for declaratory relief seeking an order that the French decision is unenforceable in the United States.²¹⁷

Judge Fogel's determination that the French defendants purposely availed themselves of the benefits of California law became a significant factor in his analysis of the third requirement for specific jurisdiction—that the exercise of jurisdiction be reasonable.²¹⁸ Citing the decision in *Bancroft & Masters*, Judge Fogel noted, "When purposeful availment has been established, Defendants have a burden of demonstrating a 'compelling case' of unreasonableness."²¹⁹ This standard sets a very high bar for a defendant. *Bancroft & Masters* enumerated seven factors relevant in determining reasonableness, all of which Judge Fogel balances in his decision.²²⁰ The seven factors are:

214. Opening Brief of Appellants at 14, *Yahoo!, Inc. v. La Ligue Contre Le Racisme et L'Antisemitisme*, 169 F. Supp. 2d 1181 (9th Cir. 2002) (No. 01-17424) (citing *Bancroft & Masters, Inc. v. Augusta Nat'l, Inc.*, 232 F.3d 1082, 1087 (9th Cir. 2000), for the holding that the purposeful availment test was satisfied "when the defendant is alleged to have engaged in *wrongful conduct* targeted at a plaintiff whom the defendant knows to be a resident of the forum state").

215. *Yahoo! I*, 145 F. Supp. 2d at 1175.

216. *Id.* at 1176 (citing *Ballard v. Savage*, 65 F.3d 1495, 1500 (9th Cir. 1995)).

217. *Id.*

218. *Id.* at 1177 (citing from *Burger King*, 471 U.S. at 476 (holding that for the exercise of jurisdiction to be reasonable, it must comport with notions of fair play and substantial justice)).

219. *Id.* (citing *Bancroft & Masters*, 223 F.3d at 1088).

220. *Id.* at 1177-80.

(1) the extent of the defendant's purposeful interjection into the forum state; (2) the burden on the defendant in defending in the forum; (3) the extent of the conflict with the sovereignty of the defendant's state; (4) the forum state's interest in adjudicating the dispute; (5) the most efficient judicial resolution of the controversy; (6) the importance of the forum to the plaintiff's interest in convenient and effective relief; and (7) the existence of an alternative forum.²²¹

In coming to a decision, Judge Fogel applied these seven factors in numerical order. Starting with the purposeful interjection factor, he acknowledged that the court must consider the degree of interjection to determine if jurisdiction is reasonable, even after finding purposeful availment.²²² He concluded that the actions of the French defendants in accessing Yahoo's U.S.-based website, mailing a letter to Yahoo in Santa Clara, using U.S. Marshals to serve Yahoo in Santa Clara, and asking a French court for an order requiring Yahoo to "reconfigure its U.S.-based servers, specifically including servers located in California" were sufficient in degree to support the Court's exercise of personal jurisdiction.²²³ As was the case in the analysis of purposeful availment in the order, the court again declined to cite any case or statutory authority that directly supported the view that these actions by the French defendants were sufficient to support a finding of purposeful availment and the imposition of personal jurisdiction.²²⁴

The court next turned to the second factor, the defendants' burden in litigating a case in a distant forum. Acknowledging that the burden for two French non-profit groups to litigate a case in California is "not trivial," Judge Fogel asserted, however, that the availability of fax machines, telephones, and e-mail greatly reduced this burden.²²⁵ As to the obvious expense for these non-profits needing to hire a U.S. law firm to defend themselves, the court concluded that this was not so severe a burden that it would deprive the French defendants of due process.²²⁶

221. *Id.* at 1177 (citing *Bancroft & Masters*, 223 F.3d at 1088).

222. *Id.* (citing *Panavision Int'l, L.P. v. Toeppen*, 141 F.3d 1316, 1323 (9th Cir. 1998)).

223. *Id.* The last item cited in this litany, that the French defendants sought an order requiring Yahoo to "reconfigure its U.S.-based servers, specifically including servers located in California," appears to be factually inaccurate. The French court's order requiring the reconfiguration of Yahoo's servers to allow for geo-location filtering came from the court in Paris, not from a request by LICRA or UEJF.

224. *Id.*

225. *Id.*

226. *Id.* at 1178.

The court's analysis leads to the question of whether there are, in the modern world, any circumstances under which a U.S. court will find that the burden for a foreign defendant of litigating a case in the United States is unreasonable. The only possible circumstance to support such a finding, it appears, requires a person to be living in a remote part of the world, without telephones, faxes, e-mail, or easy travel means. Given this interpretation of the reasonableness standard, U.S. citizens should not be surprised to find that the courts of other countries will not accept a similar "burden" argument advanced by them, and they should be prepared to shoulder the cost of defense of any claims brought against them in a foreign country.²²⁷

The third factor, that a plaintiff seeking to hale a foreign defendant into a U.S. court must meet a "higher jurisdictional threshold" than is required when the defendant is a U.S. citizen, based on respect for the sovereignty of the foreign country, is the next element the court considered.²²⁸ While acknowledging the great respect and deference due to France's sovereign interest in enforcing the orders of its courts, Judge Fogel held that courts must weigh those interests against the sovereign interests of the United States in protecting the statutory and constitutional rights of its citizens.²²⁹ In his view, French sovereignty must give way to the U.S. interest. He reaffirmed the view that in a conflict of laws situation where constitutional rights are involved, domestic law will be enforced.²³⁰

The court again addressed the issue of ripeness in analyzing the fourth factor, California's interest in adjudicating the dispute. Judge Fogel asserted that the state interest is particularly strong in this case because of

227. See, e.g., *Dow Jones & Co. v. Gutnick* (2002) HCA 56 (Austl.), available at http://www.austlii.edu.au/au/cases/cth/high_ct/2002/56.html (last visited Dec. 2, 2003). In this case, *Barrons Magazine*, a Dow Jones & Co. publication, allegedly defamed Australian businessman Joseph Gutnick through statements included in an online version of the magazine. Gutnick brought suit in the Supreme Court of Victoria, and Dow Jones objected to the jurisdiction of the court. The High Court, applying its own version of the effects test in a tort case, disagreed and held that Dow Jones must respond to the action in the Australian court system. *Id.* ¶¶ 9, 18-24.

228. *Yahoo! I*, 145 F. Supp. 2d at 1178.

229. *Id.* (citing *Bachchan v. India Abroad Publ'ns Inc.*, 585 N.Y.S.2d 661, 664 (1992) (holding a libel judgment of an English court was held unenforceable because it violated First Amendment protections afforded to the press); *Matusevitch v. Telnikoff*, 877 F. Supp. 1, 4 (D.D.C. 1995) (finding that an English libel judgment was not enforceable in the U.S. because it was "contrary to U.S. libel standards"); CAL. CIV. PROC. CODE § 1713.4(b)(3) (Deering 2002) (providing that a court need not recognize a foreign money judgment if the cause of action is repugnant to the public policy of the state)).

230. *Id.*

Yahoo's claim that "its fundamental right to free expression has been and will be affected by Defendants' forum-related activities."²³¹ Judge Fogel acknowledged that the French defendants argued that Yahoo had suffered no actual injury, that the French defendants have not sought to enforce the French Court's judgment, and that they may never seek to do so.²³² Nonetheless, he rejected the defendants' "proposed 'wait and see' approach," claiming that the lack of enforcement could have adverse consequences: "Defendants' approach would force the provider to wait indefinitely for a determination of its legal rights, effectively causing many to accept potentially unconstitutional restrictions on their content rather than face prolonged legal uncertainty."²³³

The court's concern may be valid in the abstract, but is it valid given the facts in this case? The French defendants, more than two years after receiving their judgment in France, have never sought to enforce it in the United States. Yahoo made it clear that it would not abide by the French court's order and was deservedly confident that no U.S. court would enforce it. There is, in fact, no demonstrable evidence that Yahoo's business was affected in any way by the November 2000 Order of the French court.

In sum, where online content may be unlawful in another country but lawful protected speech in the United States, there is no legal uncertainty involved. Yahoo cannot have ever seriously doubted that allowing the posting of Nazi memorabilia for auction and of related Nazi texts was protected free speech in the United States. At the time it filed its action in U.S. District Court, it also had clear notice that these same materials were illegal in France. Where was the uncertainty? What there is instead is a stalemate, with two diametrically opposed viewpoints being expressed by two different courts in two different countries. Whether this kind of stalemate has any real impact on the parties this Article will address in Part IV.²³⁴

Judge Fogel made short shrift of the remaining four factors relevant to determining reasonableness. He found that the fifth factor, the forum best suited for efficient resolution, no longer weighed heavily because of the

231. *Id.* at 1178-79 (citing CAL. CIV. PROC. CODE § 425.16, which provides a procedural mechanism to dismiss at an early stage cases that "chill the valid exercise of the constitutional right of freedom of speech"). Since no actual enforcement of the French judgment has been sought, presumably the prospect that it might be is the conduct which the Court feels chills Yahoo's freedom of speech.

232. *Id.* at 1179.

233. *Id.*

234. *See infra* Part IV.

ease of modern transportation, and that in this case it is essentially a neutral factor.²³⁵ He next found that a U.S. District Court is the more efficient and effective forum in which to resolve the issue of whether the French court's order is enforceable. Having made this finding, Judge Fogel made moot the possibility of the French court as an alternative forum.²³⁶

This analysis led Judge Fogel to conclude that the balance of factors weighs in favor of the California court's exercise of personal jurisdiction, and that the French defendants had not met their burden to make a "compelling case" necessary to rebut the presumption that jurisdiction was reasonable.²³⁷ This decision set the stage for the next procedural step in the U.S. case: a ruling on Yahoo's motion for summary judgment on First Amendment grounds.

3. *The Ruling Granting Yahoo's Motion For Summary Judgment*

Following the court's ruling denying the French defendants' Motion to Dismiss,²³⁸ Yahoo filed a motion for summary judgment on the ground that the French court's November 2000 Order violated Yahoo's First Amendment free speech rights and was therefore unenforceable in the United States. Following the submission of briefs and a September 24, 2001 hearing for oral argument, Judge Fogel granted Yahoo's motion by an order dated November 7, 2001.²³⁹ Although not as controversial or hotly contested as was the motion to dismiss, this order contains an analysis worth examining on the issue of how U.S. courts should deal with the problem of content restrictions for protected speech that originates in the United State, but is accessible online by citizens of other countries, where it is in violation of those countries' laws.

Judge Fogel began with an "Overview" section, in which he acknowledged the sovereign right of a country "to determine by law what forms of speech and conduct are acceptable within its borders."²⁴⁰ He emphasized that we should respect the French Republic's motivation in designating speech promoting the symbols or propaganda of Nazism.

235. *Yahoo! I*, 145 F. Supp. 2d at 1179 (citing *Panavision Int'l, L.P. v. Toeppen*, 141 F.3d 1316, 1323 (9th Cir. 1998)).

236. *Id.*

237. *Id.* at 1180.

238. *Id.* at 1168.

239. *See Yahoo! II*, 169 F. Supp. 2d 1181 (N.D. Cal. 2001).

240. *Id.* at 1186.

Further he asserted that “vigilance is the key to preventing atrocities such as the Holocaust from occurring again.”²⁴¹

However, he noted that the right of a sovereign state to determine the scope of allowable speech within its borders is not what is at issue in the case. He asserted that the issue is “[w]hether it is consistent with the Constitution and laws of the United States for another nation to regulate speech by a United States resident within the United States on the basis that such speech can be accessed by Internet users in that nation.”²⁴²

Judge Fogel acknowledged that since the Internet “renders the physical distance between speaker and audience virtually meaningless,” the significance of the issue as he has framed it goes “far beyond the facts of this case.”²⁴³ Posing the hypothetical situation of another government or party seeking enforcement of its laws against Yahoo or another U.S. based Internet service provider, he asked what principles should guide his, or another U.S. court’s analysis.²⁴⁴ His answer was the Constitution and the laws of the United States.²⁴⁵ His guide was the First Amendment case law that deems it preferable to permit non-violent speech, however repugnant, rather than to impose a “viewpoint” based regulation of speech.²⁴⁶

Judge Fogel concluded this Overview section by foreshadowing his ruling granting Yahoo’s motion with what comes close to an apology to the people and judiciary of France: “The government and people of France have made a different judgment based upon their own experience. In undertaking its inquiry as to the proper application of the laws of the

241. *Id.*

242. *Id.*

243. *Id.*

244. *Id.* at 1187. It is noteworthy that the court did not ponder the converse situation: what principles of law will U.S. courts apply when a foreign party or government which operates as an Internet service provider violates U.S. laws? Examples include facilitating gambling, selling alcohol or cigarettes to minors, and providing material legally deemed obscene in the United States. These activities are lawful in some countries but unlawful here. As noted previously using the gambling example, the Courts will apply the principles of U.S. law, and insist that those principles be applied extraterritorially. *See, e.g., Schwarz, supra* note 177.

245. *Yahoo! II*, 169 F. Supp. 2d at 1187.

246. *Id.* The history of this doctrine is controversial. *See, for example, Justice Brennan’s dissent in City of Renton v. Playtime Theatres, Inc.*, 475 U.S. 41, 50 (1986), in which he argued that the “enterprise zone” concept was viewpoint oriented and that the majority of the Court was engaging in a fiction to claim it was not.

United States, the Court intends no disrespect for that judgment or for the experience that has informed it.”²⁴⁷

The court next turned to its legal analysis of the merits of Yahoo’s motion. The legal issues, in the court’s view, were whether Yahoo had shown the presence of an actual controversy, justifying the need for declaratory relief,²⁴⁸ and whether Rule 56(f) of the Federal Rules of Civil Procedure, which allows postponement or denial of a motion for summary judgment due to a need to conduct further discovery, applied in this case.²⁴⁹

247. *Yahoo! II*, 169 F. Supp. 2d at 1178. This statement also foreshadows the court’s analysis of the difficult issue of comity as applied in this case, which is examined *infra*.

248. *Id.* at 1187-93; *see, e.g.*, 28 U.S.C. § 2201(a) (2000) (providing a right to a declaratory judgment from any court of the United States in any case “of actual controversy within its jurisdiction”). The U.S. Supreme Court has visited this issue on a number of occasions. In *Golden v. Zwickler*, 394 U.S. 103, 108-10 (1969), the court found no actual controversy and reversed a district court decision in a case dealing with an alleged violation of a state statute making it a crime to distribute anonymous literature in connection with an election campaign. Since the politician the literature targeted had, by the time of the hearing on the case, chosen not to stand for re-election, the court found that as of that time there was no actual threat requiring immediate action. The court first noted, “[T]he proper inquiry was whether a ‘controversy’ requisite to relief under the Declaratory Judgment Act existed at the time of the hearing on the remand.” *Zwickler*, 394 U.S. at 108. The defendant’s argument that the former Congressman could be a candidate again was rejected by the court as “hardly a substitute for evidence that this is a prospect of ‘immediacy and reality.’” *Id.* at 109. Citing its prior decision in *United Pub. Workers of Am. (C.I.O.) v. Mitchell*, 330 U.S. 75 (1947), the court repeated its statement of the rule announced in that case: “The power of courts, and ultimately of this Court, to pass upon the constitutionality of acts of Congress arises only when the interests of litigants require the use of this judicial authority for their protection against actual interference. A hypothetical threat is not enough.” *Id.* at 89-90. Applying these cases to *Yahoo v. LICRA* suggests that the question for the Court in ruling on Yahoo’s summary judgment motion is whether an actual threat existed at the time of the hearing on the motion.

249. *Yahoo! II*, 169 F. Supp. 2d at 1193-94. The French defendants argued that they needed additional discovery to develop their theory that this case was bound by dictum contained in *Desai v. Hersh*, 719 F. Supp. 670, 676 (N.D. Ill. 1989), *aff’d*, 954 F.2d 1408 (7th Cir. 1992), which stated “for purposes of suits brought in United States courts, First Amendment protections do not apply to all extraterritorial publications by persons under the protections of the Constitution.” The French defendants claimed that since Yahoo expressly targeted France, this statement suggested that further discovery might reveal a triable issue of fact. *Yahoo! II*, 169 F. Supp. 2d at 1193. That issue was the extent to which Yahoo’s modifications to its auction sites by removing all Nazi memorabilia offerings reduced its potential liability, which presumably also would affect the need for immediate action via declaratory relief. Judge Fogel rejected this argument. He found *Desai* distinguishable, because in that case the acts at issue took place in a foreign country, whereas in this case, he found, the acts at issue, the requirement that Yahoo

Judge Fogel began the “actual controversy” analysis by rejecting the French defendants’ claim that the May and November Orders do not give rise to an actual controversy. The French defendants had reasoned that because Yahoo’s subsequent actions restricting Nazi memorabilia from its auction sites may result in a French court determining that those Orders have now been substantially complied with and the need for several more procedural steps to finalize the judgment in France preclude any immediate action in the United States to enforce the orders.²⁵⁰ In short, the French defendants argued that procedurally, the case is not in an “actual controversy” position.

In rejecting this analysis, Judge Fogel noted that the French defendants took no steps to have the orders withdrawn and that in any subsequent proceeding, any penalties assessed would be retroactive for the entire period of Yahoo’s noncompliance.²⁵¹ He further noted that the substantial compliance argument lacks support, confirming his view that the French court would make a similar finding. He emphasized that Congress intended the Declaratory Judgment Act to relieve this kind of uncertainty.²⁵²

Turning to the question of whether the French orders posed any real or immediate threat to Yahoo, Judge Fogel concluded that they did.²⁵³

modify its servers, would have taken place in the United States. *Yahoo! II*, 169 F. Supp. 2d at 1193.

250. *Id.* at 1188. The French defendants argued that to finalize the judgment embodied in the May and November Orders, they would have to proceed with further evidentiary hearings to set the amount of damages. *Id.* at 1190. They also noted that the damages might not be substantial, since Yahoo removed the Nazi memorabilia within the ninety days given in Judge Gomez’s order—a fact which might cause the French court to conclude that Yahoo was in substantial compliance with its order, resulting in a refusal to order any further damages. *Id.* 1190-91.

251. *Id.*

252. *Id.* at 1189 (citing *Japan Gas Lighter Ass’n v. Ronson Corp.*, 257 F. Supp. 219, 237 (D.N.J. 1966)). The court stated:

The Declaratory Judgment Act was designed to relieve potential defendants from the Damoclean threat of impending litigation which a harassing adversary might brandish, while initiating suit at his leisure—or never. The Act permits parties so situated to forestall the accrual of potential damages by suing for a declaratory judgment, once the adverse positions have crystallized and the conflict of interests is real and immediate.

Id.

253. Judge Fogel distinguished several cases that French defendants asserted should lead the court to the opposite conclusion. *Id.* at 1189-91. The French defendants argued that the decision in *International Society for Krishna Consciousness of California, Inc. v. City of Los Angeles*, 611 F. Supp. 315, 319-20 (C.D. Cal. 1984), supported a finding that

Because the French defendants retained the power to finalize their judgment in France and to thereafter enforce that judgment including retroactive penalties in a U.S. court, Judge Fogel decided that the conflict created sufficient uncertainty to constitute a real or immediate threat. The court reasoned that Yahoo might feel compelled to take action in response to the orders and in derogation of those rights, if it was not able to obtain a declaratory judgment. Quoting from the Supreme Court's decision in *Elrod v. Burns*, Judge Fogel asserted that such a limitation creates a serious injury: "The loss of First Amendment freedoms, for even minimal periods of time, unquestionably constitutes irreparable injury."²⁵⁴

no actual controversy exists in this case. *Id.* In *International Society*, the district court rejected a claim by the City of Los Angeles that it was entitled to declaratory judgment because the resolution at issue, adopted by the Los Angeles Board of Airport Examiners limiting airport speech activities, had not yet been ratified by the City Council and was therefore unenforceable. *Id.* Judge Fogel distinguished this case on the grounds that if the Council chose not to ratify the Board's resolution, the resolution would have no effect at all, whereas in *Yahoo!*, there was no dispute that Judge Gomez's Orders were valid and enforceable, and would be enforced retroactively. *Yahoo! II*, 169 F. Supp. 2d at 1190.

Similarly, Judge Fogel rejected the French defendants' argument that there was no real or immediate threat to Yahoo because they do not presently intend to seek enforcement of the May and November Orders in the United States. *Id.* The French defendants cited the decision in *Salvation Army v. Department of Community Affairs*, 919 F.2d 183 (3d Cir. 1990), in support of this argument. In *Salvation Army*, the well known non-profit group objected to state laws regulating boarding houses as violating its right of freedom of religion. 919 F.2d at 185. In an out-of-court partial settlement, the state authorities agreed to exempt the group from part of the relevant provisions. The district court then granted summary judgment and dismissed the action. *Id.* On appeal, the Salvation Army claimed that it still faced uncertainty because the exemptions granted were not permanent and legally binding, and that the regulations at issue still violated their First Amendment rights. *Id.* The Court of Appeals for the Third Circuit rejected those arguments, agreeing with the district court that since the state had given express assurances that it would not enforce any of the exempted terms of the law, and no criminal penalties could be imposed without additional steps, and no fines could be imposed without notice, the Salvation Army's First Amendment rights would not be actually affected by the threat of future law suits. *Id.* at 192-93.

Judge Fogel distinguished this case by noting that the penalties in the *Salvation Army* case were prospective only, whereas in *Yahoo!* they were also retroactive, and by noting that the exemptions at issue in *Salvation Army* preserved the status quo, while the relevant Orders in *Yahoo!* altered it. Finally, he noted that the provisions in the Orders in *Yahoo!* have never been waived, suspended or stayed and were therefore still in full force and effect, whereas the fear in *Salvation Army* was that exemptions granted might not be still available in the future. For the court, the possibility that the French defendants might seek to enforce the Orders at some time in the future, with retroactive effect, warrants a finding of real and immediate harm. *Yahoo! II*, 169 F. Supp. 2d at 1191.

254. *Yahoo! II*, 169 F. Supp. 2d at 1189 (citing *Elrod v. Burns*, 427 U.S. 347, 373 (1976)).

The court acknowledged that it must weigh the Damoclean possibility hanging over Yahoo's head of retroactive enforcement of the French court's orders against the reality that those orders were incapable of enforcement in any U.S. court. Judge Fogel wrote,

The French order prohibits the sale or display of items based on their association with a particular political organization and bans the display of websites based on the authors' viewpoint with respect to the Holocaust and anti-Semitism. A United States court constitutionally could not make such an order. The First Amendment does not permit the government to engage in viewpoint-based regulation of speech absent a compelling governmental interest, such as averting a clear and present danger of imminent violence.²⁵⁵

In the last section of his order granting Yahoo's motion for summary judgment, Judge Fogel addressed the impact his order would have on international comity. He observed that the extent to which the United States honors the judicial decrees of foreign nations is a matter of choice, governed by the "comity of nations," and that there is no absolute obligation to accept those foreign decrees.²⁵⁶ He also noted that comity does not require a U.S. court to give effect to a foreign judicial order if it would violate American public policy or fundamental interests.²⁵⁷ This led the court to conclude that the goal of giving effect to foreign judgments under the principle of comity could not be met in this case because the French court's May 22 and November 20 Orders violated Yahoo's First Amendment rights, which are fundamental rights for U.S. citizens.²⁵⁸ Judge Fogel ended by noting that he is bound to uphold the First Amendment in this case instead of giving deference to comity because there is no international agreement or treaty providing any other guidelines

255. *Id.* (citations omitted).

256. *Id.* at 1192 (citing *Hilton v. Guyot*, 159 U.S. 113, 163 (1895)).

257. *Id.* (citing *Laker Airways Ltd. v. Sabena Belgian World Airlines*, 731 F.2d 909, 931 (D.C. Cir. 1984); *Tahan v. Hodgson*, 662 F.2d 862, 864 (D.C. Cir. 1981); *Somportex Ltd. v. Phila. Chewing Gum Corp.*, 453 F.2d 435, 440 (3d Cir. 1971)).

258. *Id.* at 1192-93. In support of this portion of the order, the court cites three cases in which U.S. courts rejected efforts to enforce in the U.S. libel judgments granted under British law. These three cases found that the British defamation law impinged on protected free speech under the First Amendment. *See Matusevitch v. Telnikoff*, 877 F. Supp. 1, 4 (D.D.C. 1995); *Abdullah v. Sheridan Square Press, Inc.*, No. 93 1994 WL 419847 (S.D.N.Y. May 4, 1994); *Bachchan v. India Abroad Publ'ns, Inc.*, 585 N.Y.S.2d 661 (Sup. Ct. 1992). As noted previously, the court cites no cases directly on point for the situation where online content violates the law in one country but is lawful in the country where the online provider maintains its principal servers and offices.

for the handling of these kinds of cases: “Absent a body of law that establishes international standards with respect to speech on the Internet and an appropriate treaty or legislation addressing enforcement of such standards to speech originating within the United States, the principle of comity is outweighed by the Court’s obligation to uphold the First Amendment.”²⁵⁹

The question of whether the creation of international standards is desirable and/or possible, or whether other proposals exist to address the unacceptable recurrence of stalemates such as have occurred in the Yahoo case, is addressed in the final section of this article.

C. The Real Issue: It’s Their Law Against Ours

Oral argument on the appellate case was presented in December 2002, and the Court of Appeal has not yet issued its decision. How might this case be resolved? One hoped for result could be that the Court of Appeal will find that there is no actual controversy here and that in essence what Yahoo was seeking was an advisory opinion from the District Court. This result would leave all of the substantive issues addressed in the District Court decision unresolved, with the questions of jurisdiction and the conflict over the regulation of online speech to be argued another day in another context. Alternatively, the Court of Appeals could find that the district court erred in its determination that the exercise of jurisdiction over the French defendants was reasonable. This result might help to clarify the extent to which jurisdiction in Internet related cases will extend, and restrain the application of the effects test to tortious conduct. A third possibility is that the district court decision could be affirmed on all counts, as the Court of Appeals may want to reassert that U.S. courts will find jurisdiction in cases where it feels it necessary to protect First Amendment free speech. Such a finding would reaffirm that U.S. courts will find jurisdiction on minimal showings of fact, particularly in declaratory relief cases where the right of a foreign party to seek enforcement of a foreign judgment whose principal ruling is contrary to U.S. policy is involved.

Whichever of these alternatives the Court of Appeals chooses will have no impact on the real controversy this case has revealed. Recently an *ABA Journal* article captured the conflict’s essence with a subheading

259. *Yahoo! II*, 169 F. Supp. 2d at 1193. In a footnote, the Court added: “The Court expresses no opinion as to whether any treaty or legislation would or could be constitutional.” *Id.* at 1193 n.12.

stating, "It's Their Law Against Ours."²⁶⁰ An attorney quoted in the article, Robert Corn-Revere, put the importance of the First Amendment in international law in this perspective. He explained,

On the Internet, the First Amendment is just a local ordinance The practice of law on the Internet is still a very new, very unsettled area, and a lot of American companies that do business on the Net are likely to find themselves in trouble somewhere in the world.²⁶¹

Counsel for Yahoo adopted a more defiant view of the likelihood of getting into trouble for posting content in the U.S. that violates the laws of other countries. Bolstered by the district court's decision, Yahoo Attorney Mary Wirth dismissed the risk of foreign claims against the company. She stated, "Now, every time I get a defamation claim from some other country about something on the U.S. site, I say 'Go ahead and sue.' . . . Even if they rule against us, it's unenforceable."²⁶²

Mary Wirth may indeed be right. So long as Yahoo stays within U.S. borders, foreign judgments against it that conflict with the company's First Amendment rights will be unenforceable in the United States. However, an ominous development in the case is the companion suit, filed in Paris by a group of Holocaust survivors, against Tim Koogle, chief executive officer of Yahoo, based on the same violation of French criminal law.²⁶³ In that case, the damages sought were symbolic, only one French franc, but it is not beyond the realm of possibility that another case may be filed in the future, seeking greater damages or even criminal penalties. If that happens, then chief executives may have to adjust their travel plans to stay out of countries where enforceable judgments have been entered against them.

U.S. authorities would be hard-pressed to object to such tactics, since they have already been deployed in this country in the highly publicized

260. Jason Krause, *Casting a Wide 'Net*, 88 A.B.A. J. 20 (2002) (discussing both the *Yahoo!* case and the negotiations which took place between Google, a Silicon Valley search engine, and the Chinese Government). Chinese authorities blocked Chinese surfers from accessing the Google site, redirecting them to government-controlled sites instead. The matter was settled following weeks of negotiation, and the government blacklist was lifted. The article does not disclose the terms of the settlement. *Id.*

261. *Id.* The article closes with this other comment from Corn-Revere: "As the Internet gets more popular around the world, Americans are going to find there's a big difference between doing business on the Net here and doing it overseas." *Id.*

262. *Id.*

263. Associated Press, *Group Sues Yahoo CEO Over Site's Nazi Auctions*, S.F. CHRON., Jan. 24, 2001, at B3.

case filed by the United States government against Russian computer engineer Dmitry Sklyarov for alleged violation of the anti-circumvention sections of the Digital Millennium Copyright Act (“DMCA”).²⁶⁴ The *Sklyarov* case and the case against Koogle are mirror images of each other. They caution that parties who host or post website information legal in their own country but illegal in another country may be safe within their own borders if enforcement of the foreign judgment would violate the laws or policies of their country. Venturing outside their borders, however, may no longer be safe.

Before Yahoo’s California lawsuit, Professor Joel Reidenberg²⁶⁵ offered the optimistic view that U.S. courts would find sufficient basis to grant enforcement of the French court’s order. He envisioned that the enforcement ordered therein, for the imposition of geo-location filtering, would usher in a new age of democratization on the Internet.²⁶⁶ As a result of the French court’s decision, Reidenberg predicted, “Internet companies and developers of infrastructure technology will be forced to recognize and accommodate varying national public values.”²⁶⁷

264. *United States v. Elcom Ltd.*, 203 F. Supp. 2d 1111 (N.D. Cal. 2002). Sklyarov developed a software program, Advanced eBook Processor (“AEBPR”) for his Russian employer Elcomsoft, which permits eBook owners to translate electronic books from Adobe’s secure eBook format to the more common Portable Document Format (“PDF”). When Adobe complained to the U.S. Government, an indictment was issued for Sklyarov’s arrest based on his actions, which were allegedly in violation of DMCA anti-circumvention and anti-trafficking statutes. *Id.* at 1119. Although Sklyarov’s actions were lawful under Russian law, they were allegedly a violation of U.S. law. When Sklyarov came to the United States in the Fall of 2001 to speak at a conference on computer security, he was arrested and jailed for a period of five months. Elec. Frontier Found., *Frequently Asked Questions (and Answers) About the Dmitry Sklyarov & ElcomSoft Prosecution*, available at http://www.eff.org/IP/DMCA/US_v_Elcomsoft/us_v_elcomsoft_faq.html (last visited Dec. 7, 2003). He was released when he agreed to provide testimony sought by the Government. *Id.* His motions, and those of Elcomsoft, to dismiss the case for lack of subject matter jurisdiction and on First Amendment grounds, were dismissed in March and May 2002, respectively. Elec. Frontier Found., “*Intellectual Property: Digital Millennium Copyright Act (DMCA): U.S. v. ElcomSoft & Sklyarov*” Archive, available at http://www.eff.org/IP/DMCA/US_v_Elcomsoft/ (last visited Dec. 7, 2003). In December 2002, the defendants were acquitted of the charges, in part as a result of the successful assertion of a fair use defense. *Id.*

265. Professor Joel Reidenberg is a professor at Fordham University.

266. Joel R. Reidenberg, *Yahoo and Democracy on the Internet*, 42 JURIMETRICS J. 261, 261 (2002).

267. *Id.* at 272. Reidenberg asserted that the French court’s rejection of Yahoo’s argument that individual countries could not attempt to impose their local laws on the Internet was a positive sign that “Internet companies cannot supplant the rule of law as established by elected representatives.” *Id.* at 275. He suggested that the French court’s

The decisions in the *Yahoo!* and *Sklyarov* cases appear to be taking us away from Reidenberg's laudable view, and instead in the opposite direction. Yahoo's counsel makes clear that Yahoo has no intention of modifying its infrastructure technology to accommodate the laws of France. The principal impediment to the democratization process espoused by Professor Reidenberg is that U.S. courts, as evidenced by the decisions in the *Yahoo!* case, will not enforce foreign laws that conflict with the First Amendment. Since the courts cannot offer us a path to democratization of the Internet, then we must look elsewhere for possible solutions to this impasse.

III. A MODEST PROPOSAL FOR INTERNATIONAL GUIDELINES, AND A FORMULA FOR NOTICE AND A REDUCTION OF UNCERTAINTY

The search for a solution requires a synthesis of Professor Geist's targeting proposal, Professor Reidenberg's suggestion that geo-location filtering be deployed as a way for nations to control content coming to their own population while leaving the rest of the Internet untouched, and one new element to be added to the mix. This last element is the creation of an international database to provide more notice to website hosts and providers of the scope and nature of each nation's content laws. The database would allow providers and hosts a greater opportunity to choose whether, and to what extent, they wish to tailor their content or impose filters. By relying on the database, they could avoid or limit liability.

A. The Movement Towards an International Treaty Regulating Jurisdiction and Enforcement of Judgments

In his footnote regarding the prospects for an international treaty regarding online content, Judge Fogel noted that he could offer no opinion as to whether any treaty or agreement restricting online content could pass constitutional scrutiny.²⁶⁸ Despite this skepticism, efforts have been underway before two international organizations to address these issues via the treaty process.

For the past seven years, a Special Commission of the Hague Conference on Private International Law has been working on a Convention on Jurisdiction and Foreign Judgments in Civil and Commercial Matters. This effort has produced several drafts. The most

position "promotes democratic pluralism on the Internet by requiring technological developments that allow states to enforce their local laws." *Id.*

268. *Yahoo! II*, 169 F. Supp. 2d 1181, 1193 & n.12 (N.D. Cal. 2001).

recent, the *Preliminary Draft Convention on Jurisdiction and Foreign Judgments in Civil and Commercial Matters* (the “Preliminary Draft”), was completed in November 1999, and remains the subject of considerable discussion.²⁶⁹ The purpose of this drafting effort is to try to reach an agreement between the forty-seven signatory nations to the Hague Conference on which rules will govern jurisdiction and recognition and enforcement of foreign judgments in civil and commercial matters.²⁷⁰

While a number of the terms of the Preliminary Draft might have changed the outcome of the *Yahoo!* case had they been in effect,²⁷¹ one key term guarantees the same outcome and negates the possibility of the Hague Convention having any significant impact on the impasse the *Yahoo* case produced. Article 28 lists the grounds for refusal of recognition or enforcement of a foreign court’s judgment. The term provides “recognition or enforcement of a judgment may be refused if . . . recognition or enforcement would be manifestly incompatible with the public policy of the State addressed.”²⁷²

This “public policy” exemption is the same ground that Judge Fogel used to deny the application of the principles of comity to the enforcement

269. Preliminary Draft Convention on Jurisdiction and Foreign Judgments in Civil and Commercial Matters, Oct. 30, 1999, amended Nov. 16, 1999 [hereinafter Preliminary Draft], at <http://www.hcch.net/e/conventions/draft36e.html> (last visited Nov. 27, 2002).

270. See generally Edward C.Y. Lau, *Update On The Hague Convention On The Recognition & Enforcement Of Foreign Judgment* (2000), at <http://www.launet.com/Hague/Update031700.html> (last visited Nov. 27, 2002).

271. See, e.g., Preliminary Draft, *supra* note 269, art. 4.1, 10, 18.1(h) & 21. Article 4.1 places jurisdiction in the forum agreed upon by the parties. *Yahoo* has argued that its Terms of Service establish jurisdiction over disputes in the United States. See *supra* note 13. Adoption of an international agreement establishing jurisdiction on this basis might be a ground for *Yahoo* to assert that the French Court could not exercise jurisdiction, unless the act arose from a tort, in which case the provisions of Article 10 might govern. Article 10 reaffirms the “effects test” as applied to tort cases, affirming that jurisdiction in tort matters may be found in the forum where the injury occurred, provided it was foreseeable that the act would result in an injury to the person in that forum. Preliminary Draft, *supra* note 269, art. 10. Article 18.2(h) provides that jurisdiction may not be established based upon the service of a writ upon the defendant in a particular forum—a term which would negate *Yahoo*’s argument that the French Defendant’s service of the Notice of Entry of Judgment of the French Court on *Yahoo* in California was a basis for finding jurisdiction over those defendants. *Id.* art. 18.2(h). Article 21 creates a priority of actions by establishing a *lis pendens* system—requiring, under certain circumstances, the court in a second forum to suspend proceedings in favor of another court which first seized jurisdiction. *Id.* art. 21. The application of this rule might, but for the “public policy” exception found and discussed in Article 28, have been a basis for suspending the U.S. District Court case until all proceedings in the French Court were completed.

272. *Id.* art. 28.1(f).

of the French court's orders.²⁷³ To date, there is no indication that the Hague Conference intends to alter Article 28 or to make any effort to reconcile the "public policy" exemption doctrine with the kind of situation posed by the Yahoo cases.

The need for a uniform system of jurisdiction and recognition of judgments in private international law, particularly where those judgments arise from intellectual property disputes, has also drawn the attention and efforts of the Geneva-based World Intellectual Property Organization ("WIPO"). In January 2001, distinguished Professors Rochelle C. Dreyfuss and Jane C. Ginsburg presented WIPO delegates with a *Draft Convention On Jurisdiction And Recognition Of Judgments In Intellectual Property Matters* (the "WIPO Draft"). The draft provided suggestions for dealing with the regulation of online content, while still retaining the same "public policy" exemption found in the Hague Convention.²⁷⁴

The authors based the WIPO Draft on the Hague Convention, but tailored the draft specifically to the issues arising from the recognition and enforcement of intellectual property judgments, as opposed to the more general approach of the Hague Convention.²⁷⁵ One such issue identified by the WIPO Draft is the problem of adjudicating multiterritorial claims.²⁷⁶ The draft's authors suggested in Article 13, adapted from Article 22 of the Brussels Convention, that such multiterritorial claims be consolidated and heard by a single court in a single forum.²⁷⁷ Article 13(3) contains a series of factors for parties to consider in deciding which court should hear the matter.²⁷⁸ Subsection 5 provides that if consolidation is not agreed to, the judgment in one action will not be preclusive of the other.²⁷⁹

273. *See supra* notes 254-255.

274. *See* Rochelle C. Dreyfus & Jane C. Ginsburg, *Draft Convention On Jurisdiction and Recognition of Judgments In Intellectual Property Matters*, 77 CHI.-KENT L. REV. 1065 (2002) [hereinafter *Draft Convention on Intellectual Property Matters*].

275. *Id.* at 1065-66.

276. *Id.* at 1073. The authors of the WIPO Draft asserted that the digital networked environment makes the likelihood of multiterritorial infringements likely, necessitating the creation of a method for adjudicating such disputes in a single forum to avoid inconsistent judgments.

277. *Id.* at 1080.

278. *Id.* at 1081.

279. *Id.* The factors for consideration in the consolidation analysis include: "the advantages of worldwide resolution of the dispute among the parties through consolidation of related pending actions, and through inviting the parties to assert all intellectual property claims related to the action in a single forum," *id.* at 1080, "whether consolidating would promote efficiency and conserve judicial resources and the resources of the parties," *id.*, "whether or not inconsistent judgments could result if multiple courts adjudicated the related claims," *id.*, and "[in patent cases, the expertise of the judicial

In further recognition of the multiterritorial nature of intellectual property disputes, the WIPO Draft authorizes the kind of injunctive relief the French court ordered in the *Yahoo!* case. Article 19(1) provides: “The court having jurisdiction under the rules of this Convention to determine the merits of the case has jurisdiction to order any provisional or protective measures, including transborder injunctions.”²⁸⁰ Had the parties agreed to the consolidation of the French and U.S. actions in the *Yahoo!* case, Article 19(1) would have made Judge Gomez’s interlocutory order directing the use of geo-location filtering software on Yahoo’s servers in Santa Clara, California an enforceable trans-border injunction. However, the reiteration of the Hague Convention’s “public policy” ground for refusal of recognition or enforcement, found in Article 25 of the WIPO Draft,²⁸¹ again leaves us with the same result reached by Judge Fogel: The French court orders are unenforceable as violative of Yahoo’s First Amendment rights.

B. Towards an Interim Solution: Effects Plus Targeting Plus Increased Foreseeability—Breaking the Egg in the Middle

Although the Hague Convention and the WIPO Draft offer many valuable proposals to move us in the direction of a more universal system of recognition and enforcement of judgments, they do not provide any viable solutions to the impasse epitomized by the *Yahoo!* cases. It appears unlikely that any treaties, conventions, or other public international legal bodies will undertake the seemingly impossible task of reaching agreement on the regulation of online content, for the simple reason that countries’ substantive laws differ on what content is acceptable. If a uniform international system of laws regulating content is unattainable, are we left just with the status quo of an endless series of wars fought between nations over which end of the egg to break?

This bleak outlook is not all we have to look forward to. The suggestions of a number of the commentators referenced previously hold out the prospect of improvement, if not solution, of this problem. Professors Geist and Reidenberg, in particular, offer a way out of our dilemma. Professor Geist points out that the foreseeability of jurisdiction

system of the Contracting State in which the court seized is located].” *Id.* The authors put all suggestions regarding patent law in brackets based on their view that patent litigation should remain outside the convention due both to the lack of universal expertise among convention members required for accurate decision making, and to the low incidence of simultaneous multinational infringements. *Id.* at 1069.

280. *Id.* at 1084.

281. *Id.* at 1086. This section of the WIPO Draft is based on Article 28 of the Hague Convention. *See supra* note 258 and accompanying text.

being imposed in a given situation is at the heart of the reasonableness standard.²⁸² In short, if a party can reasonably expect that its conduct will result in the imposition of jurisdiction in a foreign country, it can make an informed decision as to whether it wishes to take the risk of posting online content that might give rise to liability.²⁸³

Neither the *Zippo* passive/active test nor the *Calder* effects tests, Geist argues, have proven very useful or accurate in providing certainty to online content providers as tests of the foreseeability of the imposition of jurisdiction.²⁸⁴ As an alternative, he offers a “targeting test” through which jurisdiction would be based on the application of three factors: contracts, technology, and actual or implied knowledge.²⁸⁵ The contracts factor looks to the evidence of agreement between the parties on forum selection as a means of determining foreseeability. Geist acknowledges that in some contexts, particularly consumer transactions, courts must use care in interpreting “clickwrap” agreements as true acknowledgments of forum selection.²⁸⁶ He notes that the conditions of assent are also significant and bear scrutiny.²⁸⁷ He suggests that self-declaration, where a consumer affirmatively declares where jurisdiction is acceptable to them, is a better alternative.²⁸⁸ The second targeting factor, technology, refers to the use of software programs such as the geo-location filtering ordered by Judge Gomez. Geist provides a detailed analysis of some of the technology methods available at the time his article was written (early 2001), all of which allow online content providers to identify, and where appropriate, block access based on jurisdiction or user identity.²⁸⁹

Geist acknowledges that these screening methodologies are not foolproof, but points out in response that “few users of technology actually demand perfection.”²⁹⁰ As noted previously, Judge Gomez concluded that

282. Geist, *supra* note 35, 1356.

283. *Id.* at 1355-56. Professor Geist points out that worldwide Internet availability makes foreseeability difficult to gauge. *Id.* at 1357.

284. *Id.* at 1371-80.

285. *Id.* at 1380.

286. *Id.* at 1387.

287. *Id.* at 1386-87.

288. Geist, *supra* note 35, at 1391. Professor Geist acknowledges that self-declaration has its problems, notably when a mistake in the self-declaration occurs. In those circumstances, he points out that “courts have ruled that companies cannot rely on the self-declaration of a user where they know or suspect it to be false.” *Id.* at 1392 (citing *People v. World Interactive Gaming*, 714 N.Y.S.2d 844 (Sup. Ct. 1999)).

289. *Id.* at 1393-1401.

290. *Id.* at 1394 (quoting Lawrence Lessig, *The Zones of Cyberspace*, 48 STAN. L. REV. 1403, 1405, when he opined that “[a] regulation need not be absolutely effective to be sufficiently effective”).

the 80-90% level of protection the experts predicted would follow geo-location filtering and self-identification on the Yahoo U.S. site would be sufficient.²⁹¹

Professor Reidenberg enthusiastically supports Professor Geist's recommendation that courts accept and adopt the use of these technological aids.²⁹² Reidenberg hopes that through the application of technology to identify users and to protect national laws through screening of content, "states will regain their voice in the global network as participants in a pluralistic international democracy."²⁹³

At its core, the pro-filtering argument relies on the assurance that courts will choose to under-filter as opposed to over-filter. Geo-location filtering seems to offer just such an assurance, since there appears to be no evidence that a properly configured geo-location software program will screen out more than those users who come from the designated location.²⁹⁴ Further, as noted, there appears to be little opposition or dissatisfaction with the less than complete success of this software in screening out all prospective users coming from the targeted location.²⁹⁵

The last targeting factor discussed by Professor Geist is the assessment of whether the content provider has actual or implied knowledge of the geographical location of the user or participant in the online activity. Citing the tort cases that gave rise to the *Calder* "effects analysis," he notes that courts have generally accepted that the defendant either knew or should have known that defamatory content would cause injury in the plaintiff's home forum. He points out that courts have applied this same knowledge-based analysis in Internet gambling and intellectual property disputes.²⁹⁶

This factor of the targeting analysis gives rise to an issue when applied to the *Yahoo!* cases. While it is fairly easy in tort cases or infringement cases to prove knowledge, actual or implied, attributable to the defendant, the situation in the *Yahoo!* cases poses a more difficult issue of proof.

291. See *supra* notes 114-128.

292. Reidenberg, *supra* note 266, at 261-280.

293. *Id.* at 280.

294. See Beverley Earle & Gerald A. Made, *International Cyberspace: From Borderless to Balkanized*, 31 GA. J. INT'L & COMP. L. 225, 258 (2003); see also Matthew Fagin, *Regulating Speech Across Borders: Technology vs. Values*, 9 MICH. TELECOMM. & TECH. L. REV. 395, 424 (2003).

295. See Fagin, *supra* note 294, at 424.

296. Geist, *supra* note 35, at 1402-04 (citing *Star Media Network, Inc. v. Star Media, Inc.*, No. 00 Civ. 4647, 2001 WL 417118 (S.D.N.Y. Apr. 23, 2001); *People v. World Interactive Gaming*, 714 N.Y.S.2d 844 (Sup. Ct. 1999)).

How, one asks, is Yahoo or any other company to know whether the content it posts or allows to be posted will violate the laws of another country? Yahoo's efforts to shift the burden of obtaining that knowledge via its Terms of Service²⁹⁷ were rejected by the French court, but the question remains unanswered.

Geist has acknowledged that although he believes website hosts need to be aware of the laws of foreign countries, an effects test without an actual or implied knowledge requirement would pose great risks and uncertainties for providers:

First, foreign law matters. Once a company has assets or customers in a foreign country, it can ill afford to ignore the local legal system.

....

The movement toward an effects based analysis marks an important shift in the understanding of Internet jurisdiction since it may breed increased uncertainty for site operators. While a Web site [sic] operator may be aware of its effect locally, it is unrealistic to expect the site operator to identify the effects in every jurisdiction worldwide.²⁹⁸

One possible solution to this notice problem would be to develop a system for providing notice to all parties who obtain the right to host a website available to the public on the Internet. At present, any person who wishes to host a website must apply for a domain name (essentially an address on the Internet) from the Internet Corporation for Assigned Names and Numbers ("ICANN").²⁹⁹ This makes ICANN an ideal locus for the dissemination of information about the content laws of all of the nations of the world where the Internet is presently accessible.

The Internet Assigned Numbers Authority ("IANA") maintains the list of all country code Top Level Domain names ("ccTLDs"), of which 243

297. See *supra* note 13 and accompanying text.

298. Lionel Thourmyr, *The Legal Implications of the Yahoo!, Inc. Nazi Memorabilia Dispute: An Interview with Professor Michael Geist*, JURISCOM.NET, at <http://www.jurisc.com.net/en/uni/doc/yahoo/geist.htm> (last visited Nov. 27, 2002).

299. See ICANN, at <http://www.icann.org> (last visited Nov. 27, 2002). ICANN is a technical coordination body for the Internet. A coalition of Internet communities created ICANN in 1998, and assumed responsibility for a set of technical functions previously performed by IANA and other groups under the supervision of the U.S. Government. ICANN assigns all Internet domain names, IP address numbers, and coordinates the functionality of all Internet protocol parameters and port numbers.

have been assigned to the countries of the world to date.³⁰⁰ If each of those nations were required, as a condition for the right to obtain and maintain their ccTLD, to provide ICANN with a copy of or at least a link to all of the statutes identifying what forms of expressive content would, if posted and accessible to their citizens, be deemed a violation of their local laws, a gap in the present notice available to website providers would be filled.

Filling this gap would provide further foreseeability for website providers to determine whether content they host or allow to be posted would cause them to incur liability in a foreign country. The choice to post or host would then become an informed business decision of whether to incur the risk of suit in a foreign court, to employ technology to filter visitors from that country, or to employ blocking software based on user identification from that country's users. A website host who employed such technology could, as an incentive to make that use, be deemed exempt from liability for claims arising from parties who circumvented the technology and gained access to the prohibited material.

The implementation of this proposal is fraught with landmines and problems. What happens if a country fails to update its data and a party violates a new law not posted on the ICANN database? Is the burden of reviewing this database too onerous, such that imposing it on small companies is unfair? These are legitimate issues, but not so daunting as to render the proposal one not worth considering as a way to fill the foreseeability gap that now exists.

IV. CONCLUSION

Judge Fogel concluded his analysis of the comity issue in *Yahoo! v. LICRA* by noting that there are no international standards governing the regulation of online content.³⁰¹ His French counterpart, Judge Gomez, agreed that there are no such standards. Judge Gomez hoped that they would be developed to provide guidance for courts in the future.³⁰² While establishing such guidelines is undeniably difficult, the certainty that cases similar to the Yahoo-LICRA dispute will continue to challenge the international legal system mandates that the international community undertake the effort, if only because in the process we may be able to better preserve the principles of international comity—principles that the

300. See *IANA Root-Zone Whois Information*, at <http://www.iana.org/cctld/cctld-whois.htm> (last visited Nov. 27, 2002).

301. See *supra* note 239 and accompanying text.

302. Interview, Gomez, *supra* note 160, at 8.

uncertainties of Internet jurisdiction may fundamentally endanger.³⁰³ For now, the adoption of the targeting approach to jurisdiction, enhanced by the creation of a database of content related laws of each country, may prevent a repetition of the endless battles presented by the *Yahoo!* case and its progeny.

303. Thournyr, *supra* note 298.

CYBERSPACE TECHNOLOGICAL STANDARDIZATION: AN INSTITUTIONAL THEORY RETROSPECTIVE

By Daniel Benoliel[†]

ABSTRACT

Acknowledging the value of standard-setting, the Clinton administration originally made “industry self-regulation” the guiding principle for standardizing the Internet. The succeeding administration continues to use this approach. Nonetheless, historical analysis of the last two decades shows that industry self-regulation has not always been a coherent policy. Rather, it has become a rhetorical device used to legitimize the government’s agendas, particularly the mandated design of cyberspace’s architecture and infrastructure. To date, there are still too many inconsistencies in the government’s formal standardization policies. For example, the government’s policy of centralizing early infrastructure standards to mandate cyberspace’s architecture is in tension with its actions aimed at privatizing the Internet’s funding and governance. These contrasting policies demonstrate that “industry self-regulation” of cyberspace has actually included a large measure of government intervention.

This paper is a historical and conceptual assessment of the government’s standardization policies using a comparative institutional theory approach. After assessing standardization policies, this study considers the unique, multi-layered architecture of cyberspace to identify which institutional body should standardize the Internet. To do so, this study identifies a distinctive production process for cyber standards that distinguishes between the standardization of early infrastructure and the standardization of complementary applications.

© 2003 Daniel Benoliel

[†] J.S.D. candidate, University of California at Berkeley, School of Law (Boalt Hall). A version of this paper won first prize in the 30th Telecommunications Policy Research Conference (“TPRC”) Graduate Student Article Competition and was later presented at that conference and at the John M. Olin Law and Economics Workshop Seminar, at Boalt Hall. For their most helpful comments and support, I am indebted to Mark Lemley, Pamela Samuelson, Edward Rubin, David Post, Stuart Benjamin, Hal Varian, Bob Cooter, Jesse Fried, Dan Hunter, Polk Wagner, Jane Winn, and the participants at both fora. I am also grateful for the generous advice I received from Carl Cargill and Roger Martin from Sun Microsystems, Inc. Any inaccuracy in this paper is my own responsibility. For further questions or comments, please contact me at: Daniel_b@boalhall.berkeley.edu.

This study concludes with a set of policy rules. These rules are proposed with a caveat: cyberspace's present successful institutional regulative reality might not be preserved into future technological generations unless distinctive policies are maintained both sequentially in compliance with the technological standardization lifecycle, and contextually in compliance with the different categories of standards.

TABLE OF CONTENTS

I.	INTRODUCTION	1261
II.	THE TECHNOLOGICAL TRIPLE SCRUTINY ANALYSIS: TIME, SPACE, AND INSTITUTIONAL IDENTITY	1266
	A. IT Standards: The Methodological Framework.....	1266
	B. The Technological Lifecycle: The Criterion of "Time"	1267
	C. Cyberspace's Architectural Edifice: The Criterion of "Space"	1272
	1. <i>The Normative Framework</i>	1272
	2. <i>Infrastructure Standards</i>	1276
	3. <i>Application Standards</i>	1279
	D. Standard-Setting Institutions: The Criterion of "Institutional Identity"	1280
III.	THE INFRASTRUCTURE DEVELOPMENT PHASE: A POLITICAL INSTITUTIONAL INEVITABILITY	1285
	A. Infrastructure Telecommunications Services: The Collective Choice Analysis.....	1285
	B. Infrastructure Standardization: The Cost-Benefit Analysis.....	1293
	1. <i>Administrative Costs</i>	1295
	a) Quality Standardization Costs	1295
	b) Development Costs.....	1298
	c) Information Costs	1299
	2. <i>Compliance Costs</i>	1301
	a) Variety Reduction Costs.....	1301
	b) Coordination Costs	1305
	3. <i>Indirect Costs</i>	1308
IV.	THE MODIFICATION PHASE: THE COMMERCIALIZATION OF CYBER STANDARDS .	1309
V.	THE IMPLEMENTATION PHASE: THE RISE OF AUTONOMOUS INSTITUTIONS.....	1313
	A. Application Standardization: The Cost-Benefit Analysis.....	1314
	1. <i>Administrative Costs</i>	1315
	a) Eliminating Duplicative Efforts.....	1315
	b) Reducing Search Costs	1317
	2. <i>Compliance Costs</i>	1318
	3. <i>Indirect Costs</i>	1321
	B. Government Intervention	1323
	1. <i>Direct Intervention: The Problem of Efficiency</i>	1323
	2. <i>Indirect Intervention: Roles of Government Regulation</i>	1326
	a) Regulate Production Supervision Rules	1327
	b) Regulate the Process of Standardization.....	1329
	C. A Potential Deviation: The ICANN Case Study	1331
VI.	CONCLUSIONS	1336

I. INTRODUCTION

The regulative technique commonly known as technological standard-setting is one of the most effective ways to regulate cyberspace and shape its markets.¹ This regulative technique is too often seen as an overly technical discipline and has been neglected as a field of research on cyberspace regulation. Technological standard-setting has been given too little attention, both as an independent field of regulation theory and, more specifically, as a means of governing network environments such as the Internet.

As with other technological fields of mass media standardization,² cyberspace has reached the degree of comprehensiveness that makes it worthy of a wider perspective of comparative institutional analysis³ than has been given to it by the U.S. government. Notably, rather than conducting a principled analysis of the appropriate standardization approach to the Internet, the Federal Communications Commission ("FCC") has divided the network environments into incomplete legal categories by adopting the "basic" and "enhanced" service classifications prior to the widespread adoption of the Internet.

Overall, efforts to set standards in cyberspace have occasionally been ambiguous and inconsistent. The government's conflicting policies demonstrate this inconsistency. On one hand, government seeks to privatize the Internet's funding and governance particularly through the Internet Corporation for Assigned Names and Numbers ("ICANN") self-regulation

1. On the standardization discipline as an independent form of regulation, see STEPHEN BREYER, *REGULATION AND ITS REFORM* 96 (1982) (for an economic perspective); ANTHONY I. OGUS, *REGULATION: LEGAL FORM AND ECONOMIC THEORY* 150 (1994). See also CARL F. CARGILL, *OPEN SYSTEMS STANDARDIZATION: A BUSINESS APPROACH* 26-29, 137-38 (1st ed. 1997) (for an information technology ("IT") perspective); Joel R. Reidenberg, *Governing Networks and Rule-Making in Cyberspace*, 45 EMORY L.J. 911, 918, 927-28 (1996) [hereinafter Reidenberg, *Governing Networks*] (concluding that standards in cyberspace embed policy choices, thus supplementing legal rules); Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553, 570-72 (1998) [hereinafter Reidenberg, *Lex Informatica*] (for the cyberspace context).

2. For example, broadcast, cable and satellite, TV, and radio.

3. On the need for technological comprehensiveness for standard-setting, see, for example, MARTIN C. LIBICKI, *INFORMATION TECHNOLOGY STANDARDS: QUEST FOR THE COMMON BYTE* 16-18 (1995); Tineke M. Egyedi, *Institutional Dilemma in ICT Standardization: Coordinating the Diffusion of Technology*, in *INFORMATION TECHNOLOGY STANDARDS AND STANDARDIZATION: A GLOBAL PERSPECTIVE* 57 (Kai Jakobs ed., 1999); Ole Hanseth & Eric Monteiro, *Participatory Standardization of Information Infrastructure*, in *INTERNATIONAL PERSPECTIVES ON INFORMATION SYSTEMS: A SOCIAL AND ORGANISATIONAL DIMENSION* 174 (Savvas Katsikides & Graham Orange eds., 1998).

case study, while on the other, it still seeks to centralize and even monopolize the Internet's infrastructure standard-setting activities.

The commercialization, widespread use, and development of peripheral standardized software products for the Internet raise the question of who should standardize the Internet. One can answer this question by comparing the experience of analogous technological fields and retrospectively examining the experience of the Internet over the last two decades.⁴ This study undertakes this examination while confronting the unique progress of what is essentially a technological standardization process.

This study proceeds from Neil Komesar's comparative institutional theory insight. Komesar acknowledges that all standardization institutions are subject to both internal and external imperfections, and that only a comparative approach vis-à-vis the identical assignments should prevail.⁵ Accordingly, although an institution may function inefficiently, alternative institutions may function even worse. By the same token, even when the intrinsic worth of an institution might be apparent, alternative institutions may perform the same task more effectively.⁶ Thus, upon examining the market's ability to self-standardize cyberspace, the operative question is not how well the market functions, but whether political institutions, such as government branches and other autonomous institutions like industry standardization organizations, could produce a better outcome.

This study focuses predominantly on the relationship between institutional analysis and standardization *production* policy as an ex ante regulative mechanism.⁷ In addition, commercial implementation of a new technology with legal or technological standards can be viewed through the more general prism of regulation theory in cyberspace. Thus far, the ongo-

4. On the importance of precedents in designing standardization policies, see, for example, BREYER, *supra* note 1, at 99; JASON OXMAN, THE FCC AND THE UNREGULATION OF THE INTERNET (FCC OPP, Working Paper No. 31, 1999), available at http://www.fcc.gov/pub/Bureaus/OPP/working_papers/oppwp31.pdf (last visited Dec. 18, 2003); KEVIN WERBACH, DIGITAL TORNADO: THE INTERNET AND TELECOMMUNICATIONS POLICY (FCC OPP, Working Paper No. 29, 1997), available at http://www.fcc.gov/Bureaus/OPP/working_papers/oppwp29.pdf (last visited Dec. 18, 2003). Werbach cautions that there is reason to believe that analogies to familiar services may not be appropriate for the Internet due to real "category" difficulties. Henceforth, any such analogy will derive from the proposition of contextual analysis, unless claimed otherwise.

5. NEIL K. KOMESAR, IMPERFECT ALTERNATIVES: CHOOSING INSTITUTIONS IN LAW, ECONOMICS, AND PUBLIC POLICY 3-10 (1996).

6. *Id.*

7. An alternative approach would be to look at telecommunications policy ex post by examining the implications of antitrust and intellectual property law.

ing debate has focused on application standards as a regulative constraint, and has overshadowed the separate debate concerning system-oriented infrastructure standards.⁸ In the latter forms of discourse, any institutional choice is mostly a reflection of an earlier pursued policy or defined “rights,” for example, both Lessig’s constitutional urge for reducing “code control” and Johnson and Post’s freedom for regulative multiplicity. Institutional choices are mostly a reflection of an earlier pursued policy of defined legal rights.⁹ As a result, current discourse on institutional choices places legal and other normative principles above the political *production* process and costs.

The Komesarian proposition, on the other hand, suggests that the mere reflection of any social goals and “rights” on institutional choices is largely insufficient because such goals tautologically embed institutional choices of their own.¹⁰ Accordingly, social policy should become relevant only upon consideration of the proper institutional constraint.¹¹ This study follows Komesar’s theory and suggests that any institutional choice should be seen as an integral part of the general technological goal and not merely as its reflection. Indeed, the structure of cyberspace’s regulative institutions profoundly affects the prevalence and form of information technology (“IT”) standards.

8. See, e.g., LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE 101-02 (1999) (explaining that his book and the present regulative discussion is and should be aimed at the standardized application layer); see also Llewellyn J. Gibbons, *No Regulation, Government Regulation, or Self-Regulation: Social Enforcement or Social Contracting for Governance in Cyberspace*, 6 CORNELL J.L. & PUB. POL’Y 475, 481-87 (1997) (using the term cyberspace “infrastructure” while focusing on “application” standards, like e-mail or the World Wide Web, instead); David R. Johnson & David Post, *Law and Borders—The Rise of Law in Cyberspace*, 48 STAN. L. REV. 1367, 1382-91 (1996) (examining standardized applications for example, copyright and trademark regimes as the point of reference in their regulative argument); Timothy Wu, *Application-Centered Internet Analysis*, 85 VA. L. REV. 1163, 1164 (1999) (“[T]he whole Internet is rarely an appropriate level on which to generalize. Instead, legal thinking can better focus on where the variation that is apparent to the user is actually found: the application layer above the Internet’s basic protocols.”). See generally Dan L. Burk, *Federalism in Cyberspace*, 28 CONN. L.R. 1095 (1996) (focusing on consumer protection applications and other public laws to police online behavior and commerce). For a skeptical view of this trend, see CARL SHAPIRO & HAL R. VARIAN, INFORMATION RULES: A STRATEGIC GUIDE TO NETWORK ECONOMY 317-18 (1999) [hereinafter SHAPIRO & VARIAN, INFORMATION RULES].

9. LESSIG, *supra* note 8; Johnson & Post, *supra* note 8.

10. KOMESAR, *supra* note 5, at 43 (arguing that “calling something a “right” is an institutional statement”).

11. *Id.* at 271.

Based on the proposition that any institutional choice should be seen as an integral part of the general technological goal rather than a reflection of a previously pursued static policy or right, this study introduces a multi-layered production process for technological standardization. This study thus offers an alternative synthesis to the existing top-bottom, bottom-up and industry standardization organization's single-layered regulation models.¹² In general, all of the existing approaches seem to fall short with respect to the critical factor of *timing* because they fail to properly acknowledge the evolving nature of technological environments.

Technological environments such as cyberspace have processional nature. This study echoes that procession by measuring the strengths and weaknesses of each of these single-layered regulatory models in each sequential phase. It does so by looking at variations in the types of standards located on different layers of technical architecture from one production stage to the next. Overall, the institutional selection of a standard setter for each phase will be only a transitory choice among highly imperfect alternatives. Lastly, this study confronts the distinctive costs of standardization. These costs are found outside the partial scope of its institutional participatory process. Thus, this study suggests that whenever strict legal process analysis falls short in supplying policy makers with a comprehensive result, any institutional choice should be subject to production costs that exist beyond the participatory process *per se*.¹³ This study also departs from the supported proposition that current law and economics may be seen to share a joint objective and importance in standardizing cyberspace. As a result, a preliminary cost-benefit analysis is outlined in Parts II-V.

12. *Id.* at 6 (suggesting that, originally, the available institutions are political institutions i.e., executive and legislative branches of government, market and the courts). However, the cyberspace technological environment includes, in practice, additional autonomous institutions such as industry, group consortia, etc., while minimizing the role of courts. See discussion *infra* Part II.D. See generally JOSEPH FARRELL & GARTH SOLANER, COMPETITION, COMPATIBILITY AND STANDARDS: THE ECONOMICS OF HORSES, PENGUINS AND LEMMINGS (Univ. of Cal. Berkeley Dep't of Econ., Working Paper No. 8610, 1986).

13. See Edward L. Rubin, *The New Legal Process, the Synthesis of Discourse, and the Microanalysis of Institutions*, 109 HARV. L. REV. 1393, 1394, 1411-13, 1425-37 (1996) (supporting a comprehensive synthesis to law and economics and the legal process movements for comparative institutional analysis); see also JAMES G. MARCH & JOHAN P. OLSEN, REDISCOVERING INSTITUTIONS 1-2, 16-19 (1989); PAUL J. DIMAGGIO & WALTER W. POWELL, *Introduction to THE NEW INSTITUTIONALISM IN ORGANIZATIONAL ANALYSIS* 11-15 (Paul J. DiMaggio & Walter W. Powell eds., 1991); Neil K. Komesar, *Exploring the Darkness: Law, Economics, and Institutional Choice*, WIS. L. REV. 465, 466-71 (1997).

To narrow the conceptual framework of this study, Part II opens by examining how an IT standard becomes defined. It continues by defining the three technological benchmark criteria of the standardization realm: time (referring to the processional technological lifecycle), space (referring to cyberspace architecture, established by different types of technological standards), and institutional identity (referring to which institution is best situated to define, implement, and maintain standards). This sets forth a triple scrutiny analysis for the appropriate institutional choice in each of cyberspace's distinctive standardization phases.

Part III begins with the first of three standardization phases: the *infrastructure development* phase. New platform technology is typically introduced in this early technological phase, beginning with its evolution from an idea to the development of a basic system product or process. This creates the content of the first standardized infrastructure technology. In this early phase, central political institutional control, in the form of mandated government intervention in infrastructure standardization, for example, is inevitable.

Part IV continues to the *modification* phase and explains how rapid innovative changes were made to cyber technology. These changes were followed by extensive bargaining and modifications to support commercial use of Internet technology. These changes are what led to the third, and current, commercial standardization phase of the technological generation.

Part V describes the concluding technological *implementation* phase. Whenever technology matures, the explosion of new markets for both early core Internet telecommunications services and markets for application and conforming standardized products evolve and should promote the rise of autonomous standard-setting institutions. Part V also presents four arguments in support of this paper's comparative institutional analysis. The four arguments touch upon 1) standardization's chilling effect on governmental intervention, 2) government's supposedly limited direct role in application standardization, 3) government's indirect regulatory and supervisory roles in standardization, and 4) the ICANN lesson.

Part VI enunciates several suggestions as policy rationales for future technological generations in cyberspace. The main conclusion is that the unprecedented development of cyberspace provides theoreticians and decision-makers alike with a feasible opportunity to develop a comprehensive, time- and context-based institutional standardization policy. Furthermore, choosing the optimal standardization institution for each technological phase has three-fold significance in cyberspace institutional theory. First, a rational standard setter should be able to predict efficiently the de-

gree of compliance of each such standardized technology by recognizing different phases of technological development and different standardization institutions. Second, the rise of different types of IT standards demands transitory regulatory conduct for the different technological phases. Third, a comprehensive institutional framework can be established for future technological generations in cyberspace.

II. THE TECHNOLOGICAL TRIPLE SCRUTINY ANALYSIS: TIME, SPACE, AND INSTITUTIONAL IDENTITY

A. IT Standards: The Methodological Framework

Typically, there are two technical criteria used to define IT standards, such as the standards used in regulating cyberspace. First, a technological standard is primarily viewed with respect to the degree of its technical maturity. That maturity is most commonly assessed through a technological standardization process.¹⁴ Second, an IT standard is quantified as a function of its acceptance by the relevant market.¹⁵ This measurement is assessed through the intensity and the breadth of the standard's recognition and its use by consumers. That estimate typically derives from the recognition that the specified protocol or service provides significant personal and social benefits to users and the market respectively.

14. See Scott O. Bradner, *The Internet Standards Process—Revision 3*, RFC 2026 (IETF Network Working Group, Harvard University Oct. 1996) [hereinafter Bradner, *Internet Standards Process*] (noting that “specifications that are intended to become Internet Standards evolve through a set of maturity levels known as the ‘standards track’”), at <http://www.ietf.org/rfc/rfc2026.txt> (last visited Dec. 18, 2003). Eventually, the Internet Engineering Task Force defined these maturity levels as: “Proposed Standard”, “Draft Standard”, and “Standard”. *Id.* at § 4; see also FLOYD WILDER, A GUIDE TO THE TCP/IP PROTOCOL SUITE 368-70 (2d ed. 1998); *Internet Official Protocol Standards*, RFC 1800 (John Postel ed., IETF Network Working Group & Internet Architecture Board) (providing an earlier description of the maturity levels), at <http://rfc.sunsite.dk/rfc/rfc1800.html> (last visited Dec. 18, 2003); William J. Clinton & Albert Gore Jr., *A Framework for Global Electronic Commerce* (1997) [hereinafter Clinton & Gore, *A Framework*] (concluding that “[p]remature standardization . . . can lock in outdated technology”), at <http://www.technology.gov/digeconomy/framework.htm> (last visited Dec. 3, 2003); Egyedi, *supra* note 3, at 49 (offering a theoretical perspective).

15. See CARGILL, *supra* note 1, at 42 (providing a theoretical IT perspective); LIBICKI, *supra* note 3, at 18-19; Bradner, *Internet Standards Process*, *supra* note 14 (granting the strongest status, “Internet Standard”, only to those specifications that have already become widely adopted); cf. MARCH & OLSEN, *supra* note 13, at 50-52 (offering an institutional analysis perspective).

Not every technological specification meets both criteria.¹⁶ Technological developments placed on the standards track but that do not comply with both criteria are generally regarded as “non-standard” technology.¹⁷ Such non-standard technology usually lacks the minimum degree of acceptance.¹⁸ Another type of non-standard technology is found in specifications that had been previously defined as standards until they were superseded by an updated standard¹⁹ or otherwise fell into disuse.²⁰ In short, only specifications meeting both criteria are regarded as IT standards. As potentially cohesive and stable technologies, standardized specifications are the focus of cyberspace’s institutional policy planning.

B. The Technological Lifecycle: The Criterion of “Time”

The U.S. government’s formal standardization policy regarding the time constraint, either as a production investment restraint or as a function of a technological life cycle, was consistently ignored.²¹ The emerging investment literature already acknowledges that the ability of an investor to delay an irreversible investment may affect a decision on whether and

16. See Bradner, *Internet Standards Process*, *supra* note 14 (providing examples of standards which failed to attract wide acceptance, such as most of the ISO standards for data communications, and the IEEE 802.6 standard for Distributed Queue Dual-Bus data communications).

17. *Id.* §§ 4.2.1-4.2.4.

18. *Id.* § 4.2 (“Specifications that are not on the standards track are labeled with one of three ‘off-track’ maturity levels: ‘Experimental’, ‘Informational’, or ‘Historic.’”).

19. *Id.*

20. *Id.*

21. Initially, there are numerous questions regarding the timing constraint, such as the relationship between the optimal entry of infrastructure standards vis-à-vis application standards; the impact of regulatory delay in issuing first entry licenses from the diffusion of innovation to their standardization; the preemptive, immediate and long-term effects of additional entry licenses of innovations on standardization; the optimal time it takes to detect “early maturity,” the optimal time to publicly announce a standard; and the distinction between simultaneous versus sequential entry. These questions invite further research, and a reassessment of IT standardization policy at large. See generally Jay P. Choi & Marcel Thum, *Market Structure and the Timing of Technology Adoption with Network Externalities*, 42 EURO. ECON. REV. 225-44 (1998); Joseph Farrell & Garth Saloner, *Installed Base and Compatibility: Innovation Product Reannouncement and Predation*, 76 AM. ECON. REV. 940-55 (1986); Pierre Regibeau & Katherine E. Rockett, *The Timing of Product Introduction and the Creation of Compatibility Decisions*, 14 INT’L J. INDUS. ORG. 801-23 (1996). In system environments such as cyberspace, probably the most important concern regards the influence of time of entry of infrastructure standards vis-à-vis application standardization, as discussed herein.

when to invest.²² This central recognition is even more evident whenever dynamic IT standards are involved.

More specifically, in IT standardization theory and practice, it is well accepted that the technical absorption of highly technologically finished products or routine product improvement processes²³ into common usage is a progressive phenomenon. That is, the absorption of one standard or a more complex system²⁴ of standards²⁵ is neither immediate nor inclusive. Rather it is progressive and takes on different investment inputs.²⁶ Seen through a production stage model, there are three consecutive independent technological phases, in the establishment of a standardized system technology.²⁷ All three are part of what is also known as a technological life-cycle—a metaphor that typically describes the evolution of standardized technology from its emergence to technological maturity and inevitable decline.²⁸ In essence, a technological lifecycle interacts with the standardization process through the life of each standard or group of standards.²⁹

22. See Robert McDonald & Daniel R. Siegel, *The Value of Waiting to Invest*, 101 Q.J. ECON. 707 (1986); Robert S. Pindyck, *Irreversibility, Uncertainty and Investment*, 29 J. ECON. LITERATURE 1110 (1991).

23. For the difference between close-ended “product standards” and descriptive “process standards,” see CARGILL, *supra* note 1, 59-61; LOUIS G. TORNATZKY & MITCHELL FLEISCHER, *THE PROCESSES OF TECHNOLOGICAL INNOVATION* 20-22 (1990); Mark Blaug, *A Survey of the Theory of Process-Innovations*, 30 *ECONOMICA* 13-22 (1963) (commenting on the more wider context of “product innovation” and “process innovation”). Both of these IT definitions were finally adopted in The National Cooperative Production Amendments of 1993, 15 U.S.C. § 4301(a)(6)(D) (2000) (regarding the definition of a term “joint venture”) and accordingly in 15 U.S.C. § 4305(a)(3) (2000) (regarding the disclosure of the purpose of the joint venture).

24. Conceptually, systems are goods that consist of perfectly complementary components such as hardware and software. See Jeffrey Church & Neil Gandal, *Network Effects, Software Provision, and Standardization*, 40 J. INDUS. ECON. 85-103 (1992); Michael L. Katz & Carl Shapiro, *Systems Competition and Network Effects*, 4 J. ECON. PERSP. 8, 93-115 (1994).

25. See CARGILL, *supra* note 1, 142 (“Usually quite a few standards will be invoked at once.”); Andy Sloane, *The Standards Process: Tools and Methods for Standards Tracking and Implementation*, 22 *COMPUTER STANDARDS & INTERFACE* 5-12 (2000).

26. See FARRELL & SOLANER, *supra* note 12, at 3 (for the technological standardization perspective); TORNATZKY & FLEISCHER, *supra* note 23, 27-30 (for the wider technological innovation perspective); see also sources cited *infra* note 28.

27. In keeping with its public standardization production process perspective, this study will focus on a production stage model. Cf. TORNATZKY & FLEISCHER, *supra* note 23, at 28-29 (discussing a private user-oriented stage model and the interplay between both models).

28. In IT standardization literature, a variety of overlapping phases of this process were have been suggested. See, e.g., Michal J. Bonino & Michael B. Spring, *Standards as Change Agents in the Information Technology Market*, 12 *COMPUTER STANDARDS &*

For information systems like the Internet, this processional rationale is constituted in the dominant industry “performance standard”³⁰ known as the ISO/IEC 15288³¹ “Life Cycle Management—System Life Cycle Processes.”³² ISO/IEC 15288 establishes a common framework for describing the lifecycle of systems and a complete set of well-defined processes and associated terminology.³³ ISO/IEC 15288 is meant to be applicable to any type of system or system of systems, such as the Internet.³⁴ In fact, the

INTERFACES 97-107 (1991) [Weiss & Spring, *Standards as Change Agents*]; Egyedi, *supra* note 3, at 49 (suggesting developing, inventing, and diffusing phases); Robin Mansell & Richard Hawkins, *Old Roads and New Signposts: Trade Policy Objectives in Telecommunication Standards*, in TELECOMMUNICATION, NEW SIGNPOSTS TO OLD ROADS 45 (Franca Klaver & Paul Slaa eds., 1992) (suggesting planning, negotiation, and implementation phases); Martin B.H. Weiss & Michael B. Spring, *Selected Intellectual Property Issues in Standardization*, in INFORMATION TECHNOLOGY STANDARDS AND STANDARDIZATION: A GLOBAL PERSPECTIVE 63 (Kai Jakobs ed., 1999) [hereinafter Weiss & Spring, *Selected IP Issues*]. For a variety of five-phased processional descriptions, see, for example, Yesha Y. Sivan, *Knowledge Age Standards: A Brief Introduction to Their Dimensions*, in INFORMATION TECHNOLOGY STANDARDS AND STANDARDIZATION: A GLOBAL PERSPECTIVE 1 (Kai Jakobs ed., 1999) (suggesting missing, emerging, existing, declining, and dying phases). See also BREYER, *supra* note 1, at 101-09 (for an economic analysis upholding an analogous five-phased standardization process).

29. See CARGILL, *supra* note 1, at 142; Bengt Hogberg et al., *Innovation in Industrial Policy Sectors—The Cases of Remote Sensing and Bioenergy*, in ORGANIZING INDUSTRIAL DEVELOPMENT 160-62 (Rolf Wolf ed., 1986); Sloane, *supra* note 25, at 6-7.

30. “Performance standards” state requirements in terms of results with criteria for verifying compliance but without stating the methods for achieving those results. See FEDERAL PARTICIPATION IN THE DEVELOPMENT AND USE OF VOLUNTARY CONSENSUS STANDARDS AND IN CONFORMITY ASSESSMENT ACTIVITIES § 3(2) (Off. Mgmt. & Budget, Circular A-119, Feb. 10, 1998) [hereinafter OMB Circular No. A-119] (memorandum for heads of executive departments and agencies) (discussing federal participation in the development and use of voluntary consensus standards and in conformity assessment activities), at <http://www.whitehouse.gov/omb/circulars/a119/a119.html> (last visited Dec. 18, 2003).

31. “ISO” is the International Organization for Standardization. “IEC” is the International Electrotechnical Committee. See *ISO Website*, at <http://www.iso.ch/iso/en/ISOOnline.frontpage> (last visited Dec. 1, 2003).

32. Consequentially, ISO 15288 is designed to be in complete harmonization with the ISO 12207, or “Software Life Cycle Processes” standard, and the ISO 15504, or “Software Process Assessment” standard. See *ISO/IEC 15288 Website*, at <http://www.15288.com/> (last visited Dec. 18, 2003); see also *Software Productivity Consortium Website*, ISO/IEC 15288, at <http://www.software.org/quagmire/descriptions/iso-iec15288.asp> (last visited Dec. 1, 2003).

33. See *ISO/IEC 19760—System Engineering—A Guide for the Application of ISO/IEC 15288 System Life Cycle Processes*, ISO/IEC 15288 Website (Oct. 2002), at http://www.15288.com/about_15288.htm (last visited Dec. 1, 2003).

34. *Id.*

ISO/IEC 15288 system lifecycle definition is the first ISO standard to deal aggregately with system lifecycle process in the context of hardware, software and human interfaces.³⁵ Without a precise understanding of the consecutive relations between infrastructure and applications, both types of activities can be wrongly assumed to be taking place simultaneously. This assumption ignores the existence of distinct system technologies, such as encryption or security standards, thereby failing to acknowledge their distinct standards and unique standardization activities. Alternatively, with a different policy approach, a sequential institutional choice should be made to account for the differences between infrastructure and applications.

The first lifecycle phase, the infrastructure development phase, begins with idea generation and ends with the development of a basic product or process. The introduction of a new technological innovation creates the content of the core infrastructure standards. In the infrastructure development phase, standards begin to be specified in their public form.³⁶ This phase is technologically-oriented, thus any consumer-oriented price-based competition of technological knowledge is usually both technically premature and economically inefficient.³⁷ Thus, very little price-based competition transpires in this phase. In this infrastructure development phase,

35. Some definitions of an “information system life cycle” potentially include both infrastructure and applications. See U.S. Dep’t of Agric., Natural Res. Conservation Serv., *Part 511—Information System Life Cycle Oversight and Evaluation* § 511.3 (“An information system contains the components developed and maintained during its lifecycle, including hardware and software configurations, application architecture, technical architecture, custom software components, application built packages, databases, project plans, test plans, implementation and migration plans, system documentation, and user documentation.”), at http://policy.nrcs.usda.gov/scripts/lpsiis.dll/M/M_270_511.htm (last visited Dec. 1, 2003); see also Hogberg et al., *supra* note 29, at 160-62 (“[T]he life cycle concept can be applied to different levels of analysis, i.e., to a technology, a product, or a branch of industry.”).

36. See JOHN E.S. PARKER, *THE ECONOMICS OF INNOVATION* tbl. 4.5 (1978). The dynamic correlation between the creation of innovations and standards is subject to a substantive change with the beginning of standards’ commercialization. See discussion *infra* Part IV.

37. See, e.g., J. Gregory Sidak, *An Antitrust Rule for Software Integration*, 18 YALE J. ON REG. 1, 27 (2001) (“[i]n such a market, consumer knowledge is accumulating, and product demand is still immature and unstable”); Michael Whinston, *Tying, Foreclosure, and Exclusion*, 80 AM. ECON. REV. 837, 855-56 (1990) (suggesting that lack of technologic maturity leads to unclear ex ante results and to ambiguous future welfare effects); Carl Shapiro, *Antitrust in Network Industries*, *Address to the American Law Institute and the American Bar Association on Antitrust/Intellectual Property Claims in High Technology Markets* (Jan. 25, 1996), at <http://www.usdoj.gov/atr/public/speeches/shapiro.htm> (last visited Dec. 1, 2003).

radical innovations develop entirely new core standards (“infrastructure standards”). As a whole, these core standards are oriented toward increased technological performance, rather than an immediate market need. As a general rule, core or infrastructure standards usually establish a necessary technical platform for future standardized applications as well as for other complementary standardized technologies.³⁸

The second lifecycle phase is the modification phase. In the modification phase, the technology accepted in the infrastructure development phase undergoes rapid innovative changes. Because of consumer demand for enhanced complementary application products, competitors begin to challenge each other, leading to extensive bargaining over modifications.³⁹ These bargaining result in modifications to the previous technical policies that are necessary for the emergence of new markets for core Internet telecommunications services. The modification phase also serves to enhance the creation of commercial products or processes that are later finalized in the last phase of the technological lifecycle, the implementation phase. Consequently, the technological modification phase does not generate any new standards of its own because it is a short and intermediate stage.

The implementation phase is the last of three standardization stages. Here, technology matures due to technological and market limitations. In cyberspace, this leads to the final creation of new markets for both core Internet telecommunications *services* and conforming standardized *products* and *applications*.⁴⁰ The creation of new markets typically propagates the unavoidable final decline of that same technology as it attempts to adhere to its own standards, followed by the emergence of new, competitive product generations that are standards-compliant to the extent necessary for consumer adoption.⁴¹ Seen as an evolutionary production process, each product generation would then suggest a more advanced and compatible technology with existing standards and the fast growing consumer demands. In this third and final phase, complementary application standards became largely oriented towards specific market needs of improving exist-

38. This technological incentive is particularly effective when it creates entirely new markets for standards. The difficulty in maintaining this incentive after the development phase will be discussed in Part V.

39. See BREYER, *supra* note 1, at 107-08, 177-78 (describing such bargains in various industries).

40. See Sidak, *supra* note 37, at 27-28 (stating that “[i]n such a market, products are well-defined, both by the consumer demand that they satisfy and by the production technology through which firms supply them”).

41. See OMB Circular No. A-119, *supra* note 30.

ing technology and further standardizing newer applications and conforming standards (“application standards”).

C. Cyberspace’s Architectural Edifice: The Criterion of “Space”

In addition to institutionalizing the aspect of time, the technological lifecycle ultimately incorporates the creation of substantively different types of standards in each sequential phase. Infrastructure standards are created in the infrastructure development phase. This is followed by application and complementary technology standards, which are then replaced by new and improved standards.⁴² Accordingly, as a function of both technological and commercial needs, these *categories* of standards emerge as part of the overall technological standardization endeavor and, as such, serve as an additional independent regulatory constraint.

1. *The Normative Framework*

After evaluating the function of time, a rational policy planner should evaluate the long-established criterion of space, namely, types of standards by location.⁴³ In cyberspace, space is a function of architectural layer. With technological standards, the need for this criterion may be less obvious. On one hand, an overly strict definition of standards by type according to architectural layer may lead to technological rigidity, inhibiting potential standard setters from developing additional or cheaper alternative standards.⁴⁴ On the other hand, identification of standards by type may potentially lower administrative costs and thus diminish both technological and economic uncertainty.⁴⁵

Theoreticians and the FCC have tried to view setting standards according to architectural layer as a way to lower costs and diminish uncertainty. As early as 1966, the FCC commenced an inquiry (“First Computer Inquiry”) to study the interrelationship of computers and telecommunications technologies, and the use of computer-based services over telephone lines.⁴⁶ This inquiry partly influenced the Telecommunications Act of 1996 and certain of these orders are still in effect. The FCC Commission further observed that “the growing convergence of computers and communications has given rise to a number of regulatory and policy questions

42. See TORNATZKY & FLEISCHER, *supra* note 23, at 165.

43. See OGUS, *supra* note 1, at 165-68.

44. *Id.* at 167.

45. See *id.*; Shapiro, *supra* note 39, § III.B.2.

46. See *In re Regulatory & Policy Problems Presented by the Interdependence of Computer & Communication Serv. & Facilities*, 7 F.C.C.2d 11 ¶ 2 (1966) (addressing the apparent convergence between telecommunications and computing).

within the purview of the Communications Act.”⁴⁷ These policy concerns are as true today as they were more than three decades ago.⁴⁸

In its next inquiry (“Second Computer Inquiry”), the Commission reaffirmed its essential regulatory approach to the provision of computer data services, but improved its analysis.⁴⁹ By distinguishing regulated telecommunications services from unregulated data services, the Commission created the categories of *basic* services⁵⁰ (renamed “telecommunications services”) and *enhanced* services⁵¹ (renamed “information services”).⁵² The Commission also elaborated on the extent of structural separation required between the incumbent telephone provider and its enhanced services affiliate.⁵³

Understanding that the future would bring the convergence and interdependence of computers and communications, the Commission was aware of the difficulty of separating telecommunications services into two

47. *Id.* ¶ 1.

48. *Id.*

49. This distinction was then formally adopted. F.C.C., 103D CONG., REPORT IN RE FEDERAL-STATE JOINT BOARD ON UNIVERSAL SERVICE ¶ 33 (Apr. 10, 1998) [hereinafter *Second Computer Inquiry*] (stating that telecommunications services and information services are “separate, non-overlapping categories, so that information services do not constitute ‘telecommunications’ within the meaning of the 1996 Act”), at http://www.fcc.gov/Bureaus/Common_Carrier/Reports/fcc98067.pdf (last visited Dec. 1, 2003).

50. The Commission defined the term “basic” service, which referred to traditional common carrier telecommunications offerings as “the offering of transmission capacity for the movement of information.” See *In re Comsat Study*, 77 F.C.C.2d 584 ¶ 93 (1980).

51. The Commission defined “enhanced services” as “services, offered over common carrier transmission facilities used in interstate communications, which employ computer processing applications that act on the format, content, code, protocol, or similar aspects of the subscriber’s transmitted information; provide the subscriber additional, different or restructured information; or involve subscriber interaction with stored information” See 47 C.F.R. § 64.702(a) (2001).

52. The Telecommunications Act broadly defines an “information service”, specifically excluding “telecommunications services”, as

the offering of a capability for generating, acquiring, storing, transforming, processing, retrieving, utilizing, or making available information via telecommunications, and includes electronic publishing, but does not include any use of any such capability for the management, control, or operation of a telecommunications system or the management of a telecommunications service.

47 U.S.C. § 153 (2000).

53. See *Comsat Study*, 77 F.C.C.2d ¶ 190-266. For a wider discussion about the three Computer Inquiries’ genealogy, see BARBARA ESBIN, INTERNET OVER CABLE: DEFINING THE FUTURE IN TERMS OF THE PAST 25-26 (FCC OPP, Working Paper No. 30, 1998), at http://www.fcc.gov/Bureaus/OPP/working_papers/oppwp30.pdf (last visited Dec. 18, 2003).

discrete categories.⁵⁴ The Internet, in its contemporary form, did not exist at the time the FCC formed the basic/enhanced distinction and, as a result, the categories are subject to genuine interpretive ambiguities when applied to cyberspace. As the Commission acknowledged, “[p]lausible arguments can be tendered for drawing the line between the two services elsewhere. At the margin, some enhanced services are not dramatically dissimilar from basic services or dramatically different from communications as defined in the First Computer Inquiry.”⁵⁵ For example, appreciative data processing, computer memory or storage, and some advanced switching techniques typically identified as enhanced services can be components of a basic service if they are used solely to facilitate the movement of information.⁵⁶ A second interpretive ambiguity for cyberspace is that the FCC classification has focused entirely on the basic and enhanced services distinction from a telecommunications perspective with no adequate consideration of cable-based Internet services.⁵⁷ Instead, the Commission observed that because enhanced service was not explicitly referred to in the Telecommunications Act, there was no more reason to confront it with a specific traditional regulatory mechanism than there was for cable television’s formal elements of common carriage and broadcast television (then unregulated under the Act).⁵⁸

A third interpretive problem is that, even while upholding the Commission’s policy of regulating only the common carrier *basic* transmission service,⁵⁹ there is still little or no guidance about the question of how the Commission *should* act towards Internet-based services.⁶⁰ For standardiza-

54. See OXMAN, *supra* note 4, at 7; WERBACH, *supra* note 4, at 46.

55. See *Comsat Study*, 77 F.C.C.2d ¶ 434 (the Commission avoided re-drawing the line at this margin due to its concerns that such action could potentially subject the issue to constant adjudication over the status of individual services offerings. However, as such distinctions are crucial for any institutional standardization analysis, such adjudication is necessary, and will be upheld in this chapter hereinafter); see also ESBIN, *supra* note 53.

56. *Comsat Study*, 77 F.C.C.2d ¶ 419-20.

57. Traditionally, cable service has been regulated as an integrated video, information content, and conduit service under Title VI of the Telecommunications Act. See ESBIN, *supra* note 53, at 3, 83-90 (for an of integrated cable-based analysis of Internet access services); see also *Second Computer Inquiry*, *supra* note 49, at 98-167 n.140 (reserving consideration of the “regulatory classification of Internet services provided over cable television facilities”).

58. See *Comsat Study*, 77 F.C.C.2d ¶ 430.

59. See J. SCOTT MARCUS, THE POTENTIAL RELEVANCE TO THE UNITED STATES OF THE EUROPEAN UNION’S NEWLY ADOPTED REGULATORY FRAMEWORK FOR TELECOMMUNICATIONS 6 (FCC OPP, Working Paper No. 32, 2000), at http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-224213A2.doc (last visited Dec. 18, 2003).

60. See, e.g., WERBACH, *supra* note 4, at 29.

tion purposes, there was thus no adequate distinction between the question of *regulating* existing enhanced services and standardizing their *production*. This lack of direction implicitly leaves both processes to the competitive “hands off” premise of Title II of the Telecommunications Act.⁶¹ To add to this confusion, the Commission noted that it retains regulative control over enhanced services under the ancillary jurisdiction of Title I, on the grounds that the enhanced services under consideration “constitute the electronic transmission of writing, signs, signals, pictures, etc., over the interstate telecommunications network.”⁶²

Notwithstanding these interpretive difficulties, these two dependent categories, basic and enhanced, give rise to two different types of standards in present day cyberspace.⁶³ The first type is standardized physical telephony infrastructure standards and layers, which are supported by the basic packet switching, transporting, addressing, and routing protocols that establish most of cyberspace’s infrastructure standards. Originally, in the *Report In re Federal-State Joint Board on Universal Service*⁶⁴ and subsequent orders, the Commission came to address the implications of packet-switching technologies for this regulatory framework.⁶⁵ It was admitted that the use of packet switching and error control techniques⁶⁶ “that facilitate the economical, reliable movement of [such] information [do] not alter the nature of the basic service.”⁶⁷

The second type of standards, computer software standards, establish most of cyberspace’s subsequent application standards and are to be found in the user-oriented “application” layer of cyberspace’s architecture. The application layer performs many of the Internet’s information processing functions, such as authentication, email storage and retrieval, Web page hosting, and domain name server lookups. These functions involve substantial computer processing and interaction with customer-supplied information, and therefore fall squarely within the definition of enhanced

61. See OXMAN, *supra* note 4, at 13, 21.

62. *Comsat Study*, 77 F.C.C.2d ¶ 432.

63. Cf. STUART MINOR BENJAMIN ET AL., TELECOMMUNICATIONS LAW AND POLICY, 867 (2001) (implying this analysis).

64. See *supra* note 49.

65. See WERBACH, *supra* note 4, at 31 (discussing the applicability of the distinction between basic and enhanced Internet telecommunications services).

66. *Comsat Study*, 77 F.C.C.2d ¶ 420.

67. For example, in subsequent decisions the Commission determined that packet-switched networks following X.25 protocols, and frame relay service offerings provide a basic transport service. See *Application of AT&T*, 94 F.C.C.2d 48, 55-57 (1983); see also WERBACH, *supra* note 4, at 32.

services.⁶⁸ Specific enhanced services include protocol processing and electronic publishing, as well as the provision of access to data networks such as commercial online services and the Internet.⁶⁹

Ultimately, both sets of standards and their associated services should pave the way for a more comprehensive and accurate multi-layered and standards-based understanding than the current one. The higher the layer and production phase, the more specific the purposes of their standards become. The key insight is that although data is transmitted vertically, each different layered protocol is programmed horizontally to become compatible with neighboring standards and should thus be subject to different production costs. Later, with clear categories of standards, it will be possible to finalize the Internet's institutional regulative policy as a whole.⁷⁰

2. *Infrastructure Standards*

Cyberspace, and more distinctively, the Internet are commonly defined by a unified architectural backbone structure and core protocols known together as the formal TCP/IP networking reference model of the Internet's four-layer architecture.⁷¹ TCP/IP's architectural hegemony began as early as 1983, when the TCP/IP networking reference model largely overtook the formal, seven layered, Open System Interconnection ("OSI") networking reference model. The TCP/IP model refers to a large number of protocols located in four main layers of standardized architecture that formally standardize both the Internet's core infrastructure and complementary application standards by their relevant layer location. Practically, although the TCP/IP reference model is not always strictly followed with respect to keeping related functions together in a well-defined layer, most telecommunications and information systems products make an attempt to describe their functions in relation to this model. At least informally, the Internet's architecture can be described methodologically in four layers. The first layer is the "network access" layer. In its basic physical foundation lays the North American architecture, in connection with Europe through the EBONE communication supporter, which consists of three autonomous managed levels of hierarchical architecture. Each of

68. WERBACH, *supra* note 4, at 32-33.

69. *Id.*

70. *See* discussion *infra* Parts III-V.

71. A "backbone" is a telecommunications line that links one or more locations together. *See* CRAIG HUNT, TCP/IP: NETWORK ADMINISTRATION 1-22 (2d ed. 1997); Hans-Wernes Braun & Kimberly C. Claffy, *Network Analysis for a Public Internet*, in PUBLIC ACCESS TO THE INTERNET 353-56 (Brian Kahin & James Keller eds., 1995).

these levels as a whole represent a function performed when data is transferred between cooperating applications across the network, in the following hierarchical order: National Backbones (for example, NSFNET) which are attached among themselves, through (inter-) national network interconnections facilities, and down the line also to mid-level networks (for example, Midnet), which are attached to local service providers (for example, UCSD).⁷²

The latter backbone level supports five additional infrastructure levels, beginning with different IP networks (for example, 132.204.m.n), which are attached to IP sub-networks (for example, 132.204.51.n), which are attached to IP Host/end-systems (for example, 132.204.51.6)⁷³, which are attached to end-users, which are attached to networked applications (for example, X-Windows).⁷⁴

Ultimately, the different layers also differ in their standardized system-oriented specifications. The three backbone levels, consisting of the “network access” layer, contain very few protocols, as they handle relatively uncomplicated network interactions. This layer defines the network hardware and device drivers.⁷⁵

72. The FCC refers to cyberspace’s lower physical telephony infrastructure through four physical categories and formally suggested an analogous definition: backbone, middle mile, last mile and last 100 feet. *See In re Inquiry Concerning the Deployment of Advanced Telecommunications*, 15 F.C.C.R. 20913 (2000).

73. A “host” is a computer directly connected to the Internet. Still, it does not accurately reflect the actual number of Internet users, and is usually shared by groups of users and is thus smaller than them in size.

74. *See Braun & Claffy, supra note 71; cf. HUNT, supra note 71, at 1-22.* Hunt provides a functional-based description of the TCP/IP networking reference model in four levels: 1) “network access” (referring to the three backbone network levels); 2) “Internet” (referring to the IP Networks and sub-networks levels, similar to the Open System Interconnection (“OSI”) “network” layer); 3) “host-to-host transport” (referring to the IP Host/end-systems, similar to the fourth layer in the OSI model); and finally 4) the “application” level (referring to the end-users and the networked applications levels, similar to layers 5-7 in the OSI model). *HUNT, supra note 71, at 1-22.* For an analogous four-layer description, containing the Link, Network, Transport and Application layers, see *ISO/OSI Network Model Description*, at http://www.uwsg.iu.edu/usail/network/nfs/network_layers.html (last visited Dec. 1, 2002). In order to simplify, I suggest to refer the hegemonic TCP/IP labels as layers, as suggested also by Hunt. Previously, the Federal Networking Council (“FNC”) has unanimously upheld the existence of a layered architecture, as part of the Internet’s definition. *See Brian Carpenter, Architectural Principles of the Internet*, RFC 1958, at <http://www.faqs.org/rfcs/rfc1958.html> (last visited Dec. 18, 2003).

75. As far as standardization matters, these “non-consumer-oriented” levels include technologies for network management (for example, the Simple Network Management Protocol (“SNMP”)), the Ethernet standard for local area networks, the Frame Relay

More infrastructure protocols exist at the next two levels—the IP networks and sub-network layers, or the “Internet” layer, where the IP protocol prevails. As a general matter, they are both responsible for routed data interchange between hosts and across network links, through addressing and fragmentation of packets (for example, the Internet Protocol Version 6 standard (“IPv6”).⁷⁶ The third and final infrastructure layer is the “transport” layer, referring to the IP Host/end-systems level. The function of this layer is to make the Internet more useful and accessible to its users. This layer includes standardized telecommunications and transport protocols (for example, the Transmission Control Protocol (“TCP”)),⁷⁷ and more general standards for providing sufficient quality of service.⁷⁸

The various infrastructure layers are breeding grounds for strict network or system-oriented (rather than user-oriented) physical telephony infrastructure standards, which together establish the Internet’s core infrastructure. Potentially seen as “basic” services according to the meaning established in the Computer Inquiries, they should also refer to traditional common carrier telecommunications offerings as “the offering of transmission capacity for the movement of information[.]”⁷⁹ As such, the three

packet-switched data communication service, and standardized management interfaces for various classes of equipment (for example, the Fiber Distributed Data Interface (“FDDI”) for the 100 Mbps local area networks, and other operations issues). See DOUGLAS E. COMER, *INTERWORKING WITH TCP/IP (PRINCIPLES, PROTOCOLS, AND ARCHITECTURE)* 32-33 (3d ed. 1995); WILDER, *supra* note 14 (describing the TCP/IP protocol suite).

76. WILDER, *supra* note 14, at 155-64.

77. See *id.* at 165-84 (discussing the transmission control protocol); Charles Hornig, Symbolics Cambridge Research Center, RFC 894—Standard for the transmission of IP Datagrams over Ethernet networks (Apr. 1984), at <http://www.faqs.org/rfcs/rfc894.html> (last visited Dec. 2, 2003); Information Sciences Institute, University of Southern California, RFC 791—Internet Protocol (Sept. 1981), at <http://www.faqs.org/rfcs/rfc791.html> (last visited Dec. 2, 2003); Information Sciences Institute, University of Southern California, RFC 793—Transmission Control Protocol (Sept. 1981), at <http://www.faqs.org/rfcs/rfc793.html> (last visited Dec. 2, 2003); P. Mockapetris, Information Sciences Institute, University of Southern California, RFC 882—Domain names: Concepts and facilities (Nov. 1983) [hereinafter RFC 882], at <http://www.faqs.org/rfcs/rfc882.html> (last visited Dec. 2, 2003).

78. This layer is generally dominated by two different protocols, YCP and UDP, which are responsible for negotiating the flow of data between any two network hosts. See COMER, *supra* note 75, at 179-90; WILDER, *supra* note 14, at 163-64. See also Clinton & Gore, *A Framework*, *supra* note 14, § 9 (providing a U.S. governmental perspective); *Active IETF Working Group Website* (a list of working groups in these areas), at <http://ietf.org/html.charters/wg-dir.html> (last visited Dec. 18, 2003); IETF Internal Division Website (providing an industry perspective on these layers), at <http://www.ietf.org> (last visited Dec. 18, 2003).

79. *In re Comsat Study*, 77 F.C.C.2d 584 ¶ 93 (1980).

infrastructure layers are subject to separate standardization costs and a different institutional choice.

3. *Application Standards*

The “application” layer sits on top of the infrastructure layers and refers to the end-user and the networked application levels of the Internet. Application layer standards serve two major functions. First, as TCP/IP-compatible standards, they facilitate the operation of the infrastructure standards. The most familiar among these application standards are the standardized network application protocols (for example, HTTP, FTP and SMTP, NFS, DNS, arp, rlogin, talk, and ntp).⁸⁰ Functionally, application standards interact between clients (our personal computers) and servers (the relevant data storage units).⁸¹

Not all standardized technologies are directly related to earlier infrastructure technology and are not meant to facilitate operation. Thus, a second function of these standards, only indirectly related to infrastructure protocols, is as a source of innovation.⁸²

Overall, the emergence of new markets and sub-markets that the application layer has given rise to a variety of Internet software products such as browsers, encryption modules, contract infrastructures, electronic payment systems and security equipment (for example, the IP Security (“IP-Sec”) protocols⁸³ and XML Digital Signatures⁸⁴), X-Windows, Java, and

80. *See generally* WILDER, *supra* note 14, at 293-356.

81. Of central importance to this interaction are Hyper Text Transfer Protocol (“HTTP”), used to publish (and read) hypertext documents across the web; File Transfer Protocol (“FTP”), used to transfer files; and Simple Mail Transport Protocol (“SMTP”), used to transfer electronic mail. *See* COMER, *supra* note 75, at 299-304, 315-23, 344-47.

82. One example for this technological development can be examined through security technology, as it is embedded in both infrastructure and application standards. Thus, in its infrastructure lies the Secure Socket Layer (“SSL”) infrastructure protocol, which resides above the TCP layer and below application layer protocols. Examples include HTTP, LDAP, and IMAP. SSL is designed to make use of TCP to provide a reliable and end-to-end service. Because SSL is a channel security mechanism running on TCP, it can secure any protocol that can be carried by TCP. Thus, it was ideally suited for following secured applications such as SMTP, Telnet, and FTP, through independent commercial innovative software, such as NetStructure 7115 e-Commerce Accelerator (an SSL offload device) and SeeBeyond e*Xchange eBusiness Integration Suite (an interface for configuring security parameters). *See, e.g.*, Zeus Technology, *SSL Theory and Practice* (June 2000), at <http://itpapers.com/abstract.aspx?kw=ssl+&docid=25317> (last visited Dec. 2, 2003).

83. *Id.*

84. *See* W3C, *Extensible Markup Language (XML) 1.0* (2d ed. Oct. 10, 2000), at <http://www.w3.org/TR/2000/REC-xml-20001006> (last visited Dec. 2, 2003).

e-mail systems.⁸⁵ In essence, application standards are distinct from infrastructure standards in both specification and function. Potentially seen as “enhanced services” as defined by the Computer Inquiries these user-oriented standards seem to comply with the Commission’s definition of “enhanced services” as “services, offered over common carrier transmission facilities used in interstate communications, which employ computer processing applications that act on the format, content, code, protocol, or similar aspects of the subscriber’s transmitted information.” Like user-oriented standards, application standards also embed unique standardization costs and ultimately require a different institutional choice.

D. Standard-Setting Institutions: The Criterion of “Institutional Identity”

The final standard-setting criterion derives from the competing regulative regimes. Along with the criteria of time and space, this third criterion suggests a three-dimensional matrix of institutional choices for policy makers.

Institutions that regulate technological standards differ according to several variables. First, they differ by the degree of regulative formality, which indicates the degree of legality and influence of its legitimate elective legislators. Turning to comparative institutional theory, March and Olsen address this question, while questioning both the primacy of such action and its outcomes.⁸⁶ Accordingly, the core task of political institutions is to confirm the legitimacy of standardization choices by ensuring that relevant people are involved and that an appropriate control structure exists.⁸⁷ These same elements are also arguably evident in standardization ideology as they define the role of formal standards bodies as guardians of the process.⁸⁸ In all standardization bodies, specifications are merely a starting point. Functionally, standard setters also perform the role of checking the level of acceptance of their standards in relevant markets, through the breadth of their recognition and, ultimately, through actual exercise by users.

Measuring that acceptance typically requires determining whether the specified protocol or service provides significant benefit to the cyber

85. This same application layer lies exclusively (and, arguably, only for the time being) at the core of the ongoing debate regarding application software as a regulative constraint. See LESSIG, *supra* note 8, at 101-02; SHAPIRO & VARIAN, *supra* note 8; Gibbons, *supra* note 8; Johnson & Post, *supra* note 8; Wu, *supra* note 8.

86. MARCH & OLSEN, *supra* note 13, at 50-52.

87. *Id.*

88. See TORNATZKY & FLEISCHER, *supra* note 23, at 41-42.

community and market.⁸⁹ The second variable is the degree and type of monopolistic power over the right to supply, vis-à-vis the regulated status of all suppliers in a given market.⁹⁰ Third is the scope of the institution's legal status, evaluated in terms of its binding force and efficiency of enforcing a given standard.⁹¹ As a general matter, such performance is difficult to monitor for several reasons. For a start, there is no official legal means in the United States for ascertaining whether or not a standardization organization is a formal or informal standard developer.⁹² Moreover, this differentiation is empirically blurred since there have been no experiments on competitive self-regulation, no real market for the *control* of standard setters, and no easy option for principals (politicians and citizens) to dismiss ineffectual officials or market standard setters.⁹³

In one of the seminal articles on standardization, Farrell and Saloner identify five distinctive types of regimes, based on a function of their standardization endeavors.⁹⁴ The first, and least influential in cyberspace, is *de facto* standardization activity generated by internal decisions of

89. See generally MARCH & OLSEN, *supra* note 13.

90. See *id.*; TORNATZKY & FLEISCHER, *supra* note 23, at 41.

91. *Id.*

92. In the United States, alleged standards developers may request to be formally accredited by the American National Standards Institute ("ANSI"). See <http://www.ansi.org> (last visited Dec. 18, 2003). As part of a non-binding voluntary initiative, ANSI requires written procedures with strict requirements for openness, balance, consensus and other due process. Internationally, the situation is not substantively better, as alleged standards developers may be created by declaration of treaty agreements between cooperative nations, such as the International Telecommunication Union ("ITU") or by national policies which recognize a standards organization, such as the International Organization for Standardization/International Electrotechnical Commission Joint Technical Committee ("ISO/IEC JTC 1"). See *Joint Technical Committee Website*, at <http://www.jtc1.org/> (last visited Dec. 2, 2003). Drafts of "international standards" adopted by the joint technical committee are circulated to national bodies for votes. Publication as an "international standard" requires approval by at least 75% of the national bodies casting a vote. See also CARGILL, *supra* note 1, at 200-04, 269-70 (discussing the activities of ISO/IEC JTC 1).

93. See, e.g., FARRELL & SOLANER, *supra* note 12, at 5.

94. *Id.* at 2-5; see also OGUS, *supra* note 1, at 108-09; Mark A. Lemley, *Antitrust and the Internet Standardization Problem*, 28 CONN. L. REV. 1041 (1996) [hereinafter Lemley, *Antitrust*] (focusing on the government and industry players and *de facto* standards); Mark A. Lemley, *Standardizing Government Standard-Setting Policy for Electronic Commerce*, 14 BERKELEY TECH. L.J. 745, 747 (1999) [hereinafter Lemley, *Standard-Setting*]; Bob Toth, *Putting the U.S. standardization system into perspective*, 4 STANDARDVIEW 169, 169-78 (1999) (reviewing the presiding organizations inside the United States).

autonomous firms with a single vendor.⁹⁵ Closely related are standards emerging from a mutual agreement among several manufacturers, whether formal or informal, binding and/or voluntary—aimed at consolidating potentially different interests among the parties to the agreement.⁹⁶ Third, are ad hoc standards, consisting of market *de facto* and industry *gray* standards that are subsequently absorbed by consumers through accidental,⁹⁷ or strategic choice and then later adopted by the entire relevant market.⁹⁸ Standards designed by *de facto* standard setters are typically driven by self-interested profit maximizing implementers and tend to be both proprietary and closed sourced. Consequently, they are especially interesting from a legal perspective because they tend to raise a variety of issues concerning the proper scope of antitrust and intellectual property law in influencing market outcomes particularly in the implementation phase.

A fourth type of standardization institution is governmental, usually implemented through delegated regulatory agencies or organizations. Initially, governmental standards, in their meaning in section 12(d)(4) of the

95. Cf. FARRELL & SOLANER, *supra* note 12, at 2 (concluding, however, that a final analysis for both this regime and beyond is neither mutually exclusive nor independent, as it can be reflected in alternative *de facto* standard setters).

96. *Id.* (adding that this type of standard setter faces all the problems of autonomous firms, and more); see also SHAPIRO & VARIAN, *supra* note 8, at 236-37.

97. See Brian W. Arthur, *Competing Technologies and Lock-in by Historical Small Events: The Dynamics of Allocation Under Increasing Returns*, 99 ECON. J. 116-31 (1989) (modeling technological choice under increasing returns by consumers as a random process); Paul A. David, *Clio and the Economics of QWERTY*, 2 AM. ECON. REV. 75, 332-37 (explaining the development of the typewriter keyboard from this approach); Paul A. David, *Some New Standards for Economics of Standardization in the Information Age*, in ECONOMIC POLICY AND TECHNOLOGICAL PERFORMANCE 206 (Dasgupta & Stoneman eds., 1987) (confirming Arthur's study on historical lock-in under increasing returns).

98. See Michael L. Katz & Carl Shapiro, *Network Externalities, Competition and Compatibility*, 75 AM. ECON. REV. 424 (1985); Michael L. Katz & Carl Shapiro, *Product Compatibility Choice in a Market with Technological Progress*, 38 OXFORD ECON. PAPERS 145-65 (1986) (formalizing Arthur's notion into a theoretical model describing consumer's choice of technological products as a strategic consideration). These standards are possible after such a standard achieves a predominant market share over potential competitors. These three formats are commonly known as informal standards, and share the characteristic of being produced by non-legally binding autonomous market forces (*de facto*) or even particular groups (for example, non-profit organizations) or consortia (*gray*) standardizing autonomously. In some cases, companies may operate outside the established standard-setting organizations in consortia to form standards. See Roy Rada, *Consensus versus Speed*, in INFORMATION TECHNOLOGY STANDARDS AND STANDARDIZATION: A GLOBAL PERSPECTIVE 19, 30-31 (Kai Jakobs ed., 2000) (providing preliminary description of consortia); see also Egyedi, *supra* note 3, at 54-55 (describing grey standardization institutions).

National Technology Transfer and Advancement Act of 1995,⁹⁹ are developed by the government for its own uses.¹⁰⁰ As such, they are produced and made to serve, and thus penetrate, an entire industry. Finally, standards are introduced by intra-national, and more so, international standardization organizations operating jointly, through special agencies.¹⁰¹

These last two regimes, and the parties creating them, are commonly known as formal (“*de jure*”) standards and standard setters. They are processed by traditional political standard development organizations, such as the International Organization for Standardization (“ISO”), Internet Engineering Task Force (“IETF”), scientific or professional societies, trade associations or other types of industrial standard organizations, which may operate in accordance with official formal regulative bodies. Standards designed by *de jure* standard setters are typically driven by technologically-oriented implementers and tend to be both non-proprietary and open sourced. Historically, in other fields of media, standardization used to be the prefecture of international industry standardization organizations for example, the ITU, ISO, and the International Electrotechnical Commission (“IEC”). With time, standardization activity expanded to *de facto* and governmental standardization bodies.¹⁰² As in the telecommunications field, cyberspace is also subjected to all three, albeit not necessarily in that evolutionary order, as will be described in the following parts.

To conclude, the absorption of high technology, through one or more standards, into ordinary usage is sequential. Generally, there are three consecutive independent technological phases in the establishment of a standardized technology, beginning with the emergence of the technology in the development phase and ending in a full technological life-cycle. The creation of such standards varies in space, with the partition of cyberspace’s standardized architecture into four layers. Broadly, one can draw a clear distinction between the first three network or system-oriented layers,

99. National Technology Transfer and Advancement Act of 1995, Pub. L. No. 104-113, § 12(d)(4), 110 Stat. 775 (1996) (codified as amended at 15 U.S.C. 272(b) (2000)).

100. *Id.*

101. See FARRELL & SOLANER, *supra* note 12, at 4; see also FORMAL METHODS IN STANDARDS: A REPORT FROM THE BRITISH COMPUTER SOCIETY (BCS) WORKING GROUP 7-8 (C.L.N. Ruggles ed., 1990) (describing the various early European and American-based International standardization organizations); KELLEY LEE, GLOBAL TELECOMMUNICATIONS REGULATION: A POLITICAL ECONOMY PERSPECTIVE 121-22 (1995) (describing the telecommunications field main precedents: the International Telecommunications User Group (“INTUG”), Intelsat or Eutelsat).

102. Petri Mähönen, *The Standardization Process in IT—Too Slow or Too Fast?*, in INFORMATION TECHNOLOGY STANDARDS AND STANDARDIZATION: A GLOBAL PERSPECTIVE 35, 37 (Kai Jakobs ed., 2000).

and the fourth user-oriented layer. These two groups of layers consist of substantively different technological standards, and therefore create different regulative costs and concerns. First, the three lower layers imbed infrastructure standards that are intended to maintain expensive lines,¹⁰³ data networking equipment,¹⁰⁴ Internet backbone telecommunications, and cable services.¹⁰⁵ These types of standards are most common to the early development phase. The application layer involves substantively different types of standards. These application standards establish most of cyberspace's computer software products. These standards are created primarily in the implementation phase of IT standardization. In essence, although data is transmitted vertically through each layer, each functional layer is programmed horizontally and should be subject to different production costs. Ultimately, different production costs may then lead to different institutional choices, as will be discussed later.

In addition to the criteria of time and space, different types of standardization regimes can be characterized by the identity of the institution that creates them. Standards designed by *de facto* standard setters are typically driven by self-interested profit maximizing implementers and tend to be both proprietary and closed. In contrast, standards designed by *de jure* standard setters are typically driven by technologically-oriented implementers and tend to be both non-proprietary and open sourced. Conceptually different phases of technological development lead to different technological standards. With that recognition, a rational policy-maker should be able to predict efficiently the degree of regulative compliance of each such standardized technology with typical price-based Kaldor-Hicksian efficiency,¹⁰⁶ and thereby establish a comprehensive regulative policy. Therefore, any optimal institutional choice will have to consider the following three technological phases, and their distinctive standardization activities.

103. But lots of cheap routers manage a limited number of these expensive lines. *See, e.g.*, Jeffrey K. MacKie-Mason & Hal R. Varian, *Pricing the Internet*, in PUBLIC ACCESS TO THE INTERNET 269, 273 (Brian Kahin & James Keller eds., 1995) (estimating that this conclusion is reflected in the rapid decline from expensive routers to expensive transmission links).

104. Both of which are oriented at clients (modem, ISDN, cable) and servers (routers, modem pools, and call aggregators).

105. These services include, for example, hybrid fiber-coax to cable and digital cable for higher-speed PC Internet connections.

106. According to the Kaldor-Hicksian notion of efficiency, whenever at least one person is better off and at least one is worse off, gainers should compensate losers. *See* ROBERT COOTER & THOMAS ULEN, LAW AND ECONOMICS 43-44 (3d ed. 2000).

III. THE INFRASTRUCTURE DEVELOPMENT PHASE: A POLITICAL INSTITUTIONAL INEVITABILITY

The technological infrastructure development phase introduces new innovations, beginning with idea generation to the development of a basic product or process. The infrastructure development phase results in the creation of core infrastructure standards. As a whole, these infrastructure standards are oriented toward increased technological performance rather than an immediate market need. As a general rule in cyberspace, infrastructure standards establish a necessary technical platform for future standardized applications as well as for other complementary standardized technologies. This part evaluates cyberspace's infrastructure endeavor of the early 1990s in light of this technological standardization phase with a further view toward infrastructure standardization for future technological generations. Official U.S. policy regarding cyberspace standard-setting never dealt appropriately with such infrastructure standards or the difference between setting infrastructure and application standards.¹⁰⁷ It has even largely ignored the existence of the technological professional constraint on standardization itself. Instead, the United States has largely focused its standardization policy on application standard-setting. The following will provide separate analysis for the Internet architecture establishment, the separate development of infrastructure standards and the backbone transit services of cyberspace.

A. Infrastructure Telecommunications Services: The Collective Choice Analysis

Between 1986 and 1992, Internet infrastructure governance evolved from an institutional hybrid of Department of Defense ("DoD") and government-supported academic research institutions, with the DoD controlling the funding and governance of the latter.¹⁰⁸ The Internet started as a U.S. government military project. It was later transferred to the National Science Foundation ("NSF"), which operated it for another decade.¹⁰⁹ The

107. In fact, the U.S. government applied its policy broadly, and made no distinction between infrastructure standards, such as "high-speed network technologies" and application standards, such as "electronic copyright management systems," in their meaning. See Clinton & Gore, *A Framework*, *supra* note 14, § 9.

108. See Anthony M. Rutkowski, *Factors Shaping Internet Self-Governance*, in COORDINATING THE INTERNET 92, 93-94 (Brian Kahin & James H. Keller eds., 1997).

109. See Steve Bickerstaff, *Shackles on the Giant: How the Federal Government Created Microsoft, Personal Computers, and the Internet*, 78 TEXAS L. REV. 1 (1999) (describing the major developments in these years while concluding that today's Internet market exists as a result of that government intervention rather than early free market competition).

NSF officially started the process that led to the comprehensive standardization of the current cyberspace technological generation. Following the experience of other telecommunications industries, the NSF realized the need for a unified wide-area infrastructure to support the early NSFNeT.¹¹⁰ That realization immediately led to the adoption of two main *regulatory standardization* policies. The first was the 1985 decision to unify the TCP/IP set of protocols as a worldwide mandatory infrastructure for the NSFNeT program at large.¹¹¹ Over time, that decision marginalized the other competitive computer network protocols and heralded the dominance of the addressing IP system.¹¹² Second, the NSF decided to support DARPA's existing Internet organizational infrastructure hierarchy, assembled by the previous Internet Activities Board ("IAB") in the decision known as the "Request for Internet Gateways" RFC 985.¹¹³

Having successfully dominated over DARPA and NSF's interoperable infrastructures, the federal government was ready to enlist the telecommunications and cable industries to further build out the Internet's infrastructure and promote private investment.¹¹⁴ The ultimate purpose of that initia-

110. See CARGILL, *supra* note 1, at 64-65 (discussing coordinating mandatory policies, as regulatory standardization).

111. Cf. Toru Komatsu, *The history and motivation of the Internet* (Feb. 29, 2000) [hereinafter Komatsu, *History*] (describing the transition of the ARPANET host protocol from Network Control Protocol ("NCP") to TCP/IP as of Jan. 1, 1983, which required all hosts to convert simultaneously or be left having to communicate via ad-hoc mechanisms, and stating that "[t]his transition was carefully planned within the community over several years before it actually took place and went surprisingly smoothly"), at <http://www.sju.edu/~tk098681/csc5235/history.htm> (last visited Dec. 3, 2003). But it resulted in a distribution of buttons saying "I survived the TCP/IP transition." Richard W. Wiggins, *The Internet Marks Its 20th Anniversary* (Dec. 30, 2002), at <http://www.infotoday.com/newsbreaks/nb021230-1.htm> (last visited Dec. 3, 2003).

112. See Komatsu, *supra* note 110.

113. Network Technical Advisory Group, *RFC 985—Requirements for Internet Gateways* (IETF Network Working Group, May 1986), at <http://www.faqs.org/rfcs/rfc985.html>; V. Cerf, *RFC 1160 – Internet Activities Board* (IETF Network Working Group, May 1990) (defining the areas of responsibility of the Internet Activities Board), at <http://www.armware.dk/RFC/rfc/rfc1160.html> (last visited Dec. 3, 2003); see also WILDER, *supra* note 14, at 6, 366 app. A.

114. See, e.g., Mähönen, *supra* note 102, at 42 (agreeing that a "firm standard is needed for ensuring interoperability . . . before large investments are made"); see also U.S. DEP'T OF COM., *THE NATIONAL INFORMATION INFRASTRUCTURE: AGENDA FOR ACTION* (Dec. 21, 1993) [hereinafter *AGENDA FOR ACTION*] (discussing U.S. policy concerns), at <http://metalab.unc.edu/nii/NII-Agenda-for-Action.html> (last visited Dec. 18, 2003). The U.S. administration's National Information Infrastructure ("NII") initiative intends to improve access to essential services, while encouraging private sector investment in the net's development through tax and regulatory policies that encourage innovation and promote long-term investment. See discussion *infra* Part V.1.

tive was to develop an independent network, away from direct federal funding or governance.

Following the successful effort to unify infrastructure standards, the NSF went on to gradually privatize the Internet backbone providers.¹¹⁵ It began with the local and regional networks and then expanded upon successful preservation of the compatible standardized infrastructure. Its efforts eventually turned into the complete privatization of the Internet's transit infrastructure, through the final privatization of the national backbone pipes and the creation of simple economies of scale in the provision of a standardized transit service.¹¹⁶ One of the main indications of this privatization policy initiative came with the December 23, 1992 announcement by the NSF regarding the cessation of funding to the ANS T3 Internet backbone.¹¹⁷ This announcement catalyzed the transition from the government-funded engineering-oriented Internet to the commercial Internet run by private providers of telecommunications services and business.

In the beginning of this process, these efforts were aimed solely at the local and regional networks. Through its "Acceptable Use Policy" ("AUP") the NSF prohibited backbone usage for purposes "not in support of research and education," intending to encourage commercial network traffic at the local and regional levels, while temporarily denying access to the national scale transportation facilities.¹¹⁸ It was only in 1995 that the NSF finally decides to reinstitute funding of the NSFNET backbone, thus allowing the full recovery of costs through competition on buying national-scale Internet telecommunications services by private networks.¹¹⁹

With the decision to privatize these services the government also committed part of its effort to setting up and facilitating the interconnection of different future private networks to exchange traffic via Network Access Points ("NAPs").¹²⁰ This decision was meant to further encourage diverse infrastructure equipment providers, such as data networking and

115. See Jay P. Kesan & Rajiv C. Shah, *Fool Us Once Shame on You—Fool Us Twice Shame on Us: What We Can Learn From the Privatizations of the Internet Backbone Network and the Domain Name System*, 79 WASH. U. L.Q. 89, 111-17 & n.6 (2001) (describing the privatization process of the NSFNET and providing a bibliographic list of short histories that discuss the privatization of the net).

116. *Id.* at 117.

117. See MacKie-Mason & Varian, *supra* note 103, at 274.

118. Barry M. Leiner et al., *A Brief History of the Internet, version 3.31* (Internet Society) [hereinafter Leiner, *Brief History*], at <http://www.isoc.org/internet/history/brief.shtml> (last modified Dec. 4, 2000).

119. *Id.*

120. Constance K. Robinson, *Network Effects in Telecommunications Mergers MCI Worldcom Merger: Protecting the future of the Internet*, 1192 PLI/CORP 517, 530 (2000).

telecommunications equipment providers and cable operators, to participate. Any company that wished to exchange traffic at a NAP, did so after negotiating the terms and conditions of that inter-exchange through bilateral agreements.¹²¹

With the enormous growth of Internet traffic followed by a new problem of congestion of the NAPs emerged the first signs of a free ride problem. That congestion came in the form of a bottleneck of connection speed across the Internet, followed by the typical telecommunications' loss of data and quality.¹²² Acting as interest groups, the larger networks responded to this problem by investing in private dedicated connection points over which they had a monopoly (provided earlier by the government). That initiative was responsible for partly restoring faster and more accurate connections.¹²³

As individual networks grew, more actions were needed to combat free ride failure in the nationwide backbone providers' extensive network investment. As a result, larger network providers began to create policies to restrict potential peer-to-peer arrangements between small and regional ISPs that had not invested in growing their networks.¹²⁴ These individual-to-individual peering agreements were then replaced by seller-customer transit agreements, where the national backbones charged the small networks or ISPs "transit fees" for carrying and terminating their traffic, while further maintaining their inter-connectivity services monopoly.¹²⁵

The ultimate establishment of public goods in users' access rights in the Internet backbone telecommunications services, and the minimization of the free ride failure were eventually achieved as usage became essentially free to all authorized end users. While most users were connected to a backbone through a "pipe" for which a fixed access fee is charged, the user's organization nearly always started covering the access fee as over-

121. *Id.* at 531.

122. *Id.*

123. *Id.* at 532.

124. MICHAEL KENDE, THE DIGITAL HANDSHAKE: CONNECTING INTERNET BACKBONES 4-9 (FCC OPP, Working Paper No. 32, 2000) (describing the analogy between the Internet transit and peering arrangements and the bill-and-keep and sender-keeps-all arrangements in traditional telephony interconnection arrangements, respectively), available at http://www.fcc.gov/Bureaus/OPP/working_papers/oppwp32.pdf (last visited Dec. 18, 2003).

125. BENJAMIN ET AL., *supra* note 63; KENDE, *supra* note 124. In contrast, telecommunications carriers interconnecting with one another for the exchange of telecommunications traffic, still do it pursuant to Title II of the Telecommunications Act of 1996, which obliges all carriers to interconnect pursuant to reasonable terms and conditions. See 47 U.S.C. § 251 (2000).

head without any direct charge to the user. As a result, most users of the NSFNET backbone do not pay any pipeline fee to the service provider, but instead pay in order to get connected to their regional or mid-level network, and then are granted a connection to the NSFNET.¹²⁶

In essence, this provision of infrastructure backbone transit services as public goods was achieved based on a governmental policy aimed at minimizing only the negative affect of uncoordinated multiple infrastructure standardization efforts. Conceptually, this policy can be identified as a *positive default approach* to standardization. This is, in contrast to a *negative default approach*, in which not even negative externalities such as a technological arms race in infrastructure transit services are handled ex ante. For the infrastructure development phase, when a unified infrastructure is presumed to be efficient, this latter extreme approach does not seem justified, even by those who identify cyberspace as a state-free sphere complying only with alternative regimes under full consumer sovereignty.¹²⁷

In time, the creation of these first suspected monopolies by primary interest groups resulted in their vast control over access to the Internet's telecommunications services.¹²⁸ Because such vast monopoly powers

126. MacKie-Mason & Varian, *supra* note 103, at 269 n.1 (estimating that this conclusion is reflected in the rapid decline from expensive routers to expensive transmission links). For two additional secondary reasons why most Internet end-users do not pay usage charges: (1) residential local service tends to be flat-rated, and ISPs have located their POPs to maximize the number of subscribers who can reach them with a local call; and (2) ISPs typically connect to Local Exchange Carriers ("LECs") networks through business lines that have no usage charges for receiving calls. See WERBACH, *supra* note 4, at 50.

127. This regulative doctrine is analogous to Cyberanarchism referred to by Neil Netanel. See Neil Netanel, *Cyberspace Self-Governance: A Skeptical View from Liberal Democratic Theory*, 88 CALIF. L. REV. 395, 433-35 (2000); see also WERBACH, *supra* note 4, at 29.

128. See Kesan & Shah, *supra* note 115, at 148-59 (arguing that large backbone providers unfairly benefit from the lack of an interconnection policy, which penalizes smaller networks unjustly and limits new competitors). This situation typically happens when the largest backbone providers obtain the ability to dominate other backbones by threatening disconnection, degrading interconnection services, or charging monopoly prices for interconnection. See also Bickerstaff, *supra* note 109, at 101 ("[I]f left to market forces, many of the subsidies that are critical to public use of the Internet would disappear as [Independent Local Exchange Companies] would drive interconnection charges toward cost through usage-sensitive [sic] rates.").

would become almost inevitable, they would later demand counter-antitrust measures.¹²⁹

Conceptually, this governmental use of higher-quality telecommunications interest group participation conforms remarkably well to Olson's widely accepted collective choice theory.¹³⁰ According to this theory, smaller, influential, self-interested groups will be better able to organize collectively and combine their resources and minimize the free ride problem.¹³¹ This holds true for early high technology research and development ("R&D") standardization activities.¹³² To achieve access to broad dispersed public goods in cyberspace, namely users' access rights to telecommunications services in the Internet backbone networks, the standardization efforts were dominated by small influential groups of individuals or firms that were seeking to benefit themselves by establishing the Internet's infrastructure.¹³³ In practice, as in theory, then, these influential groups were small groups in the form of pre-organized homogenous telecommunications transit providers, with high per-capita stakes, rather than large heterogeneous potential customer groups with relatively small per capita stakes in the form of dispersed customers.¹³⁴ However, unlike in the earlier telephone industry, the researchers working on the ARPA's Internet working program in the early 1970s did not follow the model of the telephone system with its tightly integrated multiple networks into a centrally managed system, instead building a loose confederation of inde-

129. Lemley, *Antitrust*, *supra* note 94, at 1042; see I. DE SOLA POOL, *TECHNOLOGIES OF FREEDOM* 245 (1983); Ronald Hirshhorn, *Regulating Quality in Product Markets*, in *THE REGULATION OF QUALITY* 77 (Donald N. Dewees ed., 1983).

130. See MANCUR OLSON, *THE LOGIC OF COLLECTIVE ACTION* (2d ed. 1971). Mancur Olson's analysis of formalized models of institutional theory is one that has profoundly shaped the debate, and will be used henceforth as a theoretical point of reference.

131. See generally DANIEL FARBER & PHILIP P. FRICKEY, *LAW AND PUBLIC CHOICE* 23 (1991); Jonathan R. Macey, *Promoting Public-Regarding Legislation Through Statutory Interpretation: An Interest Group Model*, 86 *COLUM. L. REV.* 223, 231 (1986).

132. See, e.g., William J. Baumol & Janusz A. Ordover, *Antitrust: Source of Dynamic and Static Inefficiencies?*, in *ANTITRUST, INNOVATION, AND COMPETITIVENESS* 89 (Thomas M. Jorde & David J. Teece eds., 1992).

133. See OLSON, *supra* note 130, at 132-34, 162.

134. See Roger G. Noll, *Economic Perspectives on the Politics of Regulation*, in *HANDBOOK OF INDUSTRIAL ORGANIZATION* 1265 (Richard Schmalensee & Robert Willig eds., 1989); see also Hanseth & Monterio, *supra* note 3, at 173-74 (adding that the uniquely complex technological "invisibility" of the Internet infrastructure serves, in fact, as a negative incentive for participation by private users); Mähönen, *supra* note 102, at 40.

pendently managed networks.¹³⁵ Even so, ARPA assumed that such large-scale producers would be able to organize better than diffused consumers with lesser ability to reach a stable coordinated consensus needed for such new industry.¹³⁶

As a result, potential future consumers, such as the emerging Internet user community, would have found counter-organizing over-costly and substantively inefficient.¹³⁷ Accordingly, in the infrastructure development phase, legislators are generally biased in favor of these narrow economic interests at the expense of the general public.¹³⁸ Therefore, at this stage the government should have allowed and acknowledged an additional degree of monopoly power.¹³⁹

Thus, not surprisingly, the interest groups enrolled in this political action came from the leading telecommunications and cable service carriers such as AT&T, WorldCom, and Sprint.¹⁴⁰ The government's goal that only previously efficient, organized groups should be part of this preliminary commercial effort was realized.¹⁴¹ The rationale behind that policy was imported from previous experiences. The government determined that organized groups that had already paid the fixed costs of formation would have an advantage over their counterparts.¹⁴² This was especially important due to the anticipated susceptibility of these organized groups to the free ride failure. This free ride failure occurs when efficient interest groups

135. See Sharon Eisner Gillett & Mitchell Kapor, *The Self-Governing Internet: Coordination by Design*, in COORDINATING THE INTERNET 3-7 (Brian Kahin & James H. Keller eds., 1997).

136. For the theoretical perspective, see Olson, *supra* note 130, at 29; Noll, *supra* note 134, at 1264-65; see also Netanel, *supra* note 127, at 437, n.161.

137. See generally FARBER & FRICKEY, *supra* note 131, at 19 (discussing the theoretical framework); Netanel, *supra* note 127, at 438 (discussing cyberspace's application); Susan Rose-Ackerman, *Progressive Law and Economics—and the New Administrative Law*, 98 YALE L.J. 341 (1988) (same); Cass R. Sunstein, *Interest Groups in American Public Law*, 38 STAN. L. REV. 29 (1985) (same).

138. See KAY SCHLOZMAN & JOHN TIERNEY, ORGANIZED INTERESTS AND AMERICAN DEMOCRACY 317 (1986); Frank B. Cross, *Public Choice and the Judiciary*, 50 HASTINGS L.J. 355, 357 (1999); Frank H. Easterbrook, *Forward: The Court and the Economic System*, 98 HARV. L. REV. 4, 15 (1985).

139. POOL, *supra* note 129, at 245 (suggesting that “[u]nder these circumstances the best solution seemed to be to define a monopoly’s turf narrowly”).

140. Kesan & Shah, *supra* note 115, at 144 (suggesting that although thousands of companies provide Internet connectivity, they are all dependent upon MCI WorldCom, Genuity (formerly GTE), AT&T, Sprint, and Cable & Wireless); see also Neil Weinberg, *Backbone Bullies*, FORBES, Jun. 12, 2000, at 236.

141. See, e.g., Cooney et al., *Internet surge strains already shaky structure*, NETWORK WORLD, Apr. 3, 1995, at 1.

142. OLSON, *supra* note 130, at 28; cf. Noll, *supra* note 134, at 1265.

achieve an efficient outcome but have no ability to limit the benefits to its members, who originally had to confront additional, sometimes fatal, costs of formation.¹⁴³

In essence, in the infrastructure development phase of the Internet, the government took a dual regulative attitude towards what were two main purposes. The first was to coordinate a unified core standardized infrastructure, namely both the worldwide domination of the TCP/IP set of protocols, and the adoption of the hierarchical multi-layered architecture. Institutionally, research institutions were very cleverly effected in a way that allowed governmental governance of all the key aspects of TCP/IP and hierarchical multi-layered architecture adoption, while at the same time encouraging open and highly active research and academic collaboration.¹⁴⁴ Only with the success of this early regulatory goal did the government continue on to its second, substantively different goal. This second goal was the transfer of power over the Internet infrastructure to new market agents, namely, the telecommunications and cable industries. To achieve the second purpose, the government relegated itself to an indirect monitoring role, gradually pushing these interest groups to seize control over increasingly larger backbone levels. By giving away much of its power over the market for basic backbone transit *services*, the government deliberately avoided giving up hegemony over cyberspace's standardized *infrastructure*. It retained worldwide domination of the TCP/IP protocols,

143. See Noll, *supra* note 134 (suggesting that costs of formation are central to the ability to dominate the market); see also Julie E. Cohen, *Lochner in Cyberspace: The New Economic Orthodoxy of "Rights Management"*, 97 MICH. L. REV. 462, 522 (1998). A key insight of Olson that goes even further than Noll's is that such organization costs (as initial fixed costs) required to realize collective benefits, are an increasing function of the size of the group. See OLSON, *supra* note 130, at 53-57. To date, there is still no sufficient empirical evidence to apply Olson's claim to cyberspace. Furthermore, the costs of organizing groups that span international boundaries (i.e., cyberspace) are said to be decreasing in some respects due to communication technology and the Internet's radical ability to lower the significant costs associated with reproducing information and transmitting it over distance. Thus, even when Olson's application may have not yet been upheld adequately, the predictions of his theory can still serve as an early indicator in measuring cyberspace's institutional costs of formation. In fact, practice shows that earlier than 1995, private companies were merely overseen by the NSF in developing early TCP/IP hardware and software products, for example, private companies' participation in the advisory panel of scientists and engineers from academia and industry, including those involved in Computer Scientists Net ("CSNET"), to assist the Division of Networking and Communications Research and Infrastructure ("NCRI") staff in designing the first TCP/IP products. See Livinginternet.com, *NSFNET*, at http://livinginternet.com/ii_nsfnet.htm (last visited Dec. 3, 2003).

144. See Rutkowski, *supra* note 108, at 94.

and control over the root zone.¹⁴⁵ Whether this infrastructure production policy was institutionally justified will be the central question answered by the cost-benefit analysis of infrastructure standardization discussed below.

B. Infrastructure Standardization: The Cost-Benefit Analysis

Reflecting the specific costs of infrastructure standard-setting, the optimally efficient institutional choice of the infrastructure development phase was justifiably an *ex ante* governmental standard initiative propagated by federal agencies and governmentally-monitored research institutions.¹⁴⁶ Left to market agents alone, a pre-commoditized infrastructure environment bears too high a cost (and thus too high a risk of failure) to produce optimal core standards.¹⁴⁷ For the infrastructure development phase, a rationale planner should avoid any independent *de facto* or gray standardization coordination initiatives. Instead, coordination of infrastructure standardization should be entrusted to government standardization hegemony. As suggested earlier, this policy rule should be regarded as a positive default approach for minimizing the main negative externalities deriving from multi-standardization of core infrastructure. To achieve this goal, it is critical to identify the most efficient standard-setting institution.

In an early seminal work on regulation, Stephen Breyer suggests the basic theoretical framework for assessing the efficiency of such standard-setting activity.¹⁴⁸ As Breyer suggests, a rational standard setter, operating with broad statutory authority, would first define the adverse effect he seeks to control.¹⁴⁹ The standard setter would then use a preliminary rough cost-benefit analysis to identify the specific part of the general problem he intends to minimize, while obtaining the greatest improvement at the lowest cost. On balance, the standardization plan must be set at the level at which the total benefits exceed the total costs by the greatest comparative

145. See A. Michael Froomkin, *Wrong Turn in Cyberspace: Using ICANN to Route Around the APA and the Constitution*, 50 DUKE L.J. 17, 169 (2000) (suggesting that possessing control over the root zone the U.S. government had, in fact, only quasi-privatized the control over root server services); Mähönen, *supra* note 102, at 36-37.

146. See CARGILL, *supra* note 1, at 176; Hanseth & Monteiro, *supra* note 3, at 175.

147. As a general rule, that would also mean that in a few marginal cases, uniquely weak *de jure* infrastructure standards might be overcome by stronger self-regulated standards. One such example was the development of the TCP/IP infrastructure standardized set of protocols, which, jointly, quickly gained dominance over the OSI layer architecture. However, due to the coordinated hegemonic role taken by the U.S. government in that case, only a unified TCP/IP was adopted. See, e.g., Mähönen, *supra* note 102, at 43.

148. See BREYER, *supra* note 1, at 96-119.

149. *Id.* at 98.

amount and at which the marginal standardization benefits are equal to the marginal standardization costs.

Next, the standard setter would obtain information and design a standard, which would most efficiently reduce the targeted adverse effects to an economically reasonable degree.¹⁵⁰ Later, he would operate to enforce that standard through developed means that ensure compliance.¹⁵¹ Finally, he would monitor enforcement, while evaluating the standard's effectiveness, through occasional revisions.¹⁵²

After identifying the fundamental lack of interoperability in constructing cyber infrastructure products as public goods, a rational planner would then continue to categorize the preliminary rough benefit-cost analysis for each specific parts of the general problem (for example, the need for transferring, addressing and routing standards), which he wishes to minimize through possible standardization. Yet, bearing in mind the technological maturity of the TCP/IP suite of protocols, a more general cost-benefit analysis could be made for its various protocols. In providing a reasonable view of the adverse effect at issue such as the lack of interoperability and the inevitable expansion of early fragmentation of the Internet's architecture—such evolution would implicitly include the specific ingredients of the adverse problem for example, the lack of addressing, routing and transferring of this former broader framework.

Scholars have initially subdivided the costs associated with technical standardization into three types of costs. These are administrative costs, compliance costs and indirect costs.¹⁵³ *Administrative costs* are generally created by the standard setters through several subsidiary activities. The first part of the administrative costs comes from the costs of quality standardization while overcoming numerous anticompetitive concerns of the infrastructure development phase. The second part is the cost of R&D activity for standard development. The third part is the information cost or the cost of conveying information about infrastructure standard formulation. *Compliance costs* arise from the need to coordinate standard interoperability (or compatibility) and espouse technological convergence with the three general constituents of media: mass media (broadcast, cable and satellite TV, and radio), telephony (wired and wireless), and interactive computer services. Accordingly, compliance costs also include transaction

150. *Id.*

151. *Id.*

152. *Id.*

153. See OGUS, *supra* note 1, at 155-56; cf. Lemley, *Antitrust*, *supra* note 94, at 1045-51; Marcus Maher, *An Analysis of Internet Standardization*, 3 VA. J.L. & TECH. 5, 12-23 (1998).

costs and the need to reduce inefficient variety. *Indirect costs* consist of productive inefficiencies where resources are not allocated most efficiently. All these concerns still require further economic modeling, and for the purpose of this study will be broadly outlined, as follows.

1. *Administrative Costs*

a) Quality Standardization Costs

In previous telecommunications infrastructure standardization, primacy was given to technical performance-based efficiency over price-based efficiency.¹⁵⁴ This imperative quality rationale is true for any IT standardization activity, but is even more acute for infrastructure standardization to the degree that it implies a centralized and unified vision of the future of that technology.¹⁵⁵ Infrastructure quality assurances should be coordinated through delegated formal standard setters and monitored research institutions. This policy rule is based on several grounds.

First and foremost, under the alternative paradigm of price competition based on varying consumer demand,¹⁵⁶ the motivation of *de facto* standard

154. See, e.g., THE REQUEST FOR COMMENTS REFERENCE GUIDE 2 (Joyce Reynolds & Jay Postel eds., IETF Network Working Group, Aug. 1987) (giving a historical account of that policy in designing core infrastructure standards between the years 1969-1987), <http://www.ietf.org/rfc/rfc1000.txt> (last visited Dec. 3, 2003). Such policy was also embedded in the ITU rules and policies, regarding mobile telephony, HDTV, Radio Frequency Spectrum, etc. From the restructuring of the ITU in 1947 through the commercialization of telecommunications information in the 1980s, a separate federal structure was created for the “technical” organs, including the Administrative Radio Conferences (“ARC”) and International Frequency Registration Board (“IFRB”). These institutions preside next to the more “political” Plenipotentiary Conference (“PC”) and Administrative Council (“AC”), which came to maintain that technological qualitative primacy. See LEE, *supra* note 101, at 131-33; cf. THOMAS M. JORDE & DAVID J. TEECE, PRODUCT MARKET DEFINITION IN THE CONTEXT OF INNOVATION: AN EXPLANATION 236-37 (1987) (suggesting that competition in early technological phases of production is insensitive to price changes, but very sensitive to product feature changes and emphasizing that limiting new technology can bring an innovative development to an end, harming quality).

155. See, e.g., CARGILL, *supra* note 1, at 64 (justifying the central “regulatory style,” “where quality programs mandate a single procedure in order to fabricate a product that must meet stringent quality standards”).

156. See, e.g., ALAN O. SYKES, PRODUCT STANDARDS FOR INTERNATIONALLY INTEGRATED GOODS MARKETS 38 (1995) (adding that, in general, not all potential customers will be willing to pay the same for particular attributes in quality due to differences in underlying tastes or to differences in wealth). In its recent report, the National Telecommunications and Information Administration (“NTIA”) of the Department of Commerce concluded that the use of computers and the Internet varies significantly according to income, race, and ethnicity, and that the gap is increasing for certain demographic categories. See U.S. DEP’T OF COMMERCE, FALLING THROUGH THE NET:

setters would be to reduce marginal development costs. Firms will ultimately end up reducing relatively high undervalued costs of socially desirable R&D quality benefits.¹⁵⁷ Moreover, the focus on quality assurances is not only intrinsic to the development process, but would also be guaranteed if a long-term unified infrastructure production were to be preserved. In conclusion, hegemony of *de facto* standard setters in infrastructure design may not only decrease quality assurance, but may accordingly create a technologically inefficient standard market. Such market may be overcostly to maintain as seen with the formal OSI networking reference model of the Internet's architecture, developed by representatives of major computer and telecommunication companies beginning in 1983, and later overtaken by the TCP/IP networking reference model and set of protocols.¹⁵⁸

DEFINING THE DIGITAL DIVIDE § IIE, at <http://www.ntia.doc.gov/ntiahome/fttn99/contents.html> (last visited Dec. 3, 2003).

157. This core argument, limited here to the context of IT standardization, is part of a larger one, upheld by neoclassical economists, according to which average market agents tend to under-supply R&D activity (and thus benefits), resulting in these players' lack of incentive to explore different technological paths compared to their relatively lower worth. Eventually, market agents, left alone, will not generate a sufficient degree of variety of high-quality standards. See KENNETH ARROW, ECONOMIC WELFARE AND ALLOCATION OF RESOURCES FOR INVENTIONS: THE RATE AND DIRECTION OF INVENTION ACTIVITY 609-25 (1962); Richard R. Nelson, *Recent Evolutionary Theorizing about Economic Change*, 33 J. ECON. LITERATURE 48, 48-90 (1995). Nevertheless, to date, there is still no theoretical or empirical consensus that reduced competition leads to less R&D and fewer new products. Economic theory is ambiguous on this point and only industry-specific findings, for example, Internet infrastructure standardization, are relevant. See OLIVER E. WILLIAMSON, MARKETS AND HIERARCHIES: ANALYSIS AND ANTITRUST IMPLICATIONS 180 (1975) (stating that "the technological potential to innovate differs greatly across industries"); see also *F.T.C. Hearings on Global and Innovation-Based Competition*, 102d Cong. (1995) (statement of Dennis A. Yao), <http://www.ftc.gov/opp/global/yaotest.htm> (last visited Dec. 1, 2002); JOHN LIPCZYNSKI & JOHN WILSON, INDUSTRIAL ORGANIZATION: AN ANALYSIS OF COMPETITIVE MARKETS 249 (2001); CARGILL, *supra* note 1, at 37; Thomas M. Jorde & David J. Teece, *Rule Of Reason Analysis Of Horizontal Arrangements: Agreements Designed To Advance Innovation And Commercialize Technology* (1998), <http://www.ftc.gov/opp/global/jorde2.htm#5> (last visited Dec. 3, 2003).

158. Probably one of the main reasons for the final collapse of OSI was the fact that its vendors undervalued R&D investment and declined to invest in test suites for both quality assurances and compatibility. Interview with Carl Cargill, Director of Corporate Standards, Sun Microsystems, Inc. (Mar. 2003) [hereinafter Cargill Interview]. It should be added that there were other, related, reasons for the collapse of the OSI network model such as greater cost, lower efficiency and lateness to market. See *infra* note 178.

Second, the desire to propagate technology widely should also lower potential price-based competition in developing a unified infrastructure.¹⁵⁹ Initially, infrastructure standards as opposed to serial product applications such as operating systems or browsers might not be effectively exposed to premature price competition where infrastructure (and thus market) boundaries are pre-defined.¹⁶⁰ Alternatively, any inducement for premature product interchangeability and cross elasticity of demand between its substitutes will not lead to the clear differentiation of such outer boundaries.¹⁶¹ Essentially, such early price competition comes at the expense of the quality assurance that is needed in the early infrastructure development phase.¹⁶² Thus, lack of technological maturity optimally requires mandated centralized coordination.¹⁶³

Third, absence of consumer qualitative judgment is most acute with technological infrastructure because potential customers tend to overvalue the exterior *interface* of a standard at the expense of interior assurances and solid development.¹⁶⁴ In part, this is what led the U.S. government to lower its independent coordination role of *de facto* standard setters in the infrastructure development phase while continuing to monitor *de facto* standard setters' development activity.

Fourth, the combination of Akerlof's "market of lemons" insight,¹⁶⁵ with the realization that architecture and protocol designing imbed future

159. See discussion *infra* Part III.B.2.b.

160. Andrew C. Hruska, Note, *A Broad Market Approach to Product Market Definition in Innovative Industries*, 102 YALE L.J. 305, 316 (1992). In essence, this use of the potential competition doctrine in the market for infrastructure standards might apply to application standards that do not now exist but will most likely exist in the future. See generally U.S. DEP'T OF JUSTICE & F.T.C., HORIZONTAL MERGER GUIDELINES (1992) (defining market boundaries based on price competition), <http://www.ftc.gov/bc/docs/horizmer.htm> (last modified Apr. 8, 1997).

161. See ANTITRUST, INNOVATION AND COMPETITIVENESS 9 (Thomas M. Jorde & David J. Teece eds., 1992); Hruska, *supra* note 160, at 310 (noting that because product development begins years before the commercialization that would allow enforcing agencies to test claims of market power empirically, market definitions remain elusive).

162. Michael J. Trebilcock, *Regulating Service Quality in Professional Markets*, in THE REGULATION OF QUALITY 86-87 (Donald N. Dewees ed., 1983) (emphasizing that consumer ignorance of serious risks embedded in the wrong choice may call for prescribed quality standards irrespective of the price or the access affects of such standards). Later, in the implementation phase, quality standardization will serve to benefit price competition. See discussion *infra* Part V.B.

163. See Gillett & Kapor, *supra* note 135, at 18.

164. See CARGILL, *supra* note 1, at 36-37 (providing this unique technological rationale).

165. George A. Akerlof, *The Market for "Lemons": Quality Uncertainty and the Market Mechanism*, 84 Q. J. ECON. 488 (1970) (arguing that because quality is usually

market preferences,¹⁶⁶ leads to the conclusion that high-quality infrastructure standardization aimed at facilitating future transferability of information will eventually lower the anticipated informing and advertising costs of application standards that would potentially be demanded by *de facto* standard setters in the implementation phase.¹⁶⁷ Acknowledging the role of core technology in establishing future consumer preferences, it may also minimize tendencies of producers and consumers to settle for lower quality, lower-price products in the subsequent implementation phase.¹⁶⁸ This argument is relevant to any standardization activity¹⁶⁹ but it is more acute for infrastructure standardization where future market boundaries and motivations of *de facto* standard-setting are imbedded in infrastructure definitions. In that regard, *de jure* standard-setting should facilitate future competition and lower these costs.¹⁷⁰

b) Development Costs

Central governments most efficiently bear the cost of funding R&D activity and monitoring its successful performance by setting standards for

costly to produce, poor-quality products can outcompete high-quality products and the market equilibrium may entail the future production of suboptimally low-quality products exclusively, thus increasing informing and advertising costs).

166. Niva Elkin-Koren, *Copyright in Cyberspace—Rights Without Laws?*, 73 CHI-KENT L. REV. 1155, 1186 (1998); Lawrence Lessig, *Constitution and Code*, 27 CUMB. L. REV. 1, 14-15 (1997) (observing that “the architecture is the product of private interests”—whether the relatively open Internet Engineering Task Force or the absolutely closed Microsoft Corporation—and “code is political . . . the architectures that are established in cyberspace have normative significance, and . . . choices can be made about the values that this architecture will embed”); Lawrence Lessig, *The Constitution of Code: Limitations on Choice-Based Critiques of Cyberspace Regulation*, 5 COMM.LAW CONSPICUOUS 181, 184 (1997); Joel R. Reidenberg, *Rules of the Road for Global Electronic Highways: Merging the Trade and Technical Paradigms*, 6 HARV. J.L. & TECH. 287, 301-04 (1993) [hereinafter Reidenberg, *Rules of the Road*].

167. *Cf.* Akerlof, *supra* note 165.

168. *Id.*

169. *See, e.g.*, CARGILL, *supra* note 1 (discussing potential competition in future standardized technologies); James J. Anton & Dennis A. Yao, *Standard-Setting Consortia, Antitrust, and High-Technology Industries*, 64 ANTITRUST L.J. 247, 264 (1995) (“Because standard-setting is forward looking, it may involve a competitively sensitive information exchange about future technologies and products.”); David A. Balto, *Standard Setting in a Network Economy* §§ I-II (Feb. 17, 2000) (“Competition may be affected in complementary goods, or even in next generation goods.”), http://www.ftc.gov/speeches/other/standardsetting.htm#N_40_ (last visited Dec. 3, 2003); Shapiro, *supra* note 37 (describing the components of the technological arms-race competition as including anticipating user needs and foreseeing and exploiting further hardware improvements).

170. Hruska, *supra* note 160, at 316.

research institutions. As with all other types of standards, infrastructure standards tend to freeze existing technology.¹⁷¹ However, infrastructure standards tend to freeze technology for longer periods. Thus, whenever *de facto* standard setters develop infrastructure they risk losing potential long-term market revenues. Ultimately, such risks are typically reflected in the lack of sufficient firm funding for development efforts.

Therefore, deciding who should develop infrastructure standards depends upon determining which institution can most efficiently fund the development of infrastructure standards. Because infrastructure standards are necessary long-run guides, mistakes made in setting them involve high risk of future harm.¹⁷² In any defective form, infrastructure standards would have to be reviewed and modified, if they are to become functional. This development and review process in itself would impose additional development cost on standard setters, thus increasing that same risk of error. To prevent these negative side effects, there is a need for central governments to play a central role in standard-setting activity and monitor its successful performance by autonomous standard setters such as research institutions.¹⁷³ On balance, because the small-to-medium enterprise (“SMEs”) of new applications in cyberspace is usually technically excused from a pre-commitment to build infrastructure, more entrepreneurs can finance and design more applications.

c) Information Costs

Formal industry standard setters can usually better internalize information costs and risks embodied in infrastructure standardization. Nevertheless, obtaining accurate information about prospective standards is likely the most costly activity for political institutions.¹⁷⁴ Moreover, conveying

171. See BREYER, *supra* note 1, at 105-06, 115-16; Mähönen, *supra* note 102, at 39.

172. See, e.g., Gillett & Kapor, *supra* note 135 (upholding the opposite conclusion for application standardization errors); see also LIPCZYNSKI & WILSON, *supra* note 157, at 225.

173. See AGENDA FOR ACTION, *supra* note 114 (presenting U.S. government support of this policy); RESEARCH CHALLENGES FOR THE NEXT GENERATION INTERNET 5 (Jean E. Smith & Fred W. Weingarten eds., Computing Research Ass’n, May 12-14, 1997), http://www.cra.org/Policy/NGI/research_chall.pdf (last visited Dec. 3, 2003) [hereinafter NEXT GENERATION INTERNET]; SHAPIRO & VARIAN, *supra* note 8, at 314.

174. BREYER, *supra* note 1, at 103, 109, 112 (describing information collection as the main cost in standard-setting); Eric J. Iversen, *Standardization and Intellectual Property Rights: Conflicts between innovation and diffusion in new telecommunications systems*, in INFORMATION TECHNOLOGY STANDARDIZATION AND STANDARDIZATION: A GLOBAL PERSPECTIVE 85-86 (Kai Jakobs ed., 1999) (suggesting that cost-structure of manufacturers of IT standards involves high investments risks, i.e., R&D costs, regularly in excess of 10% of turnover, with low variable costs); Tim Sloane et al., *Efficient Business-to-*

information about performance-based infrastructure standards is even harder, as such information is primarily quality-based rather than price-based.¹⁷⁵ Designers of infrastructure standards have varying preferences regarding quality and other value measurements (for example, price and service) of these standards.¹⁷⁶ Theoretically, central intervention would be unnecessary whenever *de facto* standard setters 1) have perfect information on the characteristics of infrastructure; 2) behave rationally in accordance with their preferences; and 3) have not generated externalities or monopolistic control on R&D information as a result of internal R&D decisions. Unfortunately, in cyberspace, this has been far from the case. Instead, *de facto* standard setters tended to keep their standards proprietary and closed sourced. Furthermore, even when sufficient information on the quality of infrastructure standards is made available, making decisions with the necessary degree of internalized social costs is difficult and expensive to coordinate among diverse, self-interested *de facto* standard setters.

Alternatively, dominant and competitively neutral formal industry standard setters, as repeat players (as opposed to governmental one-time players), typically hold a systematic advantage with regard to conveying information about infrastructure standard-setting activity.¹⁷⁷ Thus they should be supported by a government monitoring mechanism, as seen in the development of cyberspace.¹⁷⁸ Overall, even with lesser commercial know-how than *de facto* agents,¹⁷⁹ formal industry standard setters can

Business Relationships: How Analytics and XML Can Help, at <http://www.webtechniques.com/archives/2000/11/sloane/> (last visited Dec. 1, 2002); Michel B. Spring & Martin B.H. Weiss, *Financing the Standards Development Process*, in STANDARDS DEVELOPMENT AND INFORMATION INFRASTRUCTURE 289 (Brian Kahin ed., 1994)

175. Jorde & Teece, *supra* note 154, at 9 (stating that performance changes are more difficult to measure than price changes, because performance is multidimensional and may require various criteria of measurements).

176. *Id.*

177. CARGILL, *supra* note 1, at 37 (describing the industry consensus standardization process as a means of promoting innovative market neutrality); Sloane et al., *supra* note 174 (stating that “[t]o obtain the information required to make these decisions, it’s crucial for companies to evaluate B2B operations through vendor- and market-neutral analytics”).

178. Vendors have created consortia, for example, X/OPEN, hoping not to repeat the mistakes that damaged the implementations of OSI in the development phase and encouraged by the U.S. government. To develop standards, these consortia focused on collecting information and testing the implementations of the POSIX 1003.0 standard (ISO/IEC JTC 1 9945) (Test methods ISO/IEC JTC 1 13210, et al.) and other UNIX-like commands. See CARGILL, *supra* note 1, at 76, 221-22.

179. See Anton & Yao, *supra* note 169, at 254 (suggesting that exchange of commercial know-how may not require the exchange of much marketing information). Gillett &

usually better internalize information costs and risks embodied in infrastructure standardization.¹⁸⁰

2. Compliance Costs

Standardization coordinates technological developments efficiently.¹⁸¹ The ultimate technological goal of infrastructure standardization is to achieve complete integration while minimizing congestion among various elements of a technology, such as infrastructure and application standards. At the same time, the main economic difficulty to overcome is the motivation of *de facto* standard setters to bargain among themselves, to deal with their own potential incompatibility. Still, coordinating infrastructure is a complicated process, one that requires a great deal of interaction among the different phases of any comprehensive technology. Therefore, compliance costs arise out of the need to reduce inefficient variety, coordinate standard compatibility in a scarce environment, create technological convergence with other media, and accounting for existing transaction costs, as will be elaborated hereinafter.

a) Variety Reduction Costs

One of the important costs standardization derives from inefficient variability of products. Standardization policy attempts to prevent the adoption of standards viewed as “wrong” from a social point of view whenever such variety results in too many standards.¹⁸² Theoretically, in network environments, standard setters tend to prefer a small number of

Kapor further justify that early business environment in two ways. First, unlike commercial protocols developed recently, such as IBM's SNA, Digital's DECNET, and Xerox's XNS, infrastructure standard setters did not have the marketing resources of a large company behind them; second, nor was their design oriented toward any particular vendor's hardware. See Gillett & Kapor, *supra* note 135. Consequently, in most cases, such marketing will probably not involve personnel with much knowledge or authority in the marketing area. See Anton & Yao, *supra* note 169.

180. The funding mechanism will be described *infra* Part V.B.2.a.

181. See SUZANNE K. SCHMIDT & RAYMUND WERLE, COORDINATING TECHNOLOGY: STUDIES IN THE INTERNATIONAL STANDARDIZATION OF TELECOMMUNICATION (1998); Joseph Farrell & Garth Solaner, *Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation*, 76 AM. ECON. REV. 940, 942 (1986) [hereinafter Farrell & Solaner, *Installed Base and Compatibility*].

182. See Carl Shapiro & Hal R. Varian, *The Art of Standard Wars*, CAL. MGMT. REV. 41, 8-32 (1999) [hereinafter Shapiro & Varian, *Standards Wars*]; see also SHAPIRO & VARIAN, INFORMATION RULES, *supra* note 8; W. Brian Arthur, *Competing technologies: An Overview*, in TECHNICAL CHANGE AND ECONOMIC THEORY (Giovani Dosi et al. eds., 1988).

largely dispersed standards in order to exploit network benefits.¹⁸³ Thus, compared to standards markets without network effects, network markets tend to have lower variety, which is socially desirable. Therefore, when standard setters choose a standard, they take into account the benefits from expected network size. Such markets may be characterized by a “tipping” effect where only one standard survives.¹⁸⁴ All the same, there remains a tradeoff between a network’s size and its standard variety. In practice, there is no guarantee that standard setters’ optimal solution to this tradeoff implements the socially optimal outcome.¹⁸⁵ Standard setters may adopt the wrong standard, from a social point of view, resulting in too many or too few standards.¹⁸⁶ In such cases, corrective intervention might be socially desirable. More notably, such a scenario could lead to a gravely inefficient outcome in the infrastructure development phase. Instead, infrastructure standardization homogeneity, best achieved by central policy planning, is typically more efficient.¹⁸⁷ One key way to achieve a reduction in inefficient variety was manifested in the Domain Name System (“DNS”)—the naming hierarchy that tells connected computers where to find particular web sites.¹⁸⁸ The need for such homogeneity is a direct consequence of the scarcity constraint on the production of infrastructure standards to which the DNS is subject. In regulation analysis, any definition of a regulative realm as scarce has meaningful consequences. In cyberspace, such scarcity intensifies the need to reduce inefficient variety of infrastructure standards.

To begin with, uncertainty in the future growth of users and demand for IP addresses, combined with uncertain user adoption of technical changes makes it difficult to predict how many addresses would be re-

183. See Shapiro & Varian, *Standards Wars*, *supra* note 182; SHAPIRO & VARIAN, INFORMATION RULES, *supra* note 8.

184. SHAPIRO & VARIAN, INFORMATION RULES, *supra* note 8.

185. *Id.*

186. *Id.*

187. See Gillett & Kapor, *supra* note 135.

188. The importance of DNS stability is widely agreed by different institutions. See M. Kaat, *Overview of 1999 IAB Network Layer Workshop 10* (IETF Network Working Group Oct. 2000) (providing a formal industry view), at <ftp://ftp.ietf.org/rfc/rfc2956.txt> (last visited Dec. 3, 2003); see also Management of Internet Names and Addresses, 63 Fed. Reg. 31,741 (1998) [hereinafter White Paper], available at http://www.ntia.doc.gov/ntiahome/domainname/6_5_98dns.htm (last visited Dec. 3, 2003); Internationalized Domain Names Committee, *Discussion Paper on Non-ASCII Top-Level Domain Policy Issues* (ICANN Apr. 16, 2002) (providing ICANN’s view—warning that “the sudden introduction of a massive number of new TLDs would be a bad idea”), <http://www.icann.org/committees/idn/non-ascii-tld-paper.htm> (last visited Dec. 3, 2003).

quired to satisfy this demand.¹⁸⁹ As a practical matter, successful address allocation requires administrators with strong technical skills, not just political or legal expertise.¹⁹⁰ For example, administrators must have technical knowledge to preserve Internet transparency.¹⁹¹ Regrettably, much of this traditional end-to-end infrastructure transparency has been lost in the current Internet.¹⁹² This leads to complexity in application design and inhibits the deployment of new standard applications.¹⁹³ There are multiple causes for this loss of transparency, such as the deployment of network address translation devices and the use of private addresses, firewalls and application level gateways, proxies and caches.¹⁹⁴ These mechanisms increase fragmentation of the network layer, which causes operational irregularities for many applications on the Internet. Bearing in mind that some of the mechanisms have an intrinsic value (for example firewalls) any substantive regulative intervention on the matter may be controversial, may make end-to-end transparency unreasonably costly to restore.¹⁹⁵ Thus, preferably only competitively neutral formal standard setters, as suggested above, should coordinate such technical preferences.

Secondly, the current address allocation mechanism relies on voluntary trust between the competitive allocation authorities and consumers. Both compete based on their provision of an accurate and honest assessment of their requirements, so that appropriate amounts of address space are allocated.¹⁹⁶ At the same time, the requesters must trust the allocation authorities to be fair and impartial.¹⁹⁷ A user who wants to add a network to the Internet must first obtain a guaranteed-unique IP address. This IP address is allocated to a new network through Internet connectivity providers that typically have a block of addresses allocated to them by Regional Internet Registries (“RIRs”). From time to time, the Internet As-

189. See Gillett & Kapor, *supra* note 135, at 23.

190. See also David Randy Conrad, *Administrative Infrastructure for IP Address Allocation* (Aldea Communications, Feb. 2, 1996), at <http://aldea.com/cix/randy.html> (last visited Dec. 3, 2003).

191. The concept of a single universal logical addressing scheme and the mechanisms by which packets may flow from source to destination essentially unaltered.

192. Kaat, *supra* note 188, at 2 (“[S]pecifically the assumption that IPv4 addresses are globally unique or invariant is no longer true.”).

193. *Id.* at 3 (“It was however concluded that end to end transparency is desirable and is an important issue to pursue.”).

194. *Id.*

195. *Id.*

196. ICANN Address Supporting Organization, *Memorandum of Understanding* § 1 [hereinafter MOU], at <http://www.aso.icann.org/docs/aso-mou.html> (last visited Dec. 3, 2003).

197. See Conrad, *supra* note 190.

signed Numbers Authority (“IANA”) allocates large portions of address space to suballocate to providers as the RIRs see fit. Presently, four RIRs, operating in large geopolitical regions, are signatories to the voluntary Memorandum of Understanding executed on October 18, 1999.¹⁹⁸ This geopolitical tension is likely to remain into the future, as the number of RIRs is expected to remain relatively small.¹⁹⁹ Consequently, DNS allocation is already subject to two types of centralized policies. First, based on past growth rate estimates; registries feel compelled to allocate remaining Internet Protocol version 4 (“IPv4”) address space conservatively.²⁰⁰ Second, allocation authorities are attempting to simplify the Internet routing system by allocating larger continuous blocks of addresses.²⁰¹ Conversely, with the growing dominance of self-interested commercial networks on the Internet, the trust model for resource allocation is under severe multi-institutional pressure.²⁰²

In essence, IANA’s consensual decision-making mechanism may not clearly survive challenges to its legitimacy. As long as Internet service providers and IANA or Regional Internet Registries (“RIRs”) can cooperate, such multi-institutional pressure could be resisted.²⁰³ Should the historical cooperation between these two sets of organizations break down, the ISPs, again working in concert, could conceivably create a joint self-made RIR, allocating new addresses from the unused pool without regard to IANA/RIR policies.²⁰⁴ The third reason that DNS scarcity increases variety reduction costs is procedural. Due to the clear need for consensus on infrastructure adoption, any change, as valuable as it would be, would be too slow and thus expensive to achieve.²⁰⁵ On balance, as long as DNS is subject to the present pre-designed scarcity policy, that stability will be best achieved within the non-competitive technically oriented institutional framework of formal standardization. The latter would then serve to block

198. See MOU, *supra* note 196.

199. Thus, service areas will be of continental dimensions. See K. Hubbard et al., *Internet registry IP Allocation Guidelines* § 1, RFC 2050, (IETF Network Working Group Nov. 1996), at <http://www.ietf.org/rfc/rfc2050.txt> (last visited Dec. 3, 2003).

200. See Gillett & Kapor, *supra* note 135, at 24.

201. *Id.*

202. *Id.*; David Post, *Governing Cyberspace: Where is James Madison when we need him?* (ICANNWatch 1999) [hereinafter Post, *Governing Cyberspace*], at http://www.icannwatch.org/archive/governing_cyberspace.htm (“Any entity exercising control over the DNS will be subject to immense pressure to do more than mere ‘technical management.’”).

203. Conrad, *supra* note 190, at 23.

204. *Id.*

205. See Gillett & Kapor, *supra* note 135, at 23.

the temptation to create alternative unstable, inefficient infrastructure storage of Internet names and addresses.

b) Coordination Costs

In network environments, computer communications require complete and rigid compliance with basic interface specifications.²⁰⁶ More specifically, in cyberspace, unified specifications are critical to designing both cheaper core protocols of computer communications and achieving interoperability among its hierarchical layers.²⁰⁷ Coordination costs also arise whenever standards are revised. When a network industry settles on a single standard, it may be expensive to move toward a new, superior technology even when all users would theoretically be better off doing so. These switching costs fall into four categories.

To begin with, there is the cost of identifying, and then settling on, a specific infrastructure standard.²⁰⁸ Left to strategically dominant *de facto* agents, any non-binding attempt would face competing claims from self-interested commercial stakeholders, or worse, from different national or international standard-setting agencies. This emphasizes the need to assess future technological risks, such as fragmentation of the network layer and the ultimate risk of root splitting.²⁰⁹ Since standard-setting is often a “one-shot game” (thus, non-repetitive), in coordinating such consensus infrastructure standard setters might undesirably stick to an established inferior network technology (exhibiting excess inertia) or agents might undesirably switch to a new emerging one (exhibiting excess momentum).²¹⁰ With dif-

206. See FARRELL & SOLANER, *supra* note 12, at 1; Carl Shapiro & Robert Willig, *On the Antitrust Treatment of Production Joint Ventures*, 4 J. ECON. PERSP. 113, 114 (1990) (upholding the traditional rationale for joint ventures as a form of enterprise that can enable firms to attain economies of scale or scope that they could not otherwise achieve cheaply).

207. See, e.g., A. Lyman Chapin, *The Internet standards process*, RFC 1310, (IETF Network Working Group, Mar. 1992) (suggesting that in 1992, “[m]ost IETF members agree that the greatest benefit for all members of the Internet community results from cooperative development of technically superior protocols and services”), <http://rfc.sunsite.dk/rfc/rfc1310.html> (last visited Dec. 18, 2003); David R. Johnson & David G. Post, *And How Shall the Net Be Governed?: A Meditation on the Relative Virtues of Decentralized, Emergent Law*, in COORDINATING THE INTERNET 68-69 (Brian Kahin & James H. Keller eds., 1997) (upholding that conclusion for the larger regulative perspective in cyberspace).

208. See Gillett & Kapor, *supra* note 135, at 18.

209. This “Split DNS” (or “two faced DNS”), is also a corollary of this same fragmentation, followed by the loss in communication between a particular FQDN and an IPv4 address, whenever it ceases to be universal and steady.

210. See Farrell & Solaner, *supra* note 181 (for a sequential game model suggesting that both excess inertia and excess momentum can occur. With excess inertia, new tech-

ferent parties having simultaneous moves, the nature of the problem may be equivalent to coordination games, such as the prisoner's dilemma, where no efficient incentive to cooperate prevails, and parties prefer to act strategically. Left alone, self-interested *de facto* standard setters for example, RIRs, and even nations, would have an incentive to benefit their own customers for example, regional and local ISPs and their users, at the expense of the overall DNS stability at large.²¹¹ In contrast, when parties make sequential moves, such as when several infrastructure standards are competing a "battle of the sexes" game model prevails, and the lack of incentive to cooperate is not merely strategic, but results from the existence of alternative preferences and would, once again, justify central coordination of the prevailing infrastructure standards.

In addition, in the infrastructure development phase, switching costs between technological ingredients among standard setters are typically slower to overcome and often high.²¹² Seen through the perspective of standard setters, this once again results from the purposely-low level of effective price competition.²¹³ Also, once defective but potentially stable infrastructure technology is developed, infrastructure standards become extremely costly to change.²¹⁴ Finally, the incentive of preserving a coordinated installed base, may be positively externalized to the backbone telecommunications *services* market, as it may both encourage and facilitate the ability and will of early dominant infrastructure standard setters to remain pre-committed to future infrastructure transit compatibility. An early example of this model can be seen in the decision taken by eight subsidized industrial regional networks—BARRNet, CICnet, MIDnet, NEARnet, NorthWestNet, NYSERNet, SURAnet, and WestNet—which

nology may not be adopted even if it is superior to existing technology); Joseph Farrell & Garth Solaner, *Standardization, Compatibility and Innovation*, 16 RAND J. ECON. 70, 70-83 (1985) [hereinafter Farrell & Solaner, *Standardization*]. Paul David offered the anecdotal illustration of the perseverance of the "QWERTY" keyboard as an inefficient standard that was locked-in. See Paul A. David, *Clio and the economics of QWERTY*, 75 AM. ECON. REV. 332-337.

211. Notably, this is also the risk facing ICANN's present International domain name ("IDN") policies, as will be explained in Part V.C.

212. Joseph Farrell & Garth Solaner, *Coordination Through Committees and Markets*, 19 RAND J. ECON. 235 (1988).

213. See *id.* (following the authors suggestion that when the value of coordination on a standard is large relative to the value a firm attaches to adopting its preferred standard, the standard-setting committees are very likely to reach ultimate agreement on a standard but also vice versa).

214. Reidenberg, *Lex Informatica*, *supra* note 1, at 583 (estimating that the cost of change at the local level will be imposed directly on individual users, while change at the network level will be borne by network operators).

announced, the formation of the Corporation for Regional and Enterprise Networking (“CoREN”) on May 27, 1993. This decision was eventually intended to advance interconnection and blur the distinction between regional and backbone providers, by choosing MCI (subsequently acquired by WorldCom) as their backbone provider.²¹⁵ As a practical matter, by handling large amounts of Internet traffic, the CoREN backbone achieved a wide installed base of infrastructure technology, and accordingly, was able to bill regional affiliates an efficient internal transfer prices for the transport service provided.²¹⁶

Accordingly, it is unlikely that potential price competition will efficiently motivate *de facto* standard setters to switch infrastructures. Furthermore, *de facto* standard setters will generally be unable to establish monopoly power on infrastructure technology. Thus, they should not be encouraged to compete directly on infrastructure production and coordination at large.²¹⁷ Unless a monitored switch in infrastructure is centrally coordinated, it is highly expensive and typically not in the best interest of *de facto* standard setters to switch to uncoordinated, platform-dependent standards.²¹⁸ In sum, self-interested standard setters should not be left in charge of coordinating a new superior infrastructure because of direct early coordination sunk costs, best absorbed by *de jure* standard setters, and lost values from derivative negative network externalities.²¹⁹ A moni-

215. Michael A. Einhorn, *Pricing and Competition Policies for the Internet*, in PUBLIC ACCESS TO THE INTERNET 344 (Brian Kahin & James Keller eds., 1995).

216. *Id.*

217. *See, e.g.*, NICHOLAS S. VONORTAS, COOPERATIVE RESEARCH IN R&D-INTENSIVE INDUSTRIES 244 (1991). Vonortas stated,

In contrast [to “maturing segments of R&D-intensive industries” and the case of “declining industries”], multi-firm joint ventures for research are likely to be harmful in new, fluid technology industries . . . [C]ustomers of such industries are likely to lose from broad research cooperation since it will limit healthy competition in downstream markets.

Id.

218. *See, e.g.*, Jane K. Winn, *Consumers and Standard Setting in Electronic Payments Regulation*, 5 ELEC. BANKING L. & COM. REP. 11, 15 (discussing this dilemma in electronic payment standards).

219. FARRELL & SOLANER, *supra* note 12, at 4 (emphasizing the positive correlation between the early coordination (sunk) costs hurdle and avoiding later motivation to overcome dependent switching costs). One important mean of pre-committing leading *de facto* and even gray interest group standard setters to a stable infrastructure would be to “capture” their competitive motivations by enrolling them into infrastructure R&D by funding and subsidizing their early participation as was done in practice, as explained earlier.

toring policy regarding infrastructure developers is therefore imperative to generating the indispensable infrastructure installed base.

3. *Indirect Costs*

As a general matter, network effects positively benefit marginal participants due to the addition of new systems and the growing number of users.²²⁰ Typically, these effects suggest that network goods, priced by standard producers, and thus also network access costs, valued by the network consumers, would be optimally fixed, thus facilitating the widespread adoption of a standard.

However, due to the existing differences between the two main types of standards in cyberspace, namely infrastructure and application standards, each type of standard is affected by the network externality differently. Infrastructure standards seem to carry a positive network effect because there is a strong need for stable, performance-based infrastructure. Positive network effects are not merely a reflection of infrastructure interoperability but the very reason for their central *ex ante* adoption. Moreover, due to the non-competitive environment in which they are adopted, the positive influence of infrastructure standards on the installed base is indirect rather than direct because those standards end up leading to a derivative type of standards and markets,²²¹ in the form of application standards during a later standardization phase.

Accordingly, a derivative advantage of central coordination of infrastructure standardization is the minimization of the negative costs created by multi-standardization. Traditionally, in non-technological markets, these costs are present when persons other than the purchaser consume or use the product and may result from poor quality.²²² Conversely, in the early development phase, these costs are efficiently minimized, based on the qualitative rationale discussed earlier. For these reasons, central gov-

220. For primary works on the theory of network effects, see Farrell & Solaner, *Standardization*, *supra* note 210; Farrell & Solaner, *Installed Base*, *supra* note 210; Michael L. Katz & Carl Shapiro, *Network Externalities, Competition and Compatibility*, 3 AM. ECON. REV. 73, 424-40 (1985). See also Mark A. Lemley & David McGowan, *Legal Implications of Network Economic Effects*, 86 CALIF. L. REV. 479 (1998).

221. For literature on indirect network effects, see Chien-fu Chou & Oz Shy, *Network effects without network externalities*, 8 INTL. J. INDUS. ORG. 259-70 (1990); Chien-fu Chou & Oz Shy, *Partial compatibility and supporting*, 41 ECON. LETTERS 193-97 (1993); Jeffrey Church & Neil Gandal, *Complementary Network Externalities and Technological Adoption*, 11 INTL. J. INDUS. ORG. 239-60 (1993); Jeffrey Church & Neil Gandal, *Network Effects, Software Provision, and Standardization*, 40 INTL. J. INDUS. ORG. 85-103 (1992).

222. Hirshhorn, *supra* note 129, § 3.

ernments are arguably best at achieving an ex ante critical mass of infrastructure standards, while most efficiently internalizing the costs of diffusing network effects, through funding industry organizations in research and ultimately, in endorsing and formally adopting standards.²²³

In contrast to infrastructure standards, application standards are largely subject to positive ex ante network effects in their traditional competitive form. Accordingly, they only create the need for ex post intervention meant to minimize negative effects, largely through antitrust law.²²⁴ Here it is commonly agreed that in a competitive environment, new generations of standards will work at a significant disadvantage unless they are compatible with prior generations, so that the installed base of consumers could transfer data from one product to the next subject to productive competition between standards and their *de facto* developers.

IV. THE MODIFICATION PHASE: THE COMMERCIALIZATION OF CYBER STANDARDS

During the development phase, the U.S. government directly coordinated infrastructure design. In fact, the TCP/IP model was originally born out of the Department of Defense's ARPANET.²²⁵ Sponsored and coordinated by ARPANET, TCP/IP was meant not to be owned by any corporation or individual, and was thus created outside the dominion of commercial *de facto* standard setters. As an open standard, the TCP/IP protocols are also defined by RFCs ("Requests for Comments"), which are now part of the public domain. Infrastructure production was primarily technically, rather than commercially, motivated. That policy eventually changed, with both technological and economic development taking place. These central changes are what arguably led to the substantial commercialization of standard-setting activity in cyberspace, followed by a substantive growth in the number and influence of *de facto* and gray standard setters. Empirically, it has yet to be proven whether these changes should be regarded as an independent phase in the technological standardization process or merely an intermediate technological step between two substantive phases

223. See, e.g., SHAPIRO & VARIAN, *supra* note 8, at 313-15 (discussing governments as cost-efficient in achieving critical mass of networks).

224. *Id.* at 313.

225. Dr. Vinton G. Cerf led the design team at Stanford University that developed TCP/IP and managed the U.S. governmental DARPA Internet project from 1976-1982. Along with Robert Kahn, both men are generally accepted as the early inventors of the TCP/IP as early as 1974. See Vinton G. Cerf & Robert E. Kahn, *A Protocol for Packet Network Intercommunication*, 22 IEEE TRANSACTIONS ON COMM. 637 (1974).

of standardization. Even so, the central implications of the modification phase on the institutional identity criterion are worthy of discussion.

In 1995, soon after federal funding of the North American backbone ceased, private companies began operating their own backbone networks and selling telecommunications service access to their networks and the Internet. That privatization turned the Internet into a wide-ranging linkage of interconnected public and private computer networks joined by privately owned telecommunications facilities. This development paralleled the experience of traditional telecommunications standardization where a similar shift came in the mid-1980's, with the further growth of private-sector participation supported by arguments that manufacturers, carriers, and users should take part in formulating policies directly affecting them.²²⁶

One of the consequences of the rapid growth of the user community was a growing demand for additional facilitating standards.²²⁷ Based on a largely platform-dependent approach, the resulting software products facilitated the basic Internet data communication infrastructure.²²⁸ Soon after the uniform TCP/IP infrastructure was in place, new commercial network effects provided incentive for vendors to build TCP/IP compatible products. This technological change seems to have been the main constraint leading to the commercialization of the Internet's standards.²²⁹ Yochai Benkler recently noted that there were three major, chronologically close, technological developments that marked the shift to commercialization in cyberspace: the web server, the first graphical browser, and

226. LEE, *supra* note 101, at 121-22.

227. See WERBACH, *supra* note 4, at 17 ("Like a digital tornado, the vortex continues, as the new level of demand creates the need for additional capacity, and so forth.")

228. The derivative dependence of application standards on infrastructure is of central technological importance, as it explains the shift between the first two phases, as part of the process argument, described earlier. For a discussion supporting this premise, see CARGILL, *supra* note 1, at 28-31, 119 (discussing "implementation standards," using the example of the IEEE 802.3 Ethernet infrastructure standard); WILDER, *supra* note 14, at 6, 357; Leiner, *Brief History*, *supra* note 118, at 12; Mähönen, *supra* note 102, at 42. Technically, the interchange between cyberspace's standardized infrastructure and applications is vertically coupled: network technology drives the applications, and applications drive the networks. This continuous feedback is also known as the "spiral design" process. See NEXT GENERATION INTERNET, *supra* note 173.

229. Kesan & Shah, *supra* note 115 (suggesting that the first use of the term "commercial use" was made by Stephen Wolff, Director of the NSF Division for Networking and Communications Research and Infrastructure, who had placed the term in the backdrop of his formal statement); see also COMMERCIALIZATION OF THE INTERNET SUMMARY REPORT n.93, RFC 1192 (Brian Kahin ed., IETF Network Working Group Nov. 1990), at <http://www.faqs.org/rfcs/rfc1192.html> (last visited Dec. 18, 2003).

the first proprietary service provider.²³⁰ Soon after these developments, the Internet became highly commercial in nature. According to the FCC, the Internet has grown from 10 million users in 1995 to over 40 million by 1999.²³¹ A similar pattern holds for the purchase of Internet services, which grew from \$6.2 billion in 1995 to over \$300 billion in 2002.²³²

The shift to platform-dependent compatibility was followed by consumer-oriented price competition.²³³ This shift ultimately led to the commercialization of application software.²³⁴ Eventually, cyber standardization became very closely related to the commercial decision-making done in industrial companies.²³⁵

Thus, a central change in the standardization process was a shift from centralized, government directed, standardization planning to short-term *de facto* standardization. Traditionally, industrial-age innovation followed the linear sequence from scientific discovery to applied research and development, followed by production and marketing. The standardization process is time consuming, particularly if the number of participants having divergent preferences is high. When a market exhibits rapid tech-

230. Yochai Benkler, *Net Regulation: Taking Stock and Looking Forward*, 71 U. COLO. L. REV. 1203, 1206 (2000) (“[Those three changes] changed it all. It turned out that the net was not in the future; it was here . . . in what suddenly became the new popular (not to say mass) medium.”); cf. Lemley, *Antitrust*, *supra* note 94, at 1052; Lawrence Lessig, *The Limits in Open Code: Regulatory Standards and the Future of the Net*, 14 BERKELEY TECH. L.J. 759, 760-61; Leiner, *Brief History*, *supra* note 118, at 13; Robert Cailliau, *A little History of World Wide Web: From 1945-1995* (1995), at <http://www.w3.org/history.html> (last visited Dec. 1, 2002).

231. See F.C.C., CONNECTING THE GLOBE: A REGULAR S GUIDE TO BUILDING A GLOBAL INFORMATION COMMUNITY (1999) (discussing the FCC’s policy record), at <http://www.fcc.gov/connectglobe/sec9.html> (last visited Dec. 1, 2002).

232. *Id.*

233. See FARRELL & SOLANER, *supra* note 12, at 5; Manfred M. Fisher & Börje Johanson, *Networks for Process Innovation by Firms: Conjectures from Observation in Three Countries*, in PATTERNS OF A NETWORK ECONOMY: ADVANCES IN SPATIAL AND NETWORK ECONOMIES 264 (Börje Johanson et al. eds., 1993) (arguing that “[i]n finished standardized product (or mature products) markets, which serve for mass production—price competition prevails, as a quantitative measurement, is the main criterion for the decision making of buyers”).

234. LESSIG, *supra* note 8, at 52.

235. See, e.g., Mark A. Lemley & David McGowan, *Could Java Change Everything? The Competitive Propriety of a Proprietary Standard?*, 520 PLI/PAT 453, 471; see also Ken Krechmer, *Cathedrals, Libraries and Bazaars*, Remarks at the Association of Computing Machinery (ACM) Symposium on Applied Computing Meeting (Mar. 10-13, 2002) [hereinafter Krechmer, *Cathedrals*] (suggesting that the production of standards and commercial software is becoming closely related, extracting ten lessons that may apply to standards development as well as software program development), at <http://www.csrstds.com/cathedrals.html> (last visited Dec. 2, 2003).

nological growth, as the cyber market has, the time required to develop an IT standard is typically longer than a single product or system's life cycle. To cope with this, standards bodies began, in the infrastructure development phase, to act in anticipation of the technology, developing standards before products were produced.²³⁶ In the infrastructure development phase these were known as anticipatory standards and included, for example, V.32 modem, X.25 packet interface, ISDN and ultimately TCP/IP.²³⁷

In practice, this standardization work, even at the technical committee level, was not related to R&D work or innovations. Innovations were submitted to the standards process, wherein state of the art technology froze to standards. Accordingly, in the development phase, standards were typically designed to include available high technology specifications, but did not drive towards new innovations within the standardization process itself. Yet with the growing shift to commercialization of application standards in the implementation phase, there has been an erosion in the effectiveness of the anticipatory standards process. This is evident in the rise of informal standardization. Thus, in practice, gray and *de facto* activity was initiated and driven by commercial implementers.²³⁸ Eventually, with less efficient incentives towards strategic social planning, and as a result of rapidly increasing consumer demand for application standards, the development and implementation of standards became decreasingly less anticipatory.²³⁹ Instead, anticipatory standards became responsive as they

236. CARGILL, *supra* note 1, at 45-46; Mähönen, *supra* note 102, at 38 (concluding that "the place for innovations is outside the formal standards meeting—standards just agree on the solution").

237. In fact, Bonino and Spring argue that anticipatory standards act as mechanisms for collective planning, i.e., they are an embodiment of a central industrial policy. Accordingly, their prevalence serves as an additional rationale favoring central standardization of the developmental phase. See Weiss & Spring, *Standards as Change Agents*, *supra* note 28, at 99.

238. See, e.g., Egyedi, *supra* note 3, at 55.

239. *Id.* at 54-55 (suggesting that, in fact, application standards development has begun to occur in parallel). In a conversation with Carl Cargill, Egyedi further suggested that anticipatory standard-setting activity (as in the early phase of cyberspace standardization), is now practically "dead, followed by the shift to standardize 'existing practice.'" See Cargill Interview, *supra* note 158. For example, ECMAScript, which was standardized by ECMA—based on the Javascript programming language, established earlier by Netscape. *Id.*; see also Elaine Baskin et al., *The Six Dimensions of Standards: Contribution Towards a Theory of Standardization 2* (Nov. 17, 1999) (suggesting that due to high market demand followed by the decrease in anticipatory standardization, large users with specific needs may force standards creators to become standards seekers), at <http://www.standardsresearch.org/presentations/6Dimen.doc> (last visited Dec. 18, 2003).

adapted to preceding innovative activity that was only later standardized.²⁴⁰

The major risk of such rapid, market-oriented, standard-setting is the premature stifling of innovation.²⁴¹ It may produce design standardization too early in the technological life cycle. A technology may thus be forced into “early maturity” not because of technological limitations but rather by competitor collusion.²⁴² Consequently, with the growth of application standardization, a new institutional balance had to be struck. This balance was finally established in the implementation phase.

V. THE IMPLEMENTATION PHASE: THE RISE OF AUTONOMOUS INSTITUTIONS

Whenever technology matures, the explosion of new markets for both early core services and markets for peripheral technology, in the form of application-standardized products should promote the rise of autonomous standard-setting institutions. As cyberspace entered a substantive new phase of commercialization, network effects provided incentive to vendors to build TCP/IP-compatible products. Ultimately, commercial decision-making by application producers became a controlling factor in cyberspace’s standardization reality. Thus, in the face of an increasing number of *de facto* and gray standard-setting institutions,²⁴³ government intervention would optimally take a secondary role. In practice, while the path has been neither clear nor straight, the trend of governmental interpretation of the public interest in the United States exhibits a definite pattern of decreasing intervention in standardizing cyberspace.

In the implementation phase, only self-restrained political institutions, typically governmental agencies and particularly the FCC, will come close to optimizing competition among autonomous standard setters for creating beneficial standards. However, a role for voluntary industry regulation to formalize *de facto* or gray standardization will become essential due to

240. See Krechmer, *Cathedrals*, *supra* note 235.

241. CARGILL, *supra* note 1, at 118 (concluding that “if standards, voluntary or otherwise, are introduced, they will fail, since standards act to stabilize a market”); Mähönen, *supra* note 102, at 38; Sloane, *supra* note 25, at 8 (exemplifying the HTML protocol); *cf.* LIBICKI, *supra* note 3, at 14-16 (suggesting that the optimal moment for standardizing should be “if the technology matures before the market takes off standardization can occur smoothly in between”).

242. FREDERICK BETZ, *MANAGING TECHNOLOGY* 75 fig. 2 (1987); *see also*, Lemley & McGowan, *supra* note 220, at 471. In the following part the resulting institutional consequences of this change will be examined.

243. See CARGILL, *supra* note 1, at 102-03.

new risks of ex ante technological stagnation and/or ex post anticompetitive effects on the one hand, and the need to sustain private competition, on the other. By the same token, growing concerns about governmental technological bias toward standards that best serve its policies through “code control” lead to the conclusion that formal or *de jure* industry standardization may be the most efficient mechanism for chilling direct governmental incentives for mandated standardization, beforehand or ex post. In accordance, this part presents an updated cost-benefit analysis of the different standardization costs. This cost-benefit analysis supports the comparative institutional analysis.

A. Application Standardization: The Cost-Benefit Analysis

Application standards should primarily be the product of competition, subject to the widespread need to develop, approve, and compatibly adapt applications to infrastructures and other complementary applications. These production constraints suggest a delicate balance between two seemingly contradictory processes: the creation of innovative product variety,²⁴⁴ and the reduction of that variety through user-based qualitative selection. Effective technological acceptance requires that these two processes be kept in balance.²⁴⁵ In the highly competitive environment of cyberspace, this balance is not always maintained because of genuine risks of inefficient production. As far as development per se is concerned, this balance derives from the common threat that rapid commercial activity will curtail innovation prematurely. This may lead technology into early maturity at a point where standards are still inefficient. Thus, if a leading standard is set before a technology reaches maturity, more radical innovation may be economically inefficient. As anticipatory standardization has declined, the risk of early maturity seems only to have grown. Moreover, as far as standard adaptation and approval (or formalization) are concerned, formal industry standard setters should be in charge of optimizing compatibility, so as to generate and maintain new and competitive economies of scale in application standards. Thus, the need to diminish the basic cost of early maturity resulting from anticompetitive collusion and inefficient incompatibility may necessitate a role for formal industry policy planners in optimizing *de facto* and gray standardization. Projected costs of this sort of policy planning fall into three categories: administrative, compliance, and indirect.

244. When n actors bilaterally agree to a set of standards, then $n*(n-1)/2$ rules must be defined. See SHAPIRO & VARIAN, *supra* note 8; Shapiro & Varian, *Standards Wars*, *supra* note 182.

245. See SHAPIRO & VARIAN, *supra* note 8, at 233.

1. *Administrative Costs*

a) *Eliminating Duplicative Efforts*

Standardization authorities and formal standardization organizations should encourage cooperation and prevent R&D overlap.²⁴⁶ Proceeding on this assumption, Congress passed legislation, first in 1984 through the National Cooperative Research Act,²⁴⁷ and later, in 1993 through the National Cooperative Research and Production Act (“NCRPA”),²⁴⁸ indirectly aimed at facilitating autonomous commercial innovation initiatives. Designed to accelerate existing pro-competitive policy, the NCRPA clarified how U.S. antitrust laws substantively applied to joint R&D and joint production ventures. Drafted to promote R&D activity by providing a special antitrust regime for joint ventures, the NCRPA requires U.S. courts to judge the competitive effects of a challenged joint R&D, joint production venture, or a combination of the two, in appropriate relevant markets applying a “rule-of-reason” legal standard.²⁴⁹ This rule of reason takes into account “all relevant factors affecting competition.”²⁵⁰ The NCRPA also instituted a voluntary safe harbor procedure whereby the Attorney General and the FTC may be notified in advance of a joint R&D or production venture.²⁵¹ Accordingly, numerous U.S. consortia have already filed notices under section 4305(a) of the NCRPA²⁵² in order to limit their potential penalties and liabilities for anticompetitive standardization activities.²⁵³ Overall, however, the NCRPA has not been widely exploited by

246. See CARGILL, *supra* note 1, at 131-38; Chapin, *supra* note 207; Thomas M. Jorde & David J. Teece, *Acceptable Cooperation Among Competitors in the Face of Growing International Competition*, 58 ANTITRUST L.J. 529, 538 n.28 (1989).

247. National Cooperative Research Act of 1984, Pub. L. No. 98-462, 98 Stat. 1815 (codified as amended at 15 U.S.C. §§ 4301-4306 (1993)).

248. National Cooperative Research and Production Act, 15 U.S.C. §§ 4301-4306 (1993) [hereinafter NCRPA]. By enactment of the NCRPA Congress reinforced the existing tolerance expressed for joint ventures engaged in legitimate research and development and standard-setting projects, mandating application of the rule of reason in cases brought against registered ventures. See Jack E. Brown, *Technology Joint Ventures to Set Standards or Define Interfaces*, 61 ANTITRUST L. J. 921, 933 n.55 (1990) (adding that the congressional interest in furthering joint activities was also reflected in the repeated legislative attempts to extend the NCRPA’s protections to joint production and manufacturing, as well as research ventures).

249. 15 U.S.C. § 4302.

250. *Id.*

251. 15 U.S.C. § 4305(a).

252. See DENNIS W. CARLTON & JEFFREY M. PERLOFF, *MODERN INDUSTRIAL ORGANIZATION* 522 (3d ed. 2000).

253. See also LIPCZYNSKI & WILSON, *supra* note 157, at 229; W. Rupert Maclaurin, *The Process of Technological Innovation: The Launching of a New Scientific Industry*, 40

standard setters.²⁵⁴ This can likely be attributed to the difference between the protections present in the Act and the protections Congress apparently intended to grant.²⁵⁵ Designed primarily to enhance early innovative R&D activity, the amended 1993 Act arguably still suffers from certain limitations that result from the limited view of the technological standardization process.

First, subject to a few exceptions as set forth in 15 U.S.C. §§ 4301(b)(2)(A)-(C), NCRPA excludes joint marketing and sales ventures.²⁵⁶ The Act's distinction between "pure" research and its attendant business implementation such as marketing and selling, in both time and space is potentially artificial.²⁵⁷ In order to commercialize a product effectively, development engineers must receive feedback from the production and marketing stages.²⁵⁸ Moreover, especially in the implementation phase when standard-setting committees are no longer engaged in significant anticipatory standardization, the production of application standards gets a new commercialization and marketing emphasis instead of the R&D orientation seen in the early development phase. Indeed, innovation in manufacturing processes may be just as important for the success of a new product as the design innovations conceived in the early development phase.²⁵⁹

Second, the Act exhibits a xenophobic quality by disqualifying non-American joint ventures.²⁶⁰ Thus, the new law did not protect many production ventures, because the "principal facilities for such production"

AM. ECON. REV. 90 (1950). Another significant contribution to this policy came in 1987 with the agreement between the International Organization for Standardization ("ISO") and the International Electrotechnical Commission ("IEC") to cooperate in the Joint Technical Committee ("ISO/IEC JTC 1"). The purpose of this cooperation was to eliminate the serious overlap between ISO and IEC in their standardization activity. *See* LEE, *supra* note 101, at 44-45.

254. *See* Joseph Kattan, *Contemporary Antitrust Analysis of Joint Ventures: Why It Makes Sense to Stay the Course*, Remarks at the Federal Trade Commission's Hearings on Joint Venture Project (June 5, 1997), at <http://www.ftc.gov/opp/jointvent/kattan.htm> (last visited Dec. 3, 2003); Krechmer, *Cathedrals*, *supra* note 235 (suggesting that the production of standards and commercial software is becoming closely related, extracting ten lessons that may apply to standards development as well as software program development).

255. Kattan, *supra* note 254, Krechmer, *Cathedrals*, *supra* note 235.

256. 15 U.S.C. § 4301(b)(2) (2003).

257. *See* Anthony L. Clapes, *Blinded by the Light: Antitrust Analysis of Computer Industry Alliances*, 61 ANTITRUST L.J. 899, 916 (1993).

258. *See* Jorde & Teece, *supra* note 246, at 582, 589.

259. *Id.*

260. *Id.*

clause mandates that production ventures have their principal production facilities in the United States in order to qualify for coverage.²⁶¹ Furthermore, according to the Act “each person who controls any party to such venture” must be a United States company or must be incorporated in nations that treat United States companies fairly under their own antitrust laws governing production joint ventures.²⁶² Since both ad hoc and gray standard-setting is taking on greater international significance, the Act must be amended to reflect the international realities that change introduces.

In addition to coordinating R&D cooperation, another means of preventing overlap is achieved through the industry’s expanding voluntary promotion of electronic publishing of research materials. By increasing the efficiency of information flow to manufacturers trying to improve productivity it is possible to diminish the inefficiency associated with protracted trial and error in manufacturing processes.²⁶³ At the same time, it is important for manufacturers in this environment of increasing competition to protect their R&D investments against misuse in other markets.²⁶⁴ In practice, however, there is a notable growth in the standardization community in common R&D joint ventures, entered into to share high R&D costs and uncertainty about market developments. Accordingly, several industry standardization organizations, for example IETF, make their standards (or parts of them) called RFCs, available electronically and waive all publishing or use fees.

b) Reducing Search Costs

In markets of application standards, searching for information regarding the different value measurements of standards such as costs, quality,

261. 15 U.S.C. § 4306(1).

262. 15 U.S.C. § 4306(2). H.R. 1313 passed the House of Representatives on May 18, 1993. *See House Passes Bill Easing Antitrust Law for Parties Involved in Joint Ventures*, 64 ANTITRUST & TRADE REG. REP. (BNA) No. 1615, at 600 (May 20, 1993).

263. *See, e.g.*, Paul A. David & Shane M. Greenstein, *The Economics of Compatibility Standards: An Introduction to Recent Research*, 1 ECON. INNOVATION & NEW TECH. 1, 3-42 (1990).

264. Industry standard setters may try to use the information they control to influence market trends, against the public’s interest, while raising bargaining threat costs of standardization activities to a non-cooperative level beyond the NCRPA antitrust “rule of reason.” In “unreasonable” situations there will be a need for governmental intervention. Moreover, if certain de facto standard setters are excluded from such a joint venture, they may be competitively disadvantaged because they cannot conduct research on the same scale as the members of the venture. *See, e.g.*, LAWRENCE A. SULLIVAN, HANDBOOK OF THE LAW OF ANTITRUST 298-303, nn.167-70 (1977) (discussing the anticompetitive effects of membership restrictions).

and services should be communicated to consumers in a way that is easily comprehended and used as the basis for comparison. In such markets it is difficult to measure the quality or value of these products for consumers. This intrinsic difficulty is true for information markets at large.²⁶⁵ It is thus also true for consumer-oriented markets of application standards or software products.

As a result of this tension and because information about quality is more costly to supply and process than information as to price, public agents with sufficient information to make and certify quality judgments are needed to act as monitoring proxies for consumers approaching application standard markets.²⁶⁶ In cyberspace, such formal monitoring of standard-setting was largely achieved by creating a consensus around individual technologies. While the government is best suited to subsidize such activity,²⁶⁷ it was competitively neutral industry organizations that took charge of diminishing the search costs necessary to ensure the adoption of adequately efficient standards. In such a situation, industry standard setter's monitoring proxies should avoid the situation in which the market either does not pick up momentum because of too much variety, or malfunctions due to market failures.

Standard development best contributes to the collective R&D learning process when inefficient sources of variety in design are avoided by minimizing information costs through formal industry intervention. At the same time, the choice made by the industry naturally adds to the costs of the system, since rulemaking power remains with the *de facto* standard setters. In order to preserve the competitive benefits, while maintaining the 1997 Report's self-regulatory paradigm for application standardization, formal or *de jure* industry standardization organizations should be limited only to a secondary role of monitoring and certifying *de facto* production in the implementation phase.²⁶⁸

2. Compliance Costs

To be effective, infrastructure must be hegemonic and thus require compliance with infrastructure standards. Applications standards require a different type of compliance. There is a smaller need for a unified vision of anticipatory standard-setting activity for both internalizing and uli-

265. ARROW, *supra* note 157, at 615.

266. See Akerlof, *supra* note 165; Trebilcock, *supra* note 162, at 92-95; Kip W. Viscusi, *A Note on "Lemons": Markets and Quality Certification*, 9 BELL J. ECON. 277 (1978).

267. See discussion *infra* Part V.B.2.a.

268. Cf. OGUS, *supra* note 1, at 110.

mately reducing existing costs through formal institutions. In application standard-setting, compliance with standard-setting became a regulatory *preference*, not a requirement.

However, even in this competitive environment, too great a departure from unvarying standards by *de facto* standard setters still generates significant compliance costs. Firms are presently incentivised to overstate the costs of complying with other developers in order to increase the pre-bargaining threat costs of a desired coordinated standard.²⁶⁹ Such strategic behavior on the part of self-interested standard setters takes place whenever firms make their standards proprietary, which typically translates to less compatibility. On the other hand, imposition of inflexible standards on newer firms may create barriers to market entry thereby protecting older firms from competition. Reducing such compliance costs must therefore be accomplished within a paradoxical regulatory framework: the more *de facto* activity is encouraged by formal agents, the more independent and potentially non-compliant *de facto* standards may become.

The idea that compliance costs diminish with application standards also complies with economizing transaction costs, generated in cooperative production of standards. In the world outside cyberspace, it is agreed that when transaction terms are standardized, bargaining conditions are lubricated, resulting in reduced transaction and other costs to producers and users.²⁷⁰ It is also part of the general characteristic of compliance costs, especially with respect to the implementation phase, where bargaining as a means of coordinating standards becomes feasible.²⁷¹ Indeed, one of the primary purposes of standardization is to eliminate bargaining over details of individual transactions when bargaining costs and unpredictable customized bargains would deter producers from making valuable products like application software available.

The surplus created by standardization exists for a number of reasons. First, application standards embed an *ex ante* constraint as automated, self-adapting, and self-executing regulative subject matters. Standards can therefore be encouraged to impose, prevent, or permit transactions from occurring. Because standards define the level of access to both a technol-

269. *Id.* at 155.

270. *Id.* at 122-24.

271. *See, e.g.*, Clayton P. Gillette, *Intervention and Standardization in Electronic Sales Contracts*, 53 SMU L. REV. 1431, 1435 (2000) (discussing the importance of promulgating common terms of trade in cyberspace, for example, for the XML standardized languages market); Edward L. Rubin, *Computer Languages as Networks and Power Structures: Governing the Development of XML*, 53 SMU L. REV. 1455, 1475 (2000) (same).

ogy's source code and interface, before, during, and after application standards are produced, such standards profoundly affect transactions.²⁷² Second, in cyberspace, direct product-signaling options to consumers increase in both quality and number.²⁷³ Thus, based on software design, application standards give greater regulative control to users, which allow them to more efficiently express their preferences.²⁷⁴ As signaling options increase, production based on future consumer preferences is also likely to simplify and potentially reduce the cost of transactions among producers and users alike. Third, based on consensual practices of fair competition, users of application standards may lower enforcement costs against non-compliant standard setters through monitoring and compliance technologies, for example, blocking and filtering technology or even censorship software against uncooperative behavior on the part of such non-complying standard setters. These technological constraints can be designed to increase users' ability to facilitate demand for application standards in the face of the growing need for advanced widespread application standards.²⁷⁵ The effect of the expansion in consumer diversity on demand remains, therefore, contradictory. On one hand, cyberspace increasingly allowed consumers to be involved in regulative activity, allowing users to designate preferences for standards. On the other, the need to coordinate both production efforts and compatibility intended to meet diverse prefer-

272. The IETF is a good example of this front. All standard proposals and main standards are available at their web site. See *Internet Engineering Task Force Website*, at <http://www.ietf.org>. As well, GII Virtual Roundtable is an organization-neutral online forum that includes a large set of standardization organizations such as IEC, IrDA, ISO, JTC1, POSI, VESA, ETSI, DAVIC, etc. See *GII Virtual Roundtable Website*, at <http://www.globalcollaboration.org>. Its goal is to provide a common forum for users and consumers alike for voicing their opinions on various active standardization processes. Moreover, the principle that ensures that implementers of a standard should have open access to intellectual property rights and would be required to meet the standard on a non-discriminatory, fair and reasonable basis was adopted by many standards developing organizations, for example, ANSI, ISO, IEC, ITU and the IETF. *Id.*; see also Mähönen, *supra* note 102, at 44; Rada, *supra* note 98, at 29-30, 32.

273. See Reidenberg, *Lex Informatica*, *supra* note 1, at 570 (suggesting that signaling options for completing transactions in technological products increase).

274. *Id.*

275. To date, filtering software for data and documents on the Internet is mostly related to two main standard applications, namely, e-mail and the World Wide Web. Filters can facilitate individual choice at the browser or even the server level. See generally Paul Resnick & James Miller, *PICS: Internet Access Controls Without Censorship*, 39 COMMUNICATIONS OF THE ACM 87 (1996), at <http://www.w3.org/PICS/iacwcv2.htm> (last visited Dec. 1, 2002); R. Polk Wagner, *Filters and the First Amendment*, 83 MINN. L. REV. 755, 759-69 (1999); Jonathan Weinberg, *Rating the Net*, 19 HASTINGS COMM. & ENT. L.J. 453 (1997).

ences has increased as application standards became more complex and expensive to coordinate. Thus, with such regulative costs at stake, a role for industry formalization of *de facto* and gray standards remains essential.²⁷⁶

There is an additional general justification for industry stepping up and formalizing *de facto* and gray standards. Procedurally, formal industry has an ability to maintain compliance through rules of membership and procedure. Formal industry standardization is arguably the most efficient mediator of hegemony within the overall institutional framework because the industry as a whole is more likely to remain more competitively neutral, in comparison, and thus able to serve universal interests in the face of growing self-interested competition.²⁷⁷

In summary, changes in application standards designed by *de facto* or gray standard setters can mostly be promulgated based on a formalizing role of industry standard setters that possess the authority to both articulate and implement standards of their own. More importantly, they can formalize *de facto* and gray application standards and create common terms of trade. In addition, as centralized adopters of standards, industry organizations should be relied upon to publicize content and procedures of cooperative standardization practices.

3. *Indirect Costs*

Coordinating technologies required by new standards and innovative products may be cheaper for larger, well-organized agents than for smaller firms. However, this also means that the real winners in a system are *de facto* standard setters that are both flexible enough to identify opportunities to capture an architecture and powerful enough to sustain control over it. Whenever such *de facto* control is present, economies of scale may also generate external costs of delay in production. Often, one or more such *de facto* agents might be pushed towards delaying a final innovative outcome, hoping that technological or commercial developments will overtake an undesirable result. Beyond an efficient result, there is a risk that this process will produce an answer that will end by isolating vendors and fracturing the market. That may result whenever duplicative production is made,

276. See Rada, *supra* note 98, at 33; Reidenberg, *Lex Informatica*, *supra* note 1, at 587.

277. See 47 U.S.C. § 160 (2000); *id.* § 230 (a)(4) & (b)(2). In essence, this premise is also what led to the deregulation movement in telecommunications services. See WERBACH, *supra* note 4, at 29.

rather than uniting vendors behind a standard that would be well accepted by users.²⁷⁸

Because such inefficient competition is a market failure best minimized through governmental intervention, industry standardization organizations are not a natural proxy for such regulative activity.²⁷⁹ Nonetheless, as a policy matter, they may still serve to minimize the potential need for ex ante governmental intervention.

Thus, such ex ante industry intervention could coordinate *de facto* standard setters' right to supply standards. In an attempt to avoid self-regulating firms having monopolistic control over the right to supply standards, industry may channel such activity to competition for that right. Each competing self-regulatory firm may thus be required, as part of its bid, to submit its proposed regulatory rules. A second-tier public agency, acting as proxy for consumers, might then be able to determine the right to supply implementation standards or distinct classes of such standards. In cyberspace, this theoretical approach has not yet reached its potential. At present, the industry has the more limited role of coordinating production ex ante instead of coordinating the right to supply itself.²⁸⁰

278. See Kathleen M.H. Wallman, *The Role of the Government in Telecommunications Standard-Setting*, 8 COMMLAW CONSPICUOUS 235, 239-40 (suggesting that this was the case with the wireless network standards situation that has ended up producing one set of prevailing standards for the U.S. market (Time Division Multiple Access and Code Division Multiple Access) and another for most of the rest of the world (Global System for Mobile Communications)).

279. See LEE, *supra* note 101, at 128-29 (following the example of the wireless phone industry and suggesting that their joint interoperable failure has been neither on consensual nor on technical grounds, but because of ability, given CCI rules, of core groups to dominate the standard-setting process).

280. See OGUS, *supra* note 1, at 110 (acknowledging that, to date, such a system is already used for the allocation of public franchises, for example, in broadcasting services or airline routes). This environment in which industries might control the market for the right to supply will also be subjected to antitrust law limitations. See *Joint Ventures and Standard-Setting: Problems in the Current Framework* (paper by Ernest Gellhorn & W. Todd Miller presented to the Federal Trade Commission in the hearings on the Changing Nature of Competition in a Global and Innovation-Driven Age) (Oct. 26, 1995) (suggesting that cooperation among industry participants with a collective market share of less than merely 35 percent creates few risks, especially in dynamic high-technology markets, and thus should be immune from antitrust challenges); William F. Baxter, *The Definition and Measurement of Market Power in Industries Characterized by Rapidly Developing and Changing Technologies*, 53 ANTITRUST L.J. 717, 723 (1984) (proposing that R&D joint ventures possessing market shares of lower than 20 percent be considered benign); Robert F. Bork, *The Rule of Reason and the Per Se Concept: Price-Fixing and Market Division*, 75 YALE L.J. 373, 397 (1966) (suggesting a restriction of 25 percent); Joseph F.

B. Government Intervention

Apart from the indirect support for routine and limited *infrastructure* standard-setting activity during the implementation phase, governments should continue to play a restrained indirect role in vital application standard-setting activity due to institutional barriers on efficient participation.

This study generally holds that government's role should diminish during the implementation phase. Governmental institutions should not take corrective action to minimize market failures, when the same result can be achieved through private sector standard-setting activity. Specifically, the U.S. government followed this comparative institutional guideline when it noted that "even where collective agreements or standards are necessary, private entities should, where possible, take the lead in organizing them."²⁸¹ Furthermore, under section 401 of the Telecommunications Act, the Commission must refrain from regulating, so long as forbearance is consistent with the public interest.²⁸²

1. Direct Intervention: The Problem of Efficiency

There are four main reasons why governments tend to be inefficient in standard-setting during the implementation phase. First, and most importantly from the view of an overall standardization process, rapid technological developments generally outpace the slow, ex post bureaucratic decision-making characteristic of government institutions.²⁸³ Moreover, once production patterns are adopted, they acquire a taken-for-granted quality and are not easily dismissed or changed.²⁸⁴ As a general matter, when standards are developed or ordered by governments, they tend to be designed for particular needs (especially, DoD programs), which are largely less commercially-oriented or desirable to small commercial programs with fast response times.²⁸⁵ Eventually, while the market may replace an inefficient standard by competitive technological "leapfrogging," there are typically fewer guarantees that government will do the same. Furthermore, except for the competitive self-standardization practice (backed by a governmental monitoring role), there is no market for *control* of application standard setters. Thus, the principals (politicians and citizens) cannot eas-

Brodley, *Joint Ventures and Antitrust Policy*, 95 HARV. L. REV. 1523, 1541 (1982) (proposing a restriction of 40 percent).

281. See Clinton & Gore, *A Framework*, *supra* note 14, § 9.

282. See sources cited *supra* note 277; see also WERBACH, *supra* note 4, at 29.

283. BREYER, *supra* note 1, at 106; LIBICKI, *supra* note 3, at 354; Gillett & Kapor, *supra* note 135.

284. MARCH & OLSEN, *supra* note 13, at 52.

285. LIBICKI, *supra* note 3, at 354-55.

ily dismiss ineffectual officials.²⁸⁶ Government agencies are generally composed of career public servants, not market participants and, as a result, they often do not involve the most qualified and industrially impartial individuals in the industry for the standard-setting process. This is the inherent danger of bureaucracy, particularly when it attempts to standardize a fast-moving area like the Internet.

Second, today's technology may limit the ability of governments to understand the issues in direct and substantive standardization. Particularly, in a market economy, such as that of the implementation phase, it is not apparent that the government will understand what constitutes an inefficient standard. Ultimately, the need to write standards with an eye toward enforcement raises difficulties that may potentially compel the agency to write standards that do not meet the primary policy objective of efficiency.²⁸⁷

Third, government standard-setting agencies may be slow in gathering information and may not always have access to the best information since information flows may be invisible to them.²⁸⁸ This problem also exists in the infrastructure development phase. However, in that pre-commodified phase, competition was scarcer than it is in the implementation phase and the U.S. government successfully monitored such information flows. Thus, even with the best of intentions, a government standard-setting agency may simply pick a less efficient standard, leading to stagnation of robust standards, even when they may appear technologically ill conceived. Alternatively, government agents may not always have the ability

286. MARCH & OLSEN, *supra* note 13, at 52; OGUS, *supra* note 1, at 112.

287. BREYER, *supra* note 1, at 112; OGUS, *supra* note 1, at 170.

288. See, e.g., Anton & Yao, *supra* note 169, at 252 ("Technical judgments are also critical to assessing whether the benefits of the standard outweigh the costs, but most courts and agency officials lack a technical background."); David Bazelon, *Coping with Technology Through the Legal Process*, 62 CORNELL L. REV. 817 (1977); Richard A. Posner, *Antitrust in the New Economy*, 68 ANTITRUST L.J. 925, 925 (2001) (arguing that in the new innovative economy the real legal problem lies on the institutional side because the enforcement agencies and the courts do not have adequate technical resources, and do not move fast enough, to cope effectively with a rapidly changing business sector); Howard Shelanski, *Regulating at the Technological Edge: New Challenges for the F.C.C.*, 2000 L. REV. MICH. ST. U. DET. C. L. 3 (2000). See generally Harry S. Gerla, *Federal Antitrust Law and Trade and Professional Association Standards and Certification*, 19 DAYTON L. REV. 471, 503 (1994) (examining how eventually "antitrust courts generally have been favorably disposed toward trade and professional association standards"); Howard Shelanski, *Competition and Deployment of New Technology in U.S. Telecommunications*, 68 U. CHI. L. REV. 1 (2000) (supporting a repeated-play policy in establishing the right equilibrium to the telecommunications industry in the FCC, in the face of the present technological change).

or will to perform such evaluations. As has been noted in the literature on public choice theory, government agencies in a position to influence the outcomes of market competition are highly susceptible to “capture” by private entities with an interest in the outcome. Thus, there is much less of a guarantee that a governmental standard-setting body will act efficiently in the public interest through direct standardization activity, even when it is possible for a government agency to discern what in fact that interest is.²⁸⁹

Fourth, there is the increasing trend towards globalization reflected in IT regulation and cyberspace’s standardization today. As suggested, institutional analysis refers narrowly to “collective action” and “interest groups” at the national level. These paradigms were generalized before new global regulatory realms were understood as such. These institutional paradigms assume a homogeneous national institutional structure, based on the United States or some hypothetical Western democracy rather than the borderless global arena. Accordingly, there is no good understanding of the relationship between national and international standardization, or even if this is an important or useful distinction. Indeed, fundamental to the developments affecting the institutional standardization debate is the emergence of a global economy in which the United States, as one national government, may not play the dominant role.²⁹⁰ In the future, the rise of international standardization organizations might change the current institutional balance by dominating both application and infrastructure

289. See Lemley, *Antitrust*, *supra* note 94, at 1063 n.76 (“This almost happened in the case of the United States HDTV standard. Only by an accident of timing did the government adopt a digital HDTV standard, rather than an analog standard which would have been immediately obsolete.”); see also NICHOLAS NEGROPONTE, *BEING DIGITAL* 37-40 (1995); Joseph Farrell & Carl Shapiro, *Standard Setting in High Definition Television*, 8 BROOKINGS PAPERS ON ECON. ACTIVITY 1 (1992).

290. See OFF. TECH. ASSESSMENT, U.S. CONG., *GLOBAL STANDARDS: BUILDING BLOCKS FOR THE FUTURE 90-94* (1992) (discussing the growth of international standardization efforts and the effect of multinational organizations), available at http://www.www.princeton.edu/cgi-bin/byteserv.pr1/~ota/disk1/1992/9220_n.html (last visited Dec. 4, 2003); LIBICKI, *supra* note 3, at 19, 341-42 (discussing the weakening of governments role in the midst of globalization of IT standard-setting activities).

standard-setting.²⁹¹ This potential development serves as yet another justification to support a more indirect role for the government.²⁹²

2. *Indirect Intervention: Roles of Government Regulation*

Historically, the FCC has avoided direct standardization of both Internet telecommunications services and application standards. Accordingly, the Global E-Commerce Report (“E-Commerce Report”) contains strong language concerning the proper *de facto* production of technological standards for electronic commerce.²⁹³ Although overly broad, the E-Commerce Report takes a strong position against direct government standard-setting in its section on technical standards.²⁹⁴ The E-Commerce Report states that industry groups rather than individual companies should set standards.²⁹⁵ In practice, Internet telecommunications services and application standard-setting are largely constrained by market forces, backed by general intellectual property and antitrust law.²⁹⁶ This policy should leave the government with only indirect means of standardization. Broadly, there are two governmental roles in application standard-setting. First, to regulate supervisory rules, which facilitate production of standards by market agents. Second, to regulate processes of standardization aimed at further legitimizing both decision-making and its outcomes, as will be described below.

Conceptually, there is a broad range of governmental regulative approaches to promote the production of standards. The approaches differ in extent and directness vis-à-vis the type of each technological standard. As to extent, government intervention can be promoted narrowly, for example through specific preexisting regulations such as network non-bundling of the services and facilities involved in providing advanced capability. Alternatively, government intervention could be designed widely to promote completely new regulations intended to increase competition and invest-

291. See, e.g., Linda Garcia, *A New Role for Government in Standard Setting?*, 1 STANDARDVIEW 2, 5 (1993) (suggesting that the United States may also have considerably less influence than in the past in determining the character of international standards institutions).

292. See, e.g., David C. Wood, *European Standardization Policy*, 3 STANDARDVIEW 112, 114 (1995) (suggesting that “[c]oncerned economic players should participate directly in the standardization process, without national coordination or representation”).

293. *Id.*

294. *Id.* (“The United States considers it unwise and unnecessary for governments to mandate standards for electronic commerce.”).

295. *Id.* (“We urge industry driven multilateral fora to consider technical standards in this area.”).

296. See BENJAMIN ET AL., *supra* note 63, at 915.

ment in superior services. As to directness, government intervention can be indirect, thus letting markets operate without any new specific regulation, typically with the verification of a “reasonable” procedure made in deployment. Alternatively, a government can engage in deregulating services from broad regulation, thus improving investment incentives and technological development. Then, as mentioned earlier, there is the remaining alternative, namely direct technological standardization, which would be less efficient, and is generally limited due to the restrictions discussed above.

An even more extreme scenario is the risk of a government—acting on its own or in concert with other national governments or industry members—taking over the standard-setting process by means of monopolization. It could then convene federal advisory committees for the purpose of obtaining organized industry and user input, as the U.S. government did in the sensitive case of security standards.²⁹⁷ Still, for the reasons mentioned government standard-setting intervention of this nature is routinely seen, by both the U.S. government and its critics, as the ultimate inefficient bottleneck.²⁹⁸ Thus far, the Clinton administration policy against such government intervention in application standard-setting is still valid and has not been replaced.²⁹⁹

a) Regulate Production Supervision Rules

To facilitate market production and its funding, a government should engage in three activities. First, encourage R&D activities by autonomous institutions, preferably through better-coordinated industries rather than *de facto* agents whenever such activity would not serve to duplicate *de facto* R&D initiatives. Against the backdrop of shortened technological lifecy-

297. See, e.g., LESSIG, *supra* note 8, at 35-36 (criticizing governmental code control of the use of encryption in cyberspace); A. Michael Froomkin, *The Metaphor Is the Key: Cryptography, the Clipper Chip, and the Constitution*, 143 U. PA. L. REV. 709, 712, 718-35 (1995) (describing analogous government access and regulation possibilities); Lemley, *Standard-Setting*, *supra* note 94, at 478 (criticizing the U.S. government for dictating an encryption policy).

298. Referring generally to the Internet at large, the Telecommunications Act states that 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,” and the FCC has a responsibility to implement that statute. See Telecommunications Act of 1996, 47 U.S.C. § 151 (2000); see also Clinton & Gore, *A Framework*, *supra* note 14, § 9, at 20; *The Emerging Digital Economy*, *supra* note 233; Rubin, *supra* note 271, at 1455 (concluding that, at present, the government does not want to undertake the task, private groups do not want government intrusion, and no one thinks government will develop the optimal standards).

299. Clinton & Gore, *A Framework*, *supra* note 14.

cles, autonomous self-interested standard setters, when left alone, are typically less willing to risk potential exposure to failure. This reality is what arguably makes governments most suitable for providing R&D funding, rather than forcing industry to internalize some of these development sunk costs.³⁰⁰ In addition, funds should be available to academic institutions and even small and medium sized enterprises (“SMEs”) to encourage these entities to participate in socially undervalued research activities.

A second production supervision rule is to encourage and maintain industrial competition by applying antitrust law against potential anticompetitive strategic behavior. Thus, the general antitrust rule should be applied for anticompetitive research or production co-ventures. Antitrust law should impose liability under section 1 of the Sherman Act and section 7 of the Clayton Act whenever collaborative standard-setting bodies erect barriers to entry to prevent new innovations from entering the market.³⁰¹

A third production supervision rule is indirect price regulation, whenever market competition curtails socially desirable variety in application standards.³⁰² Such price controls should be used to lower prices of otherwise undervalued social welfare benefits.³⁰³ In cyberspace, such initiatives included real-world analogous non-profit activities as research and education.³⁰⁴ In some cases such valuable benefits can be technologically inefficient. Similarly, for the purpose of preserving socially desired technologies, indirect government intervention would be essential so as to benefit low performance-sensitive applications such as remote retrieval software for text files, which benefit clients that otherwise might lose their incen-

300. See BREYER, *supra* note 1, at 10; NEXT GENERATION INTERNET, *supra* note 173, at 5.

301. See Sean P. Gates, *Standards, Innovation, and Antitrust: Integrating Innovation Concerns into the Analysis of Collaborative Standard-Setting*, 47 EMORY L.J. 583, 654-55 (1998).

302. For the U.S. government’s declaration of support for this policy, see AGENDA FOR ACTION, *supra* note 114.

303. See OGUS, *supra* note 1, at 153.

304. Accordingly, during its first seven months, the Clinton-Gore Administration followed this policy. Initially, the President’s FY 1994 budget included a \$100 million program to develop applications in areas such as education, manufacturing, health, and digital libraries. In addition, the U.S. government allocated \$50 million for National Telecommunications & Information Administration (“NTIA”) grants to demonstrate the applications of the NII for non-profit institutions such as schools, hospitals, and libraries. See *Administration NII Accomplishments* § 3, <http://www.ibiblio.org/nii/NII-Accomplishments.html> (last visited Dec. 3, 2003); see also H.R. 5983, 102d Cong. (1992); H.R. 5759, 102d Cong. (1992); S. 2937, 102d Cong. (1992) (seeking to develop high-speed technological infrastructure for schools, libraries, medicine etc.); S. 2813, 102d Cong. (1992) (seeking to facilitate public access to federal electronic information).

tive to reveal their preferences. Such inefficiency might result in the erosion of the qualitative rationale of standardization, over-emphasizing narrow performance-sensitive applications.³⁰⁵ This important supervisory rule should not be limited only to confronting strict economic productivity, but potentially technological efficiency as well. As explained, industry standard setters may not necessarily be *technologically* neutral—making government intervention in the latter not only important, but also ultimately irreplaceable.

b) Regulate the Process of Standardization

Inherent to commercial production of high technologies is the tension between efficient and fast evolving outcomes and slow and thus constraining procedures. With application standards, this difficulty is primarily the result of constant outside commercial pressure. As a result, even formal standard-setting bodies tend to focus on outcomes, as opposed to processes.³⁰⁶ Thus, as a comparative conclusion, March and Olson see the main role of fairly elected political institutions as formal standardization institutions, as best suited to confirm the legitimacy of these same undervalued procedural decisions.³⁰⁷ Translating this conclusion into practice, governments should emphasize regulating standardization processes next to technically standardizing products.³⁰⁸ As the ultimate guardians of the standards producing process, governments would thus promote the dynamics that are likely to satisfy the demand of fairness by the process used in their creation.³⁰⁹

305. Scott Shenker, *Service Models and Pricing Policies for an Integrated Services Internet*, in PUBLIC ACCESS TO THE INTERNET 326 (Brian Kahin & James Keller eds., 1995).

306. Egyedi, *supra* note 3, at 52 (“The formal standards bodies are, however, inclined to stress ‘outcome’ rather than ‘process’ results because of outside pressure.”); Joel West, *Institutional Constraints in the Initial Deployment of Cellular Telephone Service on Three Continents*, in INFORMATION TECHNOLOGY STANDARDS AND STANDARDIZATION: A GLOBAL PERSPECTIVE 198-221 (Kai Jakobs ed., 1999) (including a comparative empirical study on such external institutional pressures in the deployment of analog cellular telephone service standards, as experienced in the United States, Japan and Europe). Such outside pressure is already argued to stand in the way of ICANN, see, for example, Post, *Governing Cyberspace*, *supra* note 202 (“Any entity exercising control over the DNS will be subject to immense pressure to do more than mere ‘technical management.’”).

307. MARCH & OLSEN, *supra* note 13, at 50-52;

308. See Anton & Yao, *supra* note 169, at 248 (“[w]hile informed judgments about such procedures will still be necessary, technical judgments will be avoided”); Rubin, *supra* note 271, at 1473.

309. For a discussion on elements important in determining reasonableness of procedures in the face of potential anticompetitive effects of standard-setting, see Michael C. McCarey, Associate Director, FTC Bureau of Consumer Protection, *Industry Standards*

As indirect players, this role of governments may be achieved in different ways. First, by supervising formal industry initiatives. Beyond economic efficiency, governments would impose and encourage formal industries to maintain a democratic layer between the autonomous standardization institutions and their participants, and the standards that they design. In essence, as guardians of the process, standardization procedures would then serve, indirectly, to legitimize that same process. Second, by defining its own alternative procedures.³¹⁰ In either way, most participants in designing a standard should be made to benefit from the result.³¹¹ This role of government may then enhance the level of competitive participation in standardization itself,³¹² particularly participation of marginal participants.³¹³ The National Technology Transfer and Advancement Act of 1995³¹⁴ directed federal agencies to use standards developed by voluntary consensus bodies, except where doing so would be inconsistent with applicable law or when it would be otherwise impractical.³¹⁵ This legislation adds credence to the earlier U.S. government policy codified in the OMB Circular A-119 *Federal Participation in the Development and Use of Voluntary Standards* ("OMB Circular A-119").³¹⁶ Proposed revisions to OMB Circular A-119, characterize the nature of standards that government agencies are constrained to consider ahead of developing their own procurement or technological standards.³¹⁷ Accordingly, agencies will be submitted to an administrative burden of proof to justify why they did not

and Certification: Three Current Issues, Remarks Before the 26th Annual Symposium of the Trade Ass'n & Antitrust Law Comm. of the D.C. Bar (Feb. 1990).

310. *Id.*

311. Egyedi, *supra* note 3, at 52.

312. *Cf.* Balto, *supra* note 169, § 2 & nn.13-19 (adding that, in practice, "the antitrust jurisprudence on standard-setting focuses almost entirely on collective standard-setting and the process used to determine the standards").

313. CARGILL, *supra* note 1, at 233-34 (discussing the diversity of participants in autonomous standardization institutions both as an advantage and as a problem in reaching consensus).

314. National Technology Transfer and Advancement Act of 1995, Pub. L. No. 104-113, 110 Stat. 775 (1996)

315. *See* Rubin, *supra* note 271, at 1473 (suggesting that it would be possible for a federal statute to declare that the use of a computer language in interstate commerce that was not designed through a sufficiently cooperative process is an unfair trade practice under 15 U.S.C § 45(A)(1)). So far, however, this suggestion has not been adopted.

316. The OMB has been the guardian of the previous policy and is responsible for the more detailed regulations that all agencies will have to follow to meet the new law.

317. *See* OMB Circular No. A-119, *supra* note 30.

use a relevant “Voluntary Consensus Standard” prior to proceeding to design new government standards.³¹⁸

C. A Potential Deviation: The ICANN Case Study

One potential deviation from the supported general policy for infrastructure standard-setting may arise in the form of the federal government’s involvement with the ICANN domain name system. ICANN, a private non-profit corporation, was initially appointed to oversee the operation of the root server system. In this capacity, it was charged with supporting existing protocols and telecommunications services used to implement domain name facilities.³¹⁹ To that end, ICANN’s board of directors had two different functions. The first was to take steps towards introducing competition into the Domain Name registration system.³²⁰ The second was to uphold a policy against cybersquatting through what later became the Uniform Dispute Resolution Policy (“UDRP”) and arbitration panel.³²¹

ICANN’s establishment suggested two types of standardization concerns. The first concern regarded its potentially problematic institutional identity, the second was its wide technical mandate as a standardization organization. As a technical standardization institution, ICANN was initially constructed as a private interest group. As a private entity it exercised direct and central control with the U.S. government choosing to remain in the background.³²² As suggested earlier through public choice analysis, left alone such competitive interest groups might establish genuine public policy inefficiency. In addition, such groups might create a monopoly on the allocation of the DNS names and numbers.³²³ These latter

318. With a decrease in the governmental role in cyber standard-setting, there is some criticism of this policy as overly national in nature, in the midst of a global standardization trend. See Garcia, *supra* note 291 (for a critical analysis of OMB Circular A-119 policy, as insufficiently coordinating); Wood, *supra* note 292, at 114.

319. See RFC 882, *supra* note 77.

320. See White Paper, *supra* note 188, at 31, 749.

321. *Id.* at 31, 747. That policy was later adopted in the Uniform Domain Name Dispute Resolution Policy. See *Uniform Domain Name Dispute Resolution Policy* (Aug. 26, 1999), at <http://www.icann.org/dndr/udrp/policy.htm> (last visited Dec. 1, 2002).

322. See BENJAMIN ET AL., *supra* note 63, at 825.

323. See generally Froomkin, *supra* note 145, at 216 (suggesting that the analyses of the privatization of the DNS and TCP/IP highlights some of the reasons why the bottom-up process has failed); A. Michael Froomkin & Mark A. Lemley, *ICANN and Antitrust*, 2003 U. ILL. L. REV. 1 (discussing the various potentially anticompetitive effects of ICANN); Kesan & Shah, *supra* note 115, at 214 (suggesting that the Internet community was not able to resolve the uniformity problem through a bottom-up process, and, as a result, the U.S. government has begun to intervene).

challenges are not merely structural as they often interact with ICANN's unique technological mandate.

Thus, the second standardization concern is a function of ICANN's technical responsibilities. Arguably, ICANN was given too much technical discretion in a blurry mandate on code writing (i.e., technological standardization). ICANN controls issues such as the maintenance of the bit size of data packets, the architecture of the root services, and the preservation of unique protocol numbers for other various Internet functions.

Originally, the Department of Commerce ("DoC") initiated the *White Paper*, a non-binding statement of policy through which it informally declared that "the U.S. government should end its role in the Internet number and name address system."³²⁴ Like the *Green Paper* before it,³²⁵ the *White Paper* has conformed to the already existing vague and basic governmental "Principles for a New System"³²⁶ with no clear separation between a technical standardization policy and non-technical governance responsibilities.³²⁷ Accordingly, the DoC characterized ICANN's technical responsibility in vague terms. This new corporation was made responsible only for "technical management of the DNS," which was most likely defined as the "narrow management and administration of Internet names and numbers on an ongoing basis." Overall, most commentators still agree that the U.S. government holds *de facto* control of the root zone.³²⁸ However, it is also clear that the U.S. government has chosen not to have direct control over the root server.³²⁹ Consequently, the question of ICANN's technological standardization has still not been raised properly as a separate policy question, underscoring the need to decide on future technological risks.

324. See *White Paper*, *supra* note 188, at 31, 749.

325. NAT'L TELECOMM. & INFO. ADMIN., U.S. DEPT. COMMERCE A PROPOSAL TO IMPROVE TO TECHNICAL MANAGEMENT OF INTERNET AND ADDRESSES (1998) [hereinafter GREEN PAPER].

326. These are principles such as "stability, competition, private bottom-up coordination, and representation." *Id.*

327. See Froomkin, *supra* note 145, at 171 (suggesting that the DoC draw one henceforth).

328. See *id.* at 166 ("[T]here is no dispute that the U.S. government, through the Department of Commerce, currently enjoys *de facto* control of the DNS. Nor is there any dispute that DoC has at least temporarily ceded to ICANN, through a variety of contractual and quasi-contractual agreements, almost all the control the United States enjoys."); Steve Kettmann, *Will U.S. Release Grip on ICANN?* (Jan. 19, 2002) (emphasizing that post September 11 the gradual process by which ICANN will gain autonomy from the government has been slowed), <http://www.wired.com/news/infrastructure/0,1377,49836,00.html> (last visited Dec. 1, 2002).

329. *But see* Froomkin, *supra* note 145, at 169 (suggesting that for that reason the U.S. government had, in fact, only quasi-privatized the control on the root server).

To date, both ICANN and the DoC deny that ICANN is engaged in either regulation or governance. Instead, they suggest that ICANN is engaged in nothing more than routine standard-setting or presumably “technical coordination” or “maintenance.”³³⁰ Instead, the main controversy over ICANN’s governance mandate was limited to the question of its democratic decision-making accountability. Thus, although the DoC seems to be keeping only an indirect measure of control over the DNS and its root server, it is consistently doing so from a distance based on a rhetoric that pretends to support ICANN’s absolute self-regulation. However, with the mounting political pressures against the U.S. government’s intervention on the one hand, and commercial interests trying to “capture” ICANN’s wide technological discretion on the other, there are no guarantees that such a delicate, yet relatively successful balance can be maintained in the future. The need to balance these technological risks is not merely theoretical. Present infrastructure transparency concerns are already a good case in point for that.

In recent years, as part of ICANN’s concern with preserving end-to-end transparency, it has become involved with the various issues surrounding internationalized domain name (“IDN”) compulsory infrastructure standards. Thus, ICANN’s Board has begun to initiate inquiries about that role with regard to the various efforts to use non-ASCII characters to design international domain names supported by the domain name system at large.³³¹ On September 25, 2000, the ICANN Board approved a set of resolutions,³³² in which the Board recognized the importance of the Internet evolving to be more accessible to those who do not use the ASCII character set.³³³ Ultimately, ICANN recognized a need to specify an adequate standards track protocol based on supporting test bed findings and requirements. Upon final adoption in June 2003, IDNs became fully operational in a standards-based way.³³⁴ Consistent with ICANN’s policy,

330. See White Paper, *supra* note 188, at 31, 744.

331. The American Standard Code for Information Interchange (“ASCII”) is the most common standard for the code numbers used by computers to represent all the upper and lower-case Latin letters, numbers, punctuation, etc. There are 128 standard ASCII codes.

332. Formally relating to the 22 August Verisign Global Registry Services announcement about its introduction of the multilingual test bed.

333. See ICANN, *Minutes of Special Meeting of the Board* (Sept. 25, 2000) [hereinafter ICANN, *Minutes*], at <http://www.icann.org/minutes/minutes25sep00.htm#MultilingualDomainNames> (last visited Dec. 3, 2003).

334. See ICANN, *Guidelines for the Implementation of Internationalized Domain Names* (June 20, 2003) [hereinafter *Domain Name Guidelines*], at <http://www.icann.org/general/idn-guidelines-20jun03.htm> (last visited Dec. 3, 2003). With IDN, *de facto* designers of browsers or other Internet software would be able to program their software to convert any foreign-character domains typed in or linked to into the appropriately

the accepted standard would then have to be binding and fully compatible with the Internet's existing end-to-end model, and "preserve globally unique naming in a universally resolvable public name space."³³⁵

As the specially designed Internationalized Domain Names Committee has suggested, any Top Level Domain ("TLD") expansion should occur in a careful and controlled fashion, with regard for the overall stability of the DNS. On balance, as long as the DNS is subject to the present pre-designed scarcity policy that stability will be achieved within the limit of the total number of TLDs eligible for delegation to a given geographic unit.³³⁶ IDNs should, therefore be carefully and agreeably set at a number equal to the number of its official languages.

In part, this technological challenge was met successfully. However, it took more than a quasi-privatized ICANN to do so, as ICANN asked for the legitimacy and intervention of ISO. In fact, ICANN/IANA is still staying out of the business of determining what is and is not a country (or geographically distinct territory), and what name or abbreviation is semantically associated with any given geographic unit. Technically, the ISO-3166-1 IDS table developers at ISO, already solved the problem of what is and is not a recognized geographic unit (country or geographically distinct territory).³³⁷ However, it being a sensitive, politically oriented decision, the table only provides two- and three-letter ASCII codes for each such geographic unit. Thus, ISO's table does *not* solve the multi-faceted problem of what non-ASCII names (or abbreviations) should be assigned to each recognized geographic unit, and who should be in charge of assigning them.

In essence, the current ICANN/IANA policy permits the delegation of ASCII ccTLDs only when a given geographic unit and its associated specific 2-letter ASCII codes appear on the ISO 3166-1 list. Due to the significant political nature of this question, ICANN/IANA's policy has so far failed to authorize the use of non-ASCII characters, leaving ICANN without a given reference point for IDNs. On balance, those problems are properly excluded from the ICANN process, and resolved by a politically

convert any foreign-character domains typed in or linked to into the appropriately coded string, which could then be resolved using normal DNS queries.

335. See ICANN, *Minutes*, *supra* note 333.

336. See discussion *supra* Part III.B.2.a.

337. For most users of ISO 3166-1 the standard is the list of country names and codes. See Internet Standards Organization, *Background on ISO 3166*, <http://www.iso.org/iso/en/prods-services/iso3166ma/04background-on-iso-3166/index.html> (last visited Dec. 3, 2003).

expert, internationally recognized body, the ISO and its ISO 3166 Maintenance Agency.

ICANN's Board decided to further delegate to top-level domain registries the employment of an "inclusion-based" approach meaning that code points that are not explicitly permitted by the registry would be prohibited.³³⁸ However, in compliance with ICANN's IDN guidelines, registries are merely encouraged to work collaboratively with relevant and interested stakeholders to develop language-specific registration policies in a way that might then curtail ICANN's new main goal of privatizing today's DNS universal uniqueness in case such cooperation gets complicated.³³⁹ Left to the strategically dominant RIRs and nation-agents, in the "market of nations," ICANN/IANA's non-binding hegemony might be facing self-interested competing claims backed by different regional stakeholders, or worse, different national governments. As a one-shot game, achieving such consensus would be subject to a basic prisoner's dilemma flaw, where no efficient incentive to cooperate would prevail and parties would choose to act strategically: left alone, national registries still have the incentive to benefit their own customers at the expense of the DNS stability at large. Thus, maintaining such stability in case of lack of cooperation by RIRs, in the face of growing self-interested commercial intervention, would be a task poorly suited to a technical coordinating organization such as ICANN. Arguably, the non-binding quasi-privatized ICANN is now facing a set of political concerns for which it might not be well suited. Furthermore, ICANN might misuse its mandate of deciding when and to what neutral and authoritative arbiter such problems should be referred, thus risking potential DNS instability, politically, and ultimately, technologically.

In such critical situations, as the broadly agreed lowest common denominator rule, ICANN should attempt to enforce binding standardization policies only when there is a clear need for uniformity based on a substantive consensus among those who must implement such policies and are impacted by them.³⁴⁰ But should ICANN be faced with full responsibility

338. See *Internet Domain Name Guidelines*, *supra* note 334, at 2.

339. *Id.* at 4. Few recent examples of such complications evolved when lucrative domain-name businesses, such as VeriSign, pushed forward with their own IDN initiatives. Complicating matters even more, the Chinese government has laid claim to control over Chinese-language domain names, etc. See Greg Goth, *Engineering Complexities*, 5 IEEE INTERNET COMPUTING 8 (May 2001), <http://www.acm.org/technews/articles/2001-3/0601f.html> (last visited Oct. 23, 2003).

340. See David R. Johnson & Susan P. Crawford, *The Idea of ICANN*, at http://www.icannwatch.org/archive/the_idea_of_icann.htm (last visited Dec. 3, 2003).

for the challenge in the first place? With potentially suboptimal level of agreement on the need for DNS uniformity, as in the emerging case of IDNs, ICANN might be arguably approaching its institutional limit. Even assuming that in the long run, both economically and technically, multiple language domain names are feasible in the short run, ICANN may still have to coordinate ad hoc undesired fragmentation that might weaken the stability of the DNS, and encourage destructive collusion between name owners. Inevitably, as a policy rule, ICANN's Board may have to be more steadily backed by more authoritative agents, namely formal industry standardization organizations as with the case of ISO's 3166 Maintenance Agency and in extreme scenarios of loss of DNS hegemony, even more notably by the DoC, and the U.S. government at large; thus, overriding the US government's rhetorical policy against intervention in coordinating infrastructure standardization and its risky policy (or otherwise pretended) to support ICANN's final privatization.

Here, as potentially elsewhere, seen narrowly as mere technological standard-setting activity, no governmental guidelines were adequately established for ICANN thereby undermining the necessity for a visible and continuous technological standardization policy with respect to infrastructure standardization. Left as a technologically independent, non-binding, quasi-privatized standard-setting organization that is charged with public responsibility, an unmonitored ICANN may design or adapt standards unproductively.

VI. CONCLUSIONS

In the future, changes in cyberspace will bring about new innovative developments as part of new technological generations.³⁴¹ Both the TCP/IP suite and the Internet as a whole will continue to be standardized,

341. See NEXT GENERATION INTERNET, *supra* note 173. The Next Generation Internet was a U.S. federal initiative including experts from business, government and academia, trying to anticipate the next generation of Internet standardized applications. *Id.* It was a three-year program, which started in 1996 with a \$300M budget divided among several government agencies (with the lead role going to DARPA). *Id.* The program involved a test network with 100 sites that were linked at a speed 1,000 times greater than today for the design of revolutionary applications. *Id.* Another program, the Internet2 K20 Initiative, led by over 200 universities working in partnership with industry and government, is focused on developing an advanced network and applications. See *Internet2 K20 Initiative Website*, at <http://k20.internet2.edu/index.php> (last visited Dec. 3, 2003). Together with the NGI initiative, these two private networks may be integrated into the present Internet or remain separate. *Id.*

and that standard-setting will continue to shape new and existing cyber markets. New protocols will be designed and old ones will be revised.³⁴²

With the growth in both the community of users and the demand for sophisticated applications, a more advanced standardized architecture is already needed. New standardization challenges are already here—such as IPv6, designed and coordinated by the IETF formal standardization organization to expand address space.³⁴³ Similarly, several external trends and influences such as the deployment of wireless network technologies, mobile-networked devices and special purpose IP devices will have a large impact on the status of the infrastructure network layer.³⁴⁴ Leaving aside the question of whether these specific developments will lead to a generation leap, any adoption of such central technologies should follow the by and large positive experience of cyber standardization exemplified by the last two decades.

As for infrastructure standards, and notwithstanding strong governmental rhetoric concerning the need for regulative restraint and ICANN's potential inconsistency, this study generally supports the rationalization of the early central institutional adoption of a unified infrastructure set of standards for inter-connective transmission, as in the case of TCP/IP. Justifiably, this early endeavor was not followed by private initiatives of creating a market for infrastructure standards for inter-connectivity. Instead, only a market of basic telecommunications services evolved with the involvement of diverse infrastructure equipment providers including data networking equipment, Internet connections, telecommunications equipment providers, and cable operators. As described, a market for application standards also subsequently developed. In essence, even with the later creation of the market for backbone telecommunications services, a common stable denominator was preserved in the face of a governmental inter-connective TCP/IP naturally monopolized standard. In addition, in this

342. See Hanseth & Monterio, *supra* note 3, at 174 (explaining that new generations of infrastructure evolve by combining, extending and aligning existing infrastructure).

343. For a relevant early warning about such a possibility, see Bayh-Dole University and Small Business Patent Act of 1980 § 63, 15 U.S.C. § 3701(8) (2000).

344. See Peter Brockmann, *User Demand for Internet Services: Is the Infrastructure Ready?*, 20 COMPUTER STANDARDS AND INTERFACES 117-21 (1998) (for a broader perspective on potential infrastructure trends); Kaat, *supra* note 188, § 1; see also Lucent Technologies, *Seamless Roaming between WiFi and 3G Networks* (Mar. 2003) (describing the integration of WiFi and Third-Generation (“3G”) mobile networks to create comprehensive solutions with seamless roaming), <http://itpapers.com/abstract.aspx?kw=3g+&docid=47696> (last visited Dec. 3, 2003); Syntel, *Wireless Applications: A Natural Evolution* (Oct. 2001), <http://itpapers.com/abstract.aspx?kw=wireless&docid=16150> (last visited Dec. 3, 2003).

phase a consensual architecture was formally mandated, efficiently overshadowing the potential inter-institutional infrastructure standardization arms race.

In retrospect, only an *ex ante* governmental standard-setting initiative delegated through its early federal agencies and followed by monitored activity through research institutions was inherently efficient in the early infrastructure development phase. In this development phase of poor price-competition, the only exception was gradually giving away much of the government's power over the early market for carrying and access services, as was mostly done in the early 1990's. In essence, these infrastructure standard-setting activities were primarily technically, rather than commercially, motivated. That policy eventually changed with its move towards commercialization of application and complementary standardized products. These central changes led to substantial competition in standard-setting activity in cyberspace. In the future, such changes will continue to occur whenever such technological and economic developments take place, as part of future intermediate modification phases of different standardized technologies.

Later, the optimal institutional choice changes in the implementation phase. In this phase of application and complementary standardization, political institutions like the U.S. government through its delegated agencies, particularly the FCC, should stick to a restrained indirect role in standardization activity due to its institutional barriers to efficient participation. The government and its agencies may continue to have a role in the form of limited indirect support for infrastructure routine standard-setting activity, for example promoting increase in bandwidth on the backbone transmissions links and better physical access from homes and businesses. As a general matter, such a policy should also facilitate essential competition among autonomous standard-setting institutions. In practice, while the conduit has not always been lucid, nor followed a single fixed route, the governmental understanding of the public interest in the United States reveals a positive and definite prototype of declining centralized interference, notwithstanding an increasing number of such institutional sources.

Of special importance in this commercial environment is the role of autonomous industry institutions. New risks of *ex ante* technological stagnation and/or *ex post* anticompetitive effects mean that a role for a voluntary industry regulative approach in formalizing *gray* and *de facto* standardization will become essential. In later technological implementation phases, industry efficiently chills direct governmental incentives for intervention beforehand or *ex post*. This is subject only to indirect governmental supervision rules, which facilitate market production of standards and

procedural regulative intervention. Although many of these policies are upheld in practice, so far neither the U.S. government nor the FCC have articulated sufficiently clear or comprehensive policy on the matter. Thus far, this has left institutional choice in the Internet's standardization subject to overly general principles of marketplace competition, made to assure "reliability, interoperability, ease of use and scalability."³⁴⁵

For specific infrastructure standardization, potential risks of deviation from the government's justified proactive practices, already suggests policy conformity. One important case study for such potential digression came in the form of federal governmental involvement with ICANN. Although the DoC seems to maintain an oblique measure of control over the DNS and its root server, it is consistently doing so from a distance, almost in denial, based on a rhetoric that seems to pretend support for ICANN's final privatization. However, with the mounting political pressures against the U.S. government's intervention on the one hand, its attempt to "capture" ICANN's wide technological discretion on the other, there are no guarantees that such delicate, and relatively successful balance would be maintained in the future. The ICANN case study demonstrates that the question of technological infrastructure standardization was not raised properly *ex ante* as a policy question. This undermined the ability to confront real technological threats, such as the fragmentation of the network layer and root splitting. Such policy made no adequate division between infrastructure standards and application standards for matters of regulative intervention, hence providing ICANN, as a potentially self-interested interest group, overly broad control over both, but especially over the former.

Indeed, ICANN's technical mandate reaches potentially much further than is literally understood from existing formal documentation. Seen narrowly as mere technological routine standard-setting through technological "maintenance," no adequate governmental guidelines were put in place, thus undermining the necessity for a stable and preferably open standardization policy.

As with analogous standardization regimes, there is the risk that unless the distinctive standardization policies set forth in this study are viewed *en bloc*, and thus sequentially and contextually, cyberspace's largely successful institutional practice might not be preserved already for the present ICANN case study and to provide future benefit.

345. For the specific context of standard-setting, see Clinton & Gore, *A Framework*, *supra* note 14, § 9.

OF PATENTS AND PATH DEPENDENCY: A COMMENT ON BURK AND LEMLEY

By R. Polk Wagner[†]

ABSTRACT

This Article delves into issues surrounding the relationship between technology and the patent law. Responding to Dan Burk and Mark Lemley's earlier article, *Is Patent Law Technology-Specific?*, the piece notes that the basic question posed by Burk and Lemley's article is a relatively easy question given the several doctrines that explicitly link the subject matter context of an invention to the validity and scope of related patents. This sort of technological exceptionalism (which this Article refers to as micro-exceptionalism) is both observable and easily justifiable for a legal regime directed to technology policy. In contrast, Burk and Lemley's identification of, and advocacy for, a broader sort of exceptionalism (macro-exceptionalism) is far more troublesome, implying a role for the patent judiciary in rather detailed policy judgments, for example the optimal breadth for biotechnological or software-based patents. The Article offers a variety of reasons that macro-exceptionalism is unwarranted, and indeed, notes that a primary claim of Burk and Lemley's—that the Federal Circuit has grossly missed the mark in its (purportedly) exceptionalist approach—previews the sort of problems created by pursuing technological exceptionalism in the patent law.

TABLE OF CONTENTS

I.	INTRODUCTION	1342
II.	PATENTS VS. TECHNOLOGY	1344
	A. Macro and Micro: The Two Forms of Technological Specificity	1345
	B. The Indeterminate Scope-Effects of the PHOSITA	1348
III.	THE UNCERTAIN PATH-DEPENDENCIES OF THE PATENT LAW	1350
	A. The Dogs That Don't Bark: The Missing Evidence of Macro-Specificity ..	1351
	B. The Factual Nature of the Federal Circuit's Technological Exceptionalism	1354
	C. A Response to Burk and Lemley.....	1355
IV.	OF COASE AND COMPLEXITY: FIXING THE PATENT DOCTRINE.....	1356

© 2003 R. Polk Wagner

[†] Assistant Professor, University of Pennsylvania Law School. Thanks to Arie Michelsohn, Jeff Leftsin, Lee Petherbridge, Arti Rai, Reed Shuldiner, and participants at the Washington University Human Genome Conference and the BIO 2003 Conference for helpful comments on earlier drafts. Patrick Mirville provided research assistance. Comments appreciated: polk@law.upenn.edu. See <http://papers.pennlaw.net/> for updates and other work.

I. INTRODUCTION

In an era of accelerating technological developments, the ballooning use of patents as both swords and shields in the marketplace, and growing criticism of courts' ability to meaningfully address the challenges posed by the complex regulatory framework that is the U.S. patent system, one might expect reform-minded observers to eschew proposals based on the unbounded intervention of judges into broad, innovation policy-based analyses. Professors Dan Burk and Mark Lemley, however, boldly advocate just such a program in *Is Patent Law Technology-Specific?*¹ Inspired in large part by what they see as an emerging technological exceptionalism within the patent law, Burk and Lemley suggest important doctrinal changes that would allow judges far more latitude to establish industry-specific patent rules—a result that, they argue, will reverse the “exactly backwards” course the Federal Circuit has traveled to date.

This Article offers a different approach. To be sure, *Is Patent Law Technology-Specific?* (and its related works) marks an important and insightful contribution to the growing literature on the institutional relationships of the patent law.² And yet, in the pages that follow, I suggest an alternative view of Burk and Lemley's findings—specifically, that their exposition makes a rather compelling case *against* precisely the sort of judicial ventures into technologically-specific innovation policy that they recommend. Instead, their examples of the ongoing struggle to adapt the patent law to technological changes illuminate the undesirability of entangling the patent doctrine in broad, policy-driven technological exceptionalism. As befits an expansive regulatory regime concerned with innovation policy, the patent law is inextricably intertwined with the process and de-

1. Dan L. Burk & Mark A. Lemley, *Is Patent Law Technology-Specific?*, 17 BERKELEY TECH. L.J. 1155 (2002) [hereinafter Burk & Lemley, *Technology-Specific*]. Note that Burk & Lemley, *Technology-Specific* and this Article are part of a group of related works. See Dan L. Burk & Mark A. Lemley, *Biotechnology's Uncertainty Principle*, in F. SCOTT KIEFF, PERSPECTIVES ON PROPERTIES OF THE HUMAN GENOME PROJECT (forthcoming 2003) [hereinafter Burk & Lemley, *Uncertainty Principle*]; Dan L. Burk & Mark A. Lemley, *Policy Levers in Patent Law*, 89 VA. L. REV. 1575 (2003) [hereinafter, Burk & Lemley, *Policy-Levers*]; R. Polk Wagner, *(Mostly) Against Exceptionalism*, in F. SCOTT KIEFF, PERSPECTIVES ON PROPERTIES OF THE HUMAN GENOME PROJECT (forthcoming 2003).

2. For recent related work, see, for example, Burk & Lemley, *Technology-Specific*, *supra* note 1; Arti Kaur Rai, *Facts, Law & Policy: An Allocation of Powers Approach to Patent System Reform*, 103 COLUM. L. REV. 1035 (2003) [hereinafter Rai, *Facts, Law & Policy*]; F. Scott Kieff, *The Case for Registering Patents and the Law and Economics of Present Patent-Obtaining Rules*, 84 B.U. L. REV. (forthcoming 2004); R. Polk Wagner, *Reconsidering Estoppel: Patent Administration & the Failure of Festo*, 151 U. PA. L. REV. 159 (2002).

tails of technological development. As courts and commentators alike have long recognized, both a challenge and strength of our patent system is the ongoing effort to adapt the legal infrastructure to an ever-changing environment.³ The patent law—by explicit design—is technologically flexible, with significant adjustment points built into the system. That distinctions in treatment will exist between various technologies is both expected and unremarkable; rather than leveraging these differences for policy effect, the goal should be to embrace the flexibility while retaining the essential strengths of the unified patent system.

The argument proceeds in three parts. Part II explains that while Burk and Lemley are undoubtedly correct in noting that there is technological-specificity in the patent law—that biotechnological inventions get “treated differently” than, say, software or mechanical inventions—this observation alone is certainly no cause for alarm. Submerged in the Burk and Lemley analysis is an important conceptual distinction between two types of technological-specificity: *micro*-specificity, which applies the variable legal rules to specific technological circumstances; and *macro*-specificity, which countenances distinct legal rules across different technologies, and relatively more similar application within related technologies.⁴ Determining which of these two forms best describes modern patent jurisprudence is critically important, for this explains whether the Federal Circuit has developed (or seeks to develop) an innovation regime especially for specific industries, or whether any observable distinctions are merely the expected consequence of the patent law’s inherent flexibility.

Part III argues that micro-exceptionalism is both a more accurate description of the current patent law—as well as a normatively justifiable position. Even accepting Burk and Lemley’s analysis of the relevant case-law as correct, I argue that there remain a number of observations and explanations—such as factual misunderstandings, doctrinal confusion between facts and law, or even the unique circumstances surrounding the set of opinions that dominate biotechnology in particular—that point rather strongly in favor of a micro-specific framework rather than *macro*-exceptionalism. That is, it may well be that *Is Patent Law Technology-Specific?* has identified potentially-serious defects in the court’s jurispru-

3. See, e.g., *AT&T Corp. v. Excel Communications, Inc.*, 172 F.3d 1352, 1356 (Fed. Cir. 1999) (“As this brief review suggests, this court (and its predecessor) has struggled to make our understanding of the scope of [35 U.S.C.] § 101 responsive to the needs of the modern world.”).

4. For example, *micro*-specificity allows that invention A, in a newly-developing niche of the software field, would potentially have different application of the disclosure and obviousness standards than invention B, in a mature area of software, or invention C, in a groundbreaking biotechnological area. See *infra* Part II for more explanation.

dence as applied to this technological area. But this, I suggest, does not itself make the case that the doctrine is *macro*-specific, in part because there seems to be little reason to worry that path-dependency—the tendency of the court to continue in this direction—has emerged.

Indeed, as I argue in Part IV, the several problems with the Federal Circuit's doctrinal development identified by Burk and Lemley seem to quite strongly support the position that *macro*-exceptionalism is ultimately unjustified. *Is Patent Law Technology-Specific?* advocates substantial, policy-driven *macro*-specific changes to fundamental standards of patentability in order to adjust the breadth of patents towards their optimal level (depending upon the technology). And while this almost-Kitchian approach does seem to dominate alternatives involving the weakening of property rights in biotechnology, there is a third option—clarifying and stabilizing the patent law to reduce transaction costs—that seems to be even better. Indeed, a transaction-cost-focused analysis would suggest that it is Burk and Lemley, rather than the Federal Circuit, that have it “exactly backwards.” Given the deep uncertainties underlying the premises of Burk and Lemley's argument, as well as the promise of increasing transaction costs resulting from a shift from the *micro*-specific to *macro*-specific approach, there seem to be strong reasons to conclude that we should remain skeptical of broadening the technological-specificity of the patent law.

II. PATENTS VS. TECHNOLOGY

That the patent law is significantly “technology-specific” is both easily apparent and fully expected. As noted above, any law purporting to provide a regulatory foundation for innovation must be able to account for both the broad range of technologies and the rapid pace of change.⁵ To bind the patent law to the technological assumptions of an earlier era, or to the maturity of any particular technology, would be exceedingly foolish. And yet there is a limit: not all technological exceptionalism is benign. When the jurisprudential approach shifts from adaptation to prescription—from the application of consistent rules to variable facts to the promulgation of distinct rules to implement technology-based innovation policy—courts put at risk the very social progress they seek to enhance.⁶

5. As Burk & Lemley note, “different industries experience both innovation and the patent system in very different ways.” Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 40-41.

6. *See infra* Part IV.

A. Macro and Micro: The Two Forms of Technological Specificity

Because of the recognition that technological-specificity can be either a boon or burden to the patent system, it is critically important at the outset to determine what one means by a “technology specific” patent law. As noted above, there are two distinct conceptual schemas to consider:

Micro-Specificity: the (legal) rules applied to innovations are variable, dependent upon particular technological circumstances.

Macro-Specificity: the (legal) rules are quite distinct across different technologies—even while being relatively more similar within related technologies.⁷

To illustrate the distinction, consider three inventions, two generally in the software field, and one in biotechnology: a first deals with data management, a relatively stable and mature area of software technology; the next relates to machine learning, a relatively immature and undeveloped sub-field; and the third considers genomic research, at the high-end of biotechnology.

In a regime of micro-specificity, each of these inventions will have rather distinct patentability requirements—primarily because of the operation of the patent law’s “person having ordinary skill in the art” (“PHOSITA”) standard. As Burk and Lemley demonstrate, a higher PHOSITA standard (a greater degree of difficulty in the field) implies a lesser standard for obviousness and a greater disclosure requirement.⁸ Table 1 notes these requirements, taking the data management invention as having the lowest (baseline) difficulty level.

7. One might call this an “industry-specific” approach, but that implies an economic structure coincident within related technologies, which is perhaps—but not necessarily—the case.

8. See Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1173-82 (describing biotech jurisprudence); *id.* at 1160-73 (describing software jurisprudence). Of course, as Burk and Lemley note, the explicit coupling of the PHOSITA standard in obviousness and disclosure doctrines can obscure some small differences in the way the standard is applied to each requirement. See Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 27-29.

Table 1: The Micro-Specific Approach

Invention	PHOSITA Level	Obviousness Standard	Disclosure Requirement
Data management	Baseline	Baseline	Baseline
Machine learning	High (difficult field)	Low	High
Genomic research	Very High (very difficult field)	Very Low	Very High

In contrast, a regime of macro-specificity contemplates that different inventions will be accorded PHOSITA standards on the basis of the invention's technological field, rather than by a more nuanced or particularistic analysis.⁹ Thus, in our example, the software-related inventions will have roughly the same standards for patentability, but distinct standards from the biotechnological invention. Burk and Lemley do not address this question, but the implication is that the patentability differences in this scheme (e.g., the difference between A' and B') are more pronounced than those in the micro-specific context.

Table 2: The Macro-Specific Approach

Invention	PHOSITA Level	Obviousness Standard	Disclosure Requirement ¹⁰
Data management	A	A'	A''
Machine learning	A	A'	A''
Genomic research	B	B'	B''

9. How one might determine the appropriate level of skill for a given *field* (as opposed to the invention before the court) is open to serious question, of course. Burk and Lemley suggest that software and biotechnology require roughly similar levels of skill—or at least do not deserve the divergent treatment they report in their review of the jurisprudence. See Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1191-96; Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 20-21.

10. Note in this regard that Burk and Lemley suggest that in their ideal *macro-specific* regime, the obviousness and disclosure standards would be “decoupled,” allowing policy-based variation for particular technologies. See Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1202-05; Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 60.

As between these two forms of specificity, I take the micro form to be both a positive description of the patent law, as well as a normatively justifiable position. The chief advantage (and challenge) of the patent law is its ability to provide a set of clear background (i.e., “property”) rules upon which private parties can build to invent, invest, and commercialize. Accordingly, the patent law must always retain the flexibility to adapt to new technological developments and economic shifts. In the micro-specific context, this flexibility is realized through the use of the PHOSITA standard as the lens through which a number of critical analyses are conducted. As a question of fact that should necessarily vary from particular innovation to particular innovation, the ordinary skill in the art framework grounds the legal abstractions of the patent law to the technological facts in any given case.

The macro form of technological exceptionalism, however, is far more problematic. Here, rather than building flexibility and innovation into the stable backdrop of the law, the project is broader, typically invoking arguments related to the “nature of the technology” or the “structure of the innovation,” or perhaps even the normative profile of the participants to support essentially *sui generis* changes in the patent law. Macro-specificity shifts consideration of the patent law from a general background principle of property rights to a vehicle for particularistic, technology-specific innovation policy choices. As I note in Part IV below, there are a number of reasons why it is worth at least challenging the efficacy and appropriateness of this development in the patent law.

One important limitation of the Burk and Lemley thesis is that the distinction set forth above remains unaccounted for in their analysis. They do seem to recognize what they describe as “inherent” technological-specificity, which might be taken to correspond to what I’ve described as micro-specificity.¹¹ Yet they also quite clearly perceive (and advocate for) a broader, macro version of exceptionalism as a means by which to influence technological development in the biotechnological field.¹² Because, as I argue below, it is by no means clear that the differences they identify result from conscious macro-specific behavior on the part of the Federal Circuit, this failure to account for both forms of technological specificity weakens their argument.

11. Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1191; Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 29.

12. Burk & Lemley identify what they suggest are “extraordinary” differences in the legal standards applied to software and biotechnology, respectively. *See* Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1191; Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 19-20.

B. The Indeterminate Scope-Effects of the PHOSITA

At various points in their argument, Burk & Lemley seek to connect the PHOSITA standard directly to the scope of the patent grant.¹³ This linkage, I suggest, is tenuous at best, for the following reasons.

First, *both* the disclosure requirements and the obviousness requirement are scope-affecting. Obviously, a higher standard of disclosure will force a patentee to claim more closely to what she has described, narrowing the literal scope of the patent.¹⁴ But the obviousness standard will also affect scope: a reduced standard of (non)obviousness will allow a patentee to establish claims “closer” to any relevant prior art.¹⁵ An extremely reduced version of the obviousness requirement—call it “anticipation”—will allow claims that merely avoid the disclosure of the prior art, as well as those that cover more innovative subject matter.¹⁶ Conversely, a higher standard of (non)obviousness will yield claims that are more distinct (in physical terms, more distant) from the prior art, and thus narrower.

Importantly, note the following: (1) the inverse relationship between the obviousness and disclosure standards (at least under current doctrine); and (2) the direct relationship between the scope-effects of the obviousness and disclosure standards. This suggests that, contrary to Burk and Lemley’s assumption, the patent scope-effects of changes in the PHOSITA standard will be fundamentally indeterminate, without knowledge of the relative magnitude of the disclosure-related and obviousness-related scope effects.¹⁷

13. For example, Burk and Lemley suggest that the PHOSITA level in the software cases will yield “a relatively small number of broad patents,” and that the different standards in the biotechnological cases will result in narrower, but fewer patents. *See* Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1172-73, 1180-82.

14. *See* Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1170; Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 10-11.

15. As well as, of course, enabling the patenting of subject matter that could not otherwise be patented. Indeed, it is this “gatekeeping” function that is perhaps the most important (and apparent) contribution of obviousness. *See generally* Robert P. Merges, *Uncertainty and the Standard of Patentability*, 7 HIGH TECH. L.J. 1 (1992).

16. *See, e.g.*, 35 U.S.C. § 102 (2000).

17. Perhaps one might assume that any obviousness-based scope effects will be swamped by those related to the disclosure requirement. This seems to me to be a problematic assumption, given the typicality of scope-reducing claim amendments as a means to overcome examiner rejections. Another possibility is to assume that patentees don’t care about the scope of their claims vis-à-vis the prior art, and instead claim aggressively only towards the “outer limit” that the disclosure allows. This seems somewhat more probable, but assumes irrational behavior. In any event, neither of these assumptions seems to fit with the Burk and Lemley argument.

A second observation concerning the PHOSITA and claim scope is that the standard's effect on the scope and availability of *equivalents* infringement is also fatally indeterminate. For example, a high degree of skill in the art (a difficult field) implies:

- 1) a relatively *narrower* doctrine of equivalents, because the possibility of "known interchangeability" between claim elements and their purported equivalents will be reduced;¹⁸
- 2) a relatively *broader* doctrine of equivalents, because prior-art based limitations on the doctrine will be less available;¹⁹
- 3) a relatively *narrower* doctrine of equivalents, because of the efforts of the Federal Circuit to limit equivalents due to the patent's disclosure;²⁰
- 4) a relatively *broader* doctrine of equivalents, because of the ability to more easily overcome the *Festo* presumption against equivalents in cases where a claim amendment would eliminate infringement of a technology unforeseeable by one of ordinary skill.²¹

Thus, while it is clear that the determination of a PHOSITA standard will *affect* the scope of the doctrine of equivalents, not much more can be safely concluded. Table 3 notes the relative effects of the PHOSITA standard on the scope of the patent.

18. See, e.g., Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 23-24; see also *Graver Tank & Mfg. Co. v. Linde Air Prods. Co.*, 339 U.S. 605, 609 (1950); *Hilton Davis Chem. Corp. v. Warner-Jenkinson Co.*, 62 F.3d 1512, 1519 (Fed. Cir. 1995) (en banc), *aff'd in part & rev'd in part on other grounds*, 520 U.S. 17 (1997).

19. See, e.g., *Wilson Sporting Goods Co. v. David Geoffrey & Assocs.*, 904 F.2d 677, 684 (Fed. Cir. 1990) ("[S]ince prior art always limits what an inventor could have claimed, it limits the range of permissible equivalents of a claim.").

20. See, e.g., *Sage Prods., Inc. v. Devon Indus.*, 126 F.3d 1420, 1424-26 (Fed. Cir. 1997) (limiting equivalents due to clarity of the patent's disclosure).

21. See, e.g., *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 535 U.S. 722, 740 (2002).

Table 3: The Scope-Effects of the PHOSITA Standard

PHOSITA Level	Obviousness Standard	Disclosure Standard	Scope-effects (obviousness)	Scope-effects (disclosure)	Scope-effects (equivalents)
Low (easy field)	Higher	Lower	Lower	Higher	Indeterminate
High (hard field)	Lower	Higher	Higher	Lower	Indeterminate

Again, given the varying results in three rightmost columns (scope-effects) for each case, any conclusion concerning the overall scope-effects of a change in the PHOSITA standard would require detailed knowledge concerning the relationship among the various scope-effects. Thus, while it is clear that the PHOSITA standard influences the scope of the patent, the direction and magnitude of that effect is quite indeterminate, and no meaningful conclusions can be drawn concerning the relationship.²²

This Part has argued that, although *Is Patent Law Technology-Specific?* correctly posits a difference in the way different technologies are treated, a more nuanced analysis suggests that both the nature and scope of the technological specificity in the modern patent doctrine is perhaps less substantially clear than Burk and Lemley suggest.

III. THE UNCERTAIN PATH-DEPENDENCIES OF THE PATENT LAW

As I established in Part II.A above, two distinct forms of technological-specificity can potentially be applied to the patent law. The first, *micro*-specificity, applies varying standards of patentability according to the specific technological circumstances—meaning that each invention (in theory at least) has a unique, contextual requirement for patentability. The second, *macro*-specificity, applies similar standards of patentability to inventions in the same technological field (or “industry,” as Burk and Lemley at times refer to it), while applying distinct standards to different technological fields. This Part explores which form(s) of exceptionalism can reasonably be said to exist in the patent law.

As an initial matter, I note that there is no real question but that micro forms of technological-specificity are fundamental to the patent law. That

22. Note, of course, that Burk and Lemley’s proposal that the courts “decouple” the linkage between obviousness and disclosure, and—in the biotechnological context—relax the disclosure requirement while maintaining the obviousness standard could in some cases yield the broader patents they seek. *See infra* Part IV.

is, whether the patent law is, as Burk and Lemley ask, “technology-specific,” strikes me, then, as an easy and rather obvious question.²³ Of course it is: among other aspects, the ordinary skill in the art standard implements the micro-exceptionalism described above. Thus the analysis here considers whether the broader, macro, form is descriptive of the modern patent law.

A. The Dogs That Don’t Bark: The Missing Evidence of Macro-Specificity

As Burk and Lemley note, the Federal Circuit’s patentability jurisprudence in the biotechnological area is self-consciously distinct from that in other technological fields such as software. Indeed, Burk and Lemley juxtapose language from cases considering the disclosure requirement in the software and biotechnological fields, illuminating that in the software context, the Federal Circuit has held that the disclosure of functions is sufficient, while in the biotechnological context, the Federal Circuit has held that the disclosure of genetic function is insufficient.²⁴ Yet this demonstration, standing alone, does little more than highlight the importance of the PHOSITA standard: in each case, the court was viewing the technology through the prism of a particular level of skill in the art.²⁵ It is apparent that the court believes that the levels of skill differ in relation to the two technologies at issue; we should expect, not resist, the distinct treatment.

In order to show *macro*-specificity, Burk and Lemley must argue (as they do) that the level of skill in the art is systematically approached differently in the biotechnological and software cases and that this is an enduring (likely policy-driven) feature of the patent law rather than transient in nature.²⁶ Yet even taking the Burk and Lemley approach to the relevant caselaw as correct, there remain a variety of reasons that the approach to the PHOSITA standard in this area could appear systematic and yet result in a jurisprudence that is far more micro-specific than macro-specific. For example:

- 1) *Sample size*. One possibility is that the systematic technological specificity identified by Burk and Lemley is essentially a statistical

23. They do, of course, acknowledge that the patent law is “inherently technology specific.” Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 29.

24. See Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1183-84 (quoting *Fonar Corp. v. Gen. Elec. Co.*, 107 F.3d 1543 (Fed. Cir. 1997); *Regents of the Univ. of Cal. v. Eli Lilly & Co.*, 119 F.3d 1559 (Fed. Cir. 1997)); accord Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 20-21 (same).

25. Compare *Fonar*, 107 F.3d at 1549, with *Eli Lilly*, 119 F.3d at 1568.

26. See Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1194-95.

artifact related to the fairly small number of cases extant in the relevant jurisprudence (i.e., those analyzed by these and other commentators).²⁷ The implication here is that the purported systemization would fade or disappear as more cases are decided.

- 2) *Judicial consistency.* Another possibility is that the technological specificity identified by Burk and Lemley is related to the (remarkably) small number of judges who have authored the opinions studied in the article (and others).²⁸ This suggests that any case-to-case consistency is a reflection of one judge's uniform approach, rather than either a court-wide decision or an enduring feature of the jurisprudence.²⁹
- 3) *Factual error.* Yet another possibility is that the systematic technological specificity identified by Burk and Lemley results from a judicial misunderstanding of the relevant facts.³⁰ The judges who

27. Burk & Lemley cite eight cases as representative of the suggested problematic approach to the issue. They are: *Enzo Biochem, Inc. v. Gen-Probe, Inc.*, 296 F.3d 1316 (Fed. Cir. 2002); *Enzo Biochem, Inc. v. Calgene, Inc.*, 188 F.3d 1362, 1371 (Fed. Cir. 1999); *Eli Lilly*, 119 F.3d at 1559; *In re Deuel*, 51 F.3d 1552, 1559 (Fed. Cir. 1995); *In re Goodman*, 11 F.3d 1046, 1052 (Fed. Cir. 1993); *In re Bell*, 991 F.2d 781 (Fed. Cir. 1993); *Fiers v. Rivel*, 984 F.2d 1164 (Fed. Cir. 1993); *Amgen, Inc. v. Chugai Pharm. Co.*, 927 F.2d 1200 (Fed. Cir. 1991). These cases, or a subset thereof, appear to be the most relevant to the commentators.

28. Of the opinions cited in footnote 27 as being relevant, every one except for *Goodman* has been authored by Judge Lourie.

29. There is another important observation to be made here. Given the essentially random assignment of cases to panels of judges, it is extremely unlikely that cases focusing on obviousness or disclosure in the biotechnological area have been uniformly assigned to panels containing Judge Lourie. This fact further suggests that Judge Lourie (perhaps alone among Federal Circuit judges) actively seeks opportunities to use his particular brand of biotechnological PHOSITA analysis. That is, Judge Lourie presumably is far more likely to seize the opportunity to analyze the disclosure (or obviousness) of biotechnological inventions, while his colleagues are more likely to decide these cases on other issues, such as software. Absent these sorts of selection-effects, it would be difficult to reconcile the Federal Circuit's professed random case assignment procedure with the pattern of decision-making in this area.

30. Burk and Lemley initially note this explanation themselves, but seem to suggest that the court has not considered these factual issues in more recent cases. See Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1194-95; Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 31-32. Yet the recent cases need not undermine the intuition that the factual analysis is flawed as a matter of technology: the failure to explicitly revisit technological facts might suggest that the court continues to believe them to be correct. Note that in the Federal Circuit's most recent effort at biotechnological disclosure standards, it remanded with explicit instructions to analyze the relevant technological

have addressed these issues thus far may not fully understand the detailed, fact-based distinctions—between genomic research and small molecule chemistry, for example—and thus may have been likely to simply transfer a PHOSITA standard from one context to another. Again, this concern is neither indicative of a macro-exceptionalist approach nor difficult to remedy going forward.³¹

- 4) *Fact/Law confusion*. A fourth possibility is that any systematic technological-specificity identified by Burk and Lemley arises as a result of confusion at the Federal Circuit concerning the nature of facts, law, and *stare decisis*. For example, the court may be failing to understand the implications of the distinctly fact-based inquiry into the PHOSITA with respect to appellate review. Or the court may simply be refusing to afford factual findings any deference, in favor of factual analysis of its own.³²

Any one of these explanations for Burk and Lemley's observed patterns in the patent law is sufficient to undermine their macro-specific argument. The most likely situation, of course, is that a combination of the above exists.

facts. *See Gen-Probe, Inc.*, 296 F.3d at 1324-26 (“It is not correct, however, that all functional descriptions of genetic material fail to meet the written description requirement.”).

31. Indeed, the court seems to have importantly reaffirmed the basic factual nature of the disclosure inquiry in its most recent precedent. In *Enzo Biochem, Inc. v. Gen-Probe, Inc.*, the court stated:

Although the patent specification lacks description of the location along the bacterial DNA to which the claimed sequences bind, Enzo has at least raised a genuine issue of material fact as to whether a reasonable fact-finder could conclude that the claimed sequences are described by their ability to hybridize to structures that, while not explicitly sequenced, are accessible to the public. Such hybridization to disclosed organisms may meet the PTO's Guidelines stating that functional claiming is permissible when the claimed material hybridizes to a disclosed substrate. That is a fact question. We therefore conclude that the district court erred in granting summary judgment that the claims are invalid for failure to meet the written description requirement. On remand, the court should consider whether one of skill in the art would find the generically claimed sequences described on the basis of Enzo's disclosure of the hybridization function and an accessible structure, consistent with the PTO Guidelines. If so, the written description requirement would be met.

296 F.3d at 1328.

32. Arti Rai suggests that this is an endemic problem with the Federal Circuit's jurisprudential approach. *See Rai, Facts, Law & Policy, supra* note 2.

This is not to suggest, however, that the application of the PHOSITA standard in the biotechnological area appears anywhere near optimal. Indeed, in my view, Burk and Lemley have compellingly identified problems with the Federal Circuit's jurisprudence; at least some of these concerns—especially the judge-based effects (number two above)—suggest that the court would be well-advised to carefully consider the process by which this doctrinal development is occurring.

B. The Factual Nature of the Federal Circuit's Technological Exceptionalism

The other possibilities offered above—that the court is factually mistaken or legally misunderstanding the role and nature of the PHOSITA analysis along a number of dimensions (numbers three and four)—are also worth criticizing. And yet these problems seem quite unlikely to create the sort of path-dependencies that would raise concerns of macro-specificity.

In this vein, most commentators appear to assume the future development of this technology-specific patent doctrine will continue along the presently-observed trajectory.³³ However, the “mistakes have been made” form of criticism has an easy answer: the use of correct technological facts.³⁴ The distinctly factual nature of micro-exceptionalism provides ample opportunity for future panels of the Federal Circuit to establish their own analysis in any given case. Indeed, an appropriate understanding of the role of the PHOSITA in the patent law would seem to virtually *preclude* the creation and use of categorical rules.³⁵ The state of the art in such fields is changing rapidly. That one of ordinary skill might have been

33. See, e.g., Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1194-95; Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 33-34; see also John M. Lucas, *The Doctrine of Simultaneous Conception and Reduction to Practice in Biotechnology: A Double Standard for the Double Helix*, 26 AIPLA Q.J. 381, 418 (1998) (cited in Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1192 n.167); Arti K. Rai, *Addressing the Patent Gold Rush: The Role of Deference to PTO Denials*, 2 WASH. U. J.L. & POL'Y 199 (2000) [hereinafter Rai, *Patent Gold Rush*]; Arti K. Rai, *Intellectual Property Rights in Biotechnology: Addressing New Technology*, 34 WAKE FOREST L. REV. 827 (1999) [hereinafter Rai, *Intellectual Property Rights*].

34. As Burk and Lemley seem to acknowledge, at least in part. Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 33.

35. To this end, the court's explicit description of aspects of the disclosure requirements in the software context as a “general rule” seems distinctly unwise. See, e.g., *Fonar Corp. v. Gen. Elec. Co.*, 107 F.3d 1543, 1549 (Fed. Cir. 1997). And to the extent that the court's failure to explicitly ground PHOSITA analyses in the biotechnological area to factual considerations can be taken to infer a form of the “general rule” statement noted above, corrective actions should be taken. In this vein, the recent discussion in *Enzo Biochem* might be viewed as a nod in this direction. See discussion and citation *supra* note 31.

unable to determine the DNA sequences that would code for EPO from a few examples circa 1984³⁶ seems nearly irrelevant to the level of knowledge in DNA sequence identification in the late 1990s.³⁷ Put another way, the explicit references to the “ordinary skill in the art to which [the invention] pertains” might be said to fundamentally *require* the reconsideration of issues of technological fact at each instance,³⁸ rather than perpetuating imprecise standards, even “decoupl[ed],”³⁹ as substitutes for technological fact. The correct rule as a matter of doctrine may also be the correct rule as a matter of policy: the courts may not, and should not, “standardize” the person of ordinary skill in the art.⁴⁰ Micro-specificity will prevail, and unhelpful path-dependencies will be avoided.

C. A Response to Burk and Lemley

Before moving to a broader critique of exceptionalist schemes, a brief response is in order. That is, in *Is Patent Law Technology-Specific?*, Burk and Lemley seek to brush aside the line of analysis in Part III of this Article, suggesting that it simply results from a different understanding of the relevant caselaw.⁴¹ But this response is too facile; as I’ve repeatedly noted above, I’ve assumed in this Part that their view of the relevant doctrine is correct. The point here is that even evaluating the precedent *just as they do* does not resolve the question of whether what we see in the patent law is macro-specificity, or merely a version of micro-specificity colored by fac-

36. U.S. Patent No. 4,703,008 (issued Oct. 27, 1987). The ’008 patent was entitled “DNA sequences encoding erythropoietin” and was filed November 30, 1984. The ’008 patent was at issue in *Amgen, Inc. v. Chugai Pharma. Co.*, 927 F.3d 1200 (Fed. Cir. 1991).

37. See, e.g., Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1181-82; Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 25; Lucas, *supra* note 33, at 418.

38. See 35 U.S.C. §§ 103, 112 (2000).

39. See Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1202; Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 27.

40. One response to this assertion is that the Federal Circuit, at least, seems to consider the prior rulings as having precedential value. See Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1185. This objection is unsatisfactory. First, the court always acknowledges the factual basis of the analysis. Second, notwithstanding the backwards citations, it is difficult to determine the actual weight given to earlier factual determinations in different cases. And third, I noted above the truly remarkable homogeneity of the relevant Federal Circuit decisions, which suggests that author consistency rather than doctrinal development is at issue. See *supra* note 27 and accompanying text.

41. Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 21. They note that Lawrence M. Sung has suggested that the biotechnology cases are little different in application from cases in other technological areas. See generally Lawrence M. Sung, *On Treating Past as Prologue*, 2001 U. ILL. J.L. TECH. & POL’Y 75 (2001). This is not, however, the argument I make.

tual error or the unusual circumstances surrounding this line of cases.⁴² While it is possible that the Federal Circuit has created a policy-driven, enduring, macro-specific doctrine for the field of biotechnology, that issue has yet to be resolved.

IV. OF COASE AND COMPLEXITY: FIXING THE PATENT DOCTRINE

Even if the patent law evinces a technologically-exceptionalist approach—with disparate legal rules applied to different technological fields—there remain significant reasons to believe that the effort to formalize and tailor such exceptionalism, as Burk and Lemley advocate, is misguided.

As several commentators have observed, there is at least some concern that the field of biotechnology in particular has structural and technological features that might make it susceptible to transaction-costs and related forms of market inefficiencies.⁴³ Generally referred to by the term “anticommons,” the theory suggests that the difficulty in arranging and aggregating the patent rights necessary to actually deliver marketable goods will stymie the participants in this field—to a degree that will ultimately reduce the pace of technological development, and thus increase social losses.⁴⁴ There are a variety of responses to this perceived problem in the literature, ranging from vertical integration,⁴⁵ to the formation of collective rights organizations,⁴⁶ to the denial of patenting altogether in some areas of the field.⁴⁷ To these, Burk and Lemley add another: the expansion

42. See *supra* Part II.A.

43. See, e.g., Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons on Biomedical Research*, 280 SCI. 698 (1998); Arti K. Rai, *The Information Revolution Reaches Pharmaceuticals: Balancing Innovation Incentives, Cost, and Access in the Post-Genomics Era*, 2001 U. ILL. L. REV. 173 (2001) [hereinafter Rai, *Information Revolution*]. But see JOHN P. WALSH ET AL., THE PATENTING AND LICENSING OF RESEARCH TOOLS AND BIOMEDICAL INNOVATION (The Heinz School—Carnegie Mellon University, Working Paper No. 2, 2003).

44. See Heller & Eisenberg, *supra* note 43, at 698-99. For a full description of the anticommons theory, see Michael A. Heller, *The Tragedy of the Anticommons: Property in the Transition from Marx to Markets*, 111 HARV. L. REV. 621 (1998).

45. See, e.g., Arti K. Rai, *Fostering Cumulative Innovation in the Biopharmaceutical Industry: The Role of Patents and Antitrust*, 16 BERKELEY TECH. L.J. 813 (2001) [hereinafter, Rai, *Fostering Cumulative Innovation*] (describing and criticizing this trend)

46. See, e.g., Heller & Eisenberg, *supra* note 43. On collective rights organizations more generally, see Robert Merges, *Contracting into Liability Rules: Intellectual Property Rights and Collective Rights Organizations*, 84 CALIF. L. REV. 1293 (1996).

47. See, e.g., Heller & Eisenberg, *supra* note 43; Rai, *Fostering Cumulative Innovation*, *supra* note 45.

of patent rights in these areas, so as to create conditions more akin to Kitch's "prospect" theory.⁴⁸ In Burk and Lemley's view, such expansion—together with modestly increased standards for patentability—will yield fewer and more powerful patents,⁴⁹ thus decreasing the characteristics of the biotechnology field that might create an anticommons problem.⁵⁰

Note in this regard that there are serious problems with Burk and Lemley's assumption concerning the relationship between the frequency of patents and their enforceable scope. Their suggestion seems to be that strengthening the obviousness requirement while simultaneously weakening the disclosure requirement will yield fewer yet broader patents.⁵¹

An initial problem is the assumption that such changes will affect patent scope in the way that Burk and Lemley suggest. In Part II.B above, I described the indeterminacy of the relationship between patent scope and simultaneous changes in the obviousness and disclosure requirements—the most that can be said without a series of difficult empirical assumptions is that scope will be affected. The magnitude and direction of the effect, however, is unclear.⁵² The suggestion that those doctrines be "decoupled" might have helped their case, if they limited their proposal to loosening the disclosure requirement, for example.⁵³

Perhaps even more troubling are the assumptions about patent frequency. That there will be fewer patents of course does not logically follow from broader patents. This ignores the ex ante incentives of the patent system: broader (stronger) patents will induce additional incentives to en-

48. The prospect theory of patents analogizes patents to mineral claims—as opportunities for further investment and exploitation rather than a reward for innovation. See Edmund W. Kitch, *The Nature and Function of the Patent System*, 20 J.L. & ECON. 265, 271-80 (1977) (broad "prospect" patents allow for better resource allocation to innovation); see also Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 48 (noting the benefits of Kitch's "prospect" theory in the pharmaceutical context).

49. Note that there are serious problems with Burk and Lemley's assumption concerning the relationship between the frequency of patents and their scope. The suggestion seems to be that strengthening the obviousness requirement while simultaneously weakening the disclosure requirement will yield fewer yet broader patents. See Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1173, 1182, 1195-96; Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 62.

50. See Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 62.

51. See Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1173, 1182, 1195-96; Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 62.

52. See *supra* tbl. 3.

53. See Burk & Lemley, *Technology Specific*, *supra* note 1, at 1205 (advocating the "decoupling" of the standards for obviousness and disclosure); see also Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 62 (positing simultaneous changes).

gage in inventive behavior (or at least patenting behavior). This has been empirically verified.⁵⁴ Thus, broader patents should lead to greater, not fewer, patents.⁵⁵ To the extent that Burk and Lemley rely on their proposed increase in the obviousness standard to yield fewer patents, that simply raises the fatal indeterminacy problem noted above: Will the additional difficulty of obtaining patents (the heightened (non)obviousness requirement) outweigh the effects of broader and stronger patents (due to lower disclosure requirements)? The bottom line is that the effects of the Burk and Lemley proposals on the scope and frequency of patents is unclear.

Finally, while I am generally sympathetic to Burk and Lemley's view that strengthening biotech patents is likely to be a better solution along a number of dimensions than reducing patent scope,⁵⁶ a third option appears dominant here, especially given the uncertainties surrounding the premises of the Burk and Lemley argument. That is, the possibility of an anticommons in biotechnology (in particular) could be ameliorated, perhaps significantly, via relatively straightforward efforts to clarify and stabilize the patent jurisprudence, thereby reducing the transaction costs of combining rights.⁵⁷ This point is simply Coasean: while other commentators focus on the appropriate entitlements,⁵⁸ the reduction of transaction costs (thus diminishing the importance of those entitlements) may well provide a better payoff.⁵⁹ To be sure, we can never eliminate the transaction and related

54. See, e.g., See, e.g., Bronwyn H. Hall & Rosemarie Ham Ziedonis, *The Patent Paradox Revisited: Determinants of Patenting in the US Semiconductor Industry, 1980-94*, 32 RAND J. ECON. 101-25 (2001) (documenting an increase in patenting linked to strengthening the patent law in the 1980s).

55. Cf. Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 62.

56. See generally R. Polk Wagner, *Information Wants to Be Free: Intellectual Property and the Mythologies of Control*, 103 COLUM. L. REV. 995 (2003) (noting the benefits of broader, as opposed to narrower, property rights in information goods).

57. That is, the Federal Circuit could work to clarify and stabilize the patent law, rather than adopting substantive changes. And while there is some question concerning the success of the Federal Circuit in doing so, this approach is well-established as a part of the mission of the Federal Circuit. See, e.g., R. Polk Wagner & Lee Petherbridge, *Is the Federal Circuit Succeeding? An Empirical Assessment of Judicial Performance*, 152 U. PA. L. REV. (forthcoming 2004).

Of relevance in this context would be efforts to bring additional clarity and predictability to the obviousness and disclosure doctrines that Burk and Lemley specifically address.

58. See *supra* notes 43-47.

59. Ronald Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1 (1960) (noting the importance of transaction costs in entitlement analysis). Note that Heller and Eisenberg suggested in their original, groundbreaking article that collective rights organizations—

costs inherent in the patent system, and efforts to clarify the law can only achieve part of this goal at best. But all options in this context are “second best” in nature: the goal is to improve the current situation, given the world as we find it.

One important question is how one might seek to clarify patent rights in this context. And while a full treatment of that question is well beyond the scope of this Article, one thing we would surely *not* want to do is create the additional jurisprudential and doctrinal confusion that will result from introducing macro-exceptionalism to the patent law. In this sense *Is Patent Law Technology-Specific?* is a virtually perfect indictment of itself: Burk and Lemley go to great lengths to demonstrate the troublesome aspects of macro-specificity in the patent law, reserving special criticism for the Federal Circuit’s inability to adequately understand the innovation policy needs of the modern biotechnology and software industries.⁶⁰ Having demonstrated (they suggest) the error of the court’s technology-specific ways, one might expect that the next step is to suggest doctrinal adjustments to eliminate the inter-industry differences they suggest are harming technological development. Instead, they suggest that the court apply an entirely new jurisprudential framework, based on underlying principles of innovation policy. (Not to mention that the new doctrine would “decouple” the relatively uniform PHOSITA standard, thus requiring *at least* two detailed factual analyses to replace one.⁶¹) Implicit in their proposal is the idea that this new jurisprudential framework would (a) require reconsideration of obviousness and disclosure standards in all technological fields served by the patent law⁶²; and (b) be subject to (and under obligation of) revision by the court anytime the background conditions for innovation in a particular field change.⁶³ And all this policy-driven, unconstrained decisionmaking with incredible importance for the future of technological development is being placed into the hands of the Federal Circuit—the very body that Burk and Lemley suggest has done such a poor job to this point.

In sum, this does not seem to be a proposal that is likely to increase the certainty and stability of the patent law in the biotechnological nor other

another transaction-cost-reducing mechanism—might be a solution to any anticommons problem. *See* Heller & Eisenberg, *supra* note 43.

60. *See* Burk & Lemley, *Technology-Specific*, *supra* note 1, at 1195-96.

61. *See id.* at 1202.

62. That is to say *all* technological fields.

63. One conjures up troubling images of the Federal Circuit reviewing evidence concerning the availability of early-stage capital investment vehicles, interest rates, market conditions, and the like in a particular field prior to establishing the levels of (non)obviousness and disclosure required.

areas. Indeed, a transaction-cost-focused analysis would suggest that it is Burk and Lemley, rather than the Federal Circuit, who have it “exactly backwards.”⁶⁴ Given the deep uncertainties underlying the premises of the argument,⁶⁵ as well as the promise of increasing transaction costs resulting from a shift from the *micro*-specific to *macro*-specific approach, there seem to be strong reasons to conclude that we should instead work diligently to reduce the very sort technological-specificity that *Is Patent Law Technology Specific?* advocates.

64. Cf. Burk & Lemley, *Uncertainty Principle*, *supra* note 1, at 3.

65. See *supra* Parts II, III.