JOURNAL ON TELECOMMUNICATIONS & HIGH TECHNOLOGY LAW
is published semi-annually by the
Journal on Telecommunications & High Technology Law,
Campus Box 401, Boulder, CO 80309-0401

ISSN: 1543-8899

POSTMASTER: Please send address changes to JTHTL,
Campus Box 401, Boulder, CO 80309-0401

Subscriptions

Domestic volume subscriptions are available for $45.00. City of Boulder subscribers please add $3.74 sales tax. Boulder County subscribers outside the City of Boulder please add $2.14 sales tax. Metro Denver subscribers outside of Boulder County please add $1.85 sales tax. Colorado subscribers outside of Metro Denver please add $1.31 sales tax.

International volume subscriptions are available for $50.00.

Inquiries concerning ongoing subscriptions or obtaining an individual issue should be directed to the attention of JTHTL Managing Editor at JTHTL@colorado.edu or by writing JTHTL Managing Editor, Campus Box 401, Boulder, CO 80309-0401.

Back issues in complete sets, volumes, or single issues may be obtained from: William S. Hein & Co., Inc., 1285 Main Street, Buffalo, NY 14209. Back issues may also be found in electronic format for all your research needs on HeinOnline http://heinonline.org/.

Manuscripts

JTHTL invites the submission of unsolicited manuscripts. Please send softcopy manuscripts to the attention of JTHTL Articles Editors at JTHTL@colorado.edu in Word or PDF formats or through ExpressO at http://law.bepress.com/expresso. Hardcopy submissions may be sent to JTHTL Articles Editors, Campus Box 401, Boulder, CO 80309-0401. Unfortunately, JTHTL cannot return manuscripts. JTHTL uses THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (18th ed. 2005) for citation format and THE CHICAGO MANUAL OF STYLE (15th ed. 2003) for a style guide.

Cite as: 7 J. ON TELECOMM. & HIGH TECH. L. __ (2009).

J. ON TELECOMM. & HIGH TECH. L.

JOURNAL ON TELECOMMUNICATIONS & HIGH TECHNOLOGY LAW

Volume 7 *Spring 2009*

J. ON TELECOMM. & HIGH TECH. L.

THE UNIVERSITY OF COLORADO LAW SCHOOL

FACULTY, 2008–09

Deans

DAVID H. GETCHES, *DEAN AND RAPHAEL J. MOSES PROFESSOR OF NATURAL RESOURCES LAW*. A.B., OCCIDENTAL COLLEGE; J.D., UNIVERSITY OF SOUTHERN CALIFORNIA.

DAYNA BOWEN MATTHEW, *ASSOCIATE DEAN FOR ACADEMIC AFFAIRS AND PROFESSOR OF LAW,* A.B., HARVARD-RADCLIFFE; J.D., UNIVERSITY OF VIRGINIA.

PHILIP J. WEISER, *ASSOCIATE DEAN FOR RESEARCH, PROFESSOR OF LAW, AND EXECUTIVE DIRECTOR OF THE SILICON FLATIRONS TELECOMMUNICATIONS PROGRAM*. B.A., SWARTHMORE COLLEGE; J.D., NEW YORK UNIVERSITY.

LORENZO A. TRUJILLO, *ASSISTANT DEAN FOR STUDENTS AND PROFESSIONAL PROGRAMS AND PROFESSOR ATTENDANT RANK*, B.A., UNIVERSITY OF COLORADO; M.A., UNIVERSITY OF COLORADO; ED.D., UNIVERSITY OF SAN FRANCISCO; J.D., UNIVERSITY OF COLORADO.

SUSANI CLAYTON, *ASSISTANT DEAN FOR CAREER DEVELOPMENT*. B.A., UNIVERSITY OF COLORADO; M.A., UNIVERSITY OF ARIZONA; J.D., UNIVERSITY OF DENVER COLLEGE OF LAW.

KRISTINE H. MCCORD, *ASSISTANT DEAN FOR ADMISSIONS AND FINANCIAL AID*. B.S., UNIVERSITY OF NORTH CAROLINA; J.D., GEORGE MASON UNIVERSITY.

Emeritus Faculty

CLIFFORD J. CALHOUN, *PROFESSOR EMERITUS*. A.B., LL.B., HARVARD UNIVERSITY.

HOMER H. CLARK, JR., *PROFESSOR EMERITUS*. A.B., LL.D., AMHERST COLLEGE; LL.B., LL.M., HARVARD UNIVERSITY.

JAMES N. CORBRIDGE, JR., *PROFESSOR EMERITUS*. A.B., BROWN UNIVERSITY; LL.B., YALE UNIVERSITY.

TED J. FIFLIS, *PROFESSOR EMERITUS*. B.S., NORTHWESTERN UNIVERSITY; LL.B., HARVARD UNIVERSITY

David S. Hill, *Professor Emeritus*. B.S., J.D., University of Nebraska.

J. Dennis Hynes, *Professor Emeritus*. B.A., LL.B., University of Colorado.

Howard C. Klemme, *Professor Emeritus*. B.A., LL.B., University of Colorado; LL.M., Yale University.

Courtland H. Peterson, *Nicholas Doman Professor of International Law Emeritus*. B.A., LL.B., University of Colorado; M. Comp. L., University of Chicago; Dr. Jur., University of Freiburg (Germany).

William E. Rentfro, *Professor Emeritus*. B.A., University of Colorado; Th.M., LL.B., University of Denver.

Don W. Sears, *Professor Emeritus*. B.S., J.D., Ohio State University.

Peter N. Simon, *Professor Emeritus*. B.S., M.D., University of Wisconsin; J.D., University of California, Berkeley.

Norton L. Steuben, *Professor Emeritus*. A.B., J.D., University of Michigan.

Arthur H. Travers, Jr., *Professor Emeritus*. B.A., Grinnell College; LL.B., Harvard University.

Tenured and Tenure-Track Faculty

Barbara A. Bintliff, *Nicholas Rosenbaum Professor of Law and Law Library Director*. B.A., Central Washington State College; J.D., M.L.L., University of Washington.

William Boyd, *Associate Professor of Law*. Ph.D., University of California, Berkeley; J.D., Stanford University.

Harold H. Bruff, *Charles Inglis Thomson Professor of Law*. B.A., Williams College; J.D., Harvard University.

Maxine Burkett, *Associate Professor of Law*. B.A., Williams College; J.D., University of California, Berkeley.

Emily M. Calhoun, *Professor of Law*. B.A., M.A., Texas Tech University; J.D., University of Texas.

Paul F. Campos, *Professor of Law*. A.B., M.A., J.D., University of Michigan.

Richard B. Collins, *Professor of Law and Director of the Byron R. White Center for the Study of American Constitutional Law*. B.A., Yale College; LL.B., Harvard University.

Nestor Davidson, *Associate Professor of Law*. A.B., Harvard University; J.D., Columbia University.

Wayne M. Gazur, *Professor of Law.* B.S., University of Wyoming; J.D., University of Colorado; LL.M., University of Denver.

Lakshman D. Guruswamy, *Professor of Law.* LL.B., Sri Lanka; Ph.D., University of Durham, U.K.

Melissa Hart, *Associate Professor of Law.* B.A., Harvard-Radcliffe College; J.D., Harvard University.

Clare Huntington, *Associate Professor of Law.* B.A., Oberlin College; J.D., Columbia University.

Sarah A. Krakoff, *Associate Professor of Law.* B.A., Yale University; LL.B, University of California, Berkeley.

Mark J. Loewenstein, *Nicholas A. Rosenbaum Professor of Law.* A.B., J.D., University of Illinois.

Scott A. Moss, *Associate Professor of Law.* B.A., M.A., Stanford University; J.D., Harvard University.

Christopher B. Mueller, *Henry S. Lindsley Professor of Procedure and Advocacy.* A.B., Haverford College; J.D., University of California, Berkeley.

Robert F. Nagel, *Ira C. Rothgerber, Jr. Professor of Constitutional Law.* B.A., Swarthmore College; J.D., Yale University.

Helen L. Norton, *Associate Professor of Law.* B.A., Stanford University; J.D., University of California, Berkeley.

Paul Ohm, *Associate Professor of Law.* B.S./B.A., Yale University; J.D., University of California, Los Angeles.

Scott R. Peppet, *Associate Professor of Law.* B.A., Cornell University; J.D., Harvard University.

William T. Pizzi, *Professor of Law.* A.B., Holy Cross College; M.A., University of Massachusetts; J.D., Harvard University.

Carolyn B. Ramsey, *Associate Professor of Law.* B.A., University of California, Irvine; A.M., Stanford University; J.D., Stanford University.

Pierre J. Schlag, *Byron White Professor of Constitutional Law.* B.A., Yale University; J.D., University of California, Los Angeles.

Amy J. Schmitz, *Associate Professor of Law.* B.A., Drake University; J.D., University of Minnesota.

Laura Spitz, *Associate Professor of Law.* B.A., University of Toronto; LL.B., University of British Columbia Faculty of Law; J.S.D., Cornell University.

ANDREW SCHWARTZ, *ASSOCIATE PROFESSOR OF LAW.* B.A., BROWN UNIVERSITY; J.D., COLUMBIA UNIVERSITY.

MARK SQUILLACE, *PROFESSOR OF LAW AND DIRECTOR OF THE NATURAL RESOURCES LAW CENTER.* B.S., MICHIGAN STATE UNIVERSITY; J.D., UNIVERSITY OF UTAH.

HARRY SURDEN, *ASSOCIATE PROFESSOR OF LAW.* B.A., CORNELL UNIVERSITY; J.D., STANFORD UNIVERSITY.

MICHAEL J. WAGGONER, *ASSOCIATE PROFESSOR OF LAW.* A.B., STANFORD UNIVERSITY; LL.B., HARVARD UNIVERSITY.

MARIANNE WESSON, *PROFESSOR OF LAW AND WOLF-NICHOL FELLOW.* A.B., VASSAR COLLEGE; J.D., UNIVERSITY OF TEXAS.

AHMED A. WHITE, *ASSOCIATE PROFESSOR OF LAW.* B.A., SOUTHERN UNIVERSITY AND A & M COLLEGE; J.D., YALE UNIVERSITY.

CHARLES F. WILKINSON, *UNIVERSITY'S DISTINGUISHED PROFESSOR AND MOSES LASKY PROFESSOR OF LAW.* B.A., DENISON UNIVERSITY; LL.B., STANFORD UNIVERSITY.

Research and Clinical Faculty

NORMAN F. AARONSON, *CLINICAL PROFESSOR EMERITUS, LEGAL AID AND DEFENDER PROGRAM.* A.B., BRANDEIS UNIVERSITY; J.D., BOSTON UNIVERSITY.

J. BRAD BERNTHAL, *ASSOCIATE CLINICAL PROFESSOR OF LAW.* B.A., UNIVERSITY OF KANSAS; J.D., UNIVERSITY OF COLORADO.

DEBORAH J. CANTRELL, *DIRECTOR OF CLINICAL PROGRAMS & ASSOCIATE PROFESSOR OF LAW.* B.A., SMITH COLLEGE; M.A., UNIVERSITY OF CALIFORNIA, LOS ANGELES; J.D., UNIVERSITY OF SOUTHERN CALIFORNIA.

MARGARET ANN ENGLAND, ASSOCIATE *CLINICAL PROFESSOR, LEGAL AID AND DEFENDER PROGRAM.* B.A., UNIVERSITY OF MICHIGAN; J.D., UNIVERSITY OF DENVER.

H. PATRICK FURMAN, *CLINICAL PROFESSOR EMERITUS, LEGAL AID AND DEFENDER PROGRAM, AND DIRECTOR OF CLINICAL PROGRAMS.* B.A., J.D., UNIVERSITY OF COLORADO.

COLENE ROBINSON, ASSOCIATE *CLINICAL PROFESSOR, JUVENILE AND FAMILY LAW.* B.A., VALPARAISO UNIVERSITY; J.D., LOYOLA UNIVERSITY SCHOOL OF LAW, CHICAGO.

JILL E. TOMPKINS, *CLINICAL PROFESSOR AND DIRECTOR OF THE AMERICAN INDIAN LAW CLINIC.* B.A., THE KING'S COLLEGE; J.D., UNIVERSITY OF MAINE.

LEGAL WRITING AND APPELLATE ADVOCACY FACULTY

AL CANNER, *LEGAL WRITING PROFESSOR*. B.A., BRANDEIS UNIVERSITY; J.D., UNIVERSITY OF COLORADO.

CATHERINE DU BOIS, *LEGAL WRITING PROFESSOR*.

DEREK H. KIERNAN-JOHNSON, *LEGAL WRITING PROFESSOR*. A.B. PRINCETON UNIVERSITY; J.D., UNIVERSITY OF MICHIGAN.

NATALIE MACK, *LEGAL WRITING PROFESSOR*. B.S., UNIVERSITY OF SOUTH CAROLINA; J.D., UNIVERSITY OF COLORADO.

GABRIELLE M. STAFFORD, *LEGAL WRITING PROFESSOR*. B.A., UNIVERSITY OF PENNSYLVANIA; J.D., BOSTON UNIVERSITY.

TODD M. STAFFORD, *LEGAL WRITING PROFESSOR*. B.A., SOUTHERN METHODIST UNIVERSITY; J.D., DUKE UNIVERSITY.

Law Library Faculty

BARBARA A. BINTLIFF, *NICHOLAS ROSENBAUM PROFESSOR OF LAW AND LAW LIBRARY DIRECTOR*. B.A., CENTRAL WASHINGTON STATE COLLEGE; J.D., M.L.L., UNIVERSITY OF WASHINGTON.

ALICIA BRILLON, *REFERENCE LIBRARIAN*. B.A., M.L.I.S., UNIVERSITY OF WASHINGTON; J.D., SEATTLE UNIVERSITY.

GEORGIA K. BRISCOE, *ASSOCIATE DIRECTOR AND HEAD OF TECHNICAL SERVICES*. B.S., WASHINGTON STATE UNIVERSITY; M.A., UNIVERSITY OF SAN DIEGO; M.L.S., UNIVERSITY OF MICHIGAN.

YUMIN JIANG, *TECHNICAL SERVICES LIBRARIAN*. M.S., UNIVERSITY OF ILLINOIS, URBANA-CHAMPAIGN; M.A., UNIVERSITY OF WISCONSIN.

ROBERT LINZ, *HEAD OF PUBLIC SERVICES AND ASSOCIATE DIRECTOR OF LAW LIBRARY*.

ALAN PANNELL, *REFERENCE LIBRARIAN*. B.A. UNIVERSITY OF OKLAHOMA; J.D. WESTERN NEW ENGLAND COLLEGE; M.A. UNIVERSITY OF ARIZONA.

KAREN SELDEN, *CATALOG LIBRARIAN*. B.S., PENNSYLVANIA STATE UNIVERSITY; M.L.S., SIMMONS COLLEGE.

JANE E. THOMPSON, *HEAD OF FACULTY SERVICES*. B.A., UNIVERSITY OF MISSOURI; M.A., J.D., UNIVERSITY OF DENVER.

Adjunct, Adjoint and Visiting Faculty

GARRY R. APPEL, *ATTORNEY AT LAW, APPEL & LUCAS, P.C., DENVER, COLORADO*. B.A., J.D., UNIVERSITY OF COLORADO.

DOUGLAS S. KENNEY, *RESEARCH ASSOCIATE, NATURAL RESOURCES LAW CENTER*. B.A., UNIVERSITY OF COLORADO; M.S., UNIVERSITY OF MICHIGAN SCHOOL OF NATURAL RESOURCES AND ENVIRONMENT; PH.D., CORNELL UNIVERSITY.

KATHRYN M. MUTZ, *RESEARCH ASSOCIATE, NATURAL RESOURCES LAW CENTER*. B.A., UNIVERSITY OF CHICAGO; M.S., UTAH STATE UNIVERSITY; J.D., UNIVERSITY OF COLORADO.

ADAM REED, *RESEARCH FELLOW, CENTER FOR ENERGY & ENVIRONMENTAL SECURITY (CEES)*. B.A., SOUTHERN METHODIST UNIVERSITY; J.D., UNIVERSITY OF COLORADO.

JILL RENNERT, *RESEARCH ASSOCIATE, SILICON FLATIRONS TELECOMMUNICATIONS PROGRAM*.

JULIE TEEL, *SENIOR RESEARCH ASSOCIATE, CENTER FOR ENERGY & ENVIRONMENTAL SECURITY*. B.S., UNIVERSITY OF CALIFORNIA, BERKELEY; J.D., NEW YORK UNIVERSITY.

JILL VAN MATRE, *RESEARCH ASSOCIATE, SILICON FLATIRONS TELECOMMUNICATIONS PROGRAM*. B.S., INDIANA UNIVERSITY; J.D., UNIVERSITY OF COLORADO.

FROM THE EDITOR

The world of telecommunications is primed to experience many changes. With the dawn of a new administration come changes in executive policy and personnel. For example, President Obama recently announced that a new official would oversee security of the digital infrastructure.[1] President Obama has also "borrowed" Professor Phil Weiser and appointed him to the U.S. Department of Justice. Though we will miss Phil immensely, we wish him all the best. He has been the rock of this journal since its inception and we are thankful that he has brought together such a great support network, including Paul Ohm, Dale Hatfield, Brad Bernthal, Harry Surden and Andrew Schwartz, to continue guiding the *Journal*. It is with these topics in the foreground that we are pleased to present the second and final issue of the seventh volume of the *Journal on Telecommunications & High Technology Law*.

This issue begins with an article from Professor F.M. Scherer discussing the evolution of patent reform in the United States and abroad. The next article is a contribution from authors Rick Whitt and Stephen Schultze. Their article discusses "Emergence Economics" and applies this concept to communication policy. The next two articles conclude the pieces presented at the Silicon Flatirons Digital Broadband Migration Conference, which focused on "Information Policy for the Next Administration." Professor Christopher Sprigman's article advocates applying antitrust principles to copyright law. Professor Ellen Goodman discusses how the goals of communication policy might not be best served through the auction process used in the 700 MHz auction.

As in every issue, we are also pleased to present student notes. Dana Jozefczyk, a Casenote and Comment Editor, writes about Apple's Fairplay and the antitrust cases brought against Apple alleging illegal tying. In his article, John Bergmayer, the Lead Production Editor, argues against using lawyers or legislatures to create policy solutions with respect to "software monocultures" when these issues may be resolved through technical advancements. Finally, we are thankful to Faegre & Benson for sponsoring the 2008 Silicon Flatirons Writing Competition. David Wilson, JTHTL's Editor in Chief for volume six, won the competition for his article regarding the Navajo Nation and the Internet to the

---

[1] Cam Simpson & August Cole, *Obama Moves to Curb Data-System Attacks*, WALL ST.J., May 30-31, 2009, at A3.

Hogan technology plan. In his article, he discusses jurisdictional questions surrounding tribal sovereignty regarding the plan and advocates policy and statutory changes.

I thank these authors for their contributions. I also cannot express enough gratitude to the entire board and the members for their time and commitment to the *Journal*. A special thanks to our Production Editors, John Bergmayer and Chris Larson, for their countless hours of work on these articles.

As I mentioned above, volume eight marks a transition year for the *Journal*. Judging by the work they have already done, I know there is no group of people better equipped to handle the changes than the board of volume eight, led by Blake Reid, Ty Martinez, Avi Loewenstein, Per Larsen and Jeff O'Holleran.

Finally, I have to thank my family and friends for their support and patience. I also must thank the many professors at the University of Colorado Law School who have provided astounding expertise and support. I am proud to call them my mentors and friends.

Along with the entire board of editors, I am pleased to present the second and final issue of the seventh volume of the *Journal on Telecommunications & High Technology Law*.

Hiwot Molla
Editor-in-Chief

JOURNAL ON
TELECOMMUNICATIONS & HIGH TECHNOLOGY LAW

Volume 7 *Spring 2009*

CONTENTS

ARTICLES

STUDENT NOTES

SILICON FLATIRONS WRITING COMPETITION

THE POLITICAL ECONOMY OF PATENT POLICY REFORM IN THE UNITED STATES

F. M. SCHERER*

INTRODUCTION

During the 1980s and 1990s, important legislative, judicial, and diplomatic initiatives emanated from the United States, strengthening patent and copyright enforcement systems both domestically and in the broader world economy. The political influences that led to these changes are interesting in their own right.[1] Even more interesting, however, is the fact that governmental emphasis on patent systems increased in the wake of impressive new findings from economic studies showing that patents played a surprisingly minor role in well-established corporations' decisions to invest in research, development, and

* Aetna Professor Emeritus, John F. Kennedy School of Government, Harvard University. The author thanks Wesley J. Cohen, Cecil Quillen, and Phil Weiser for valuable comments.

1. For a contribution with a similar focus and some similar conclusions, see WILLIAM M. LANDES & RICHARD A. POSNER, THE POLITICAL ECONOMY OF INTELLECTUAL PROPERTY LAW (2004), which is derived from WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW (2003).

technological innovation. The opposing movements of the political and behavioral science currents will be a principal theme of this article.

I. THE TURBULENT EARLY HISTORY

Governments' policies toward patents on inventions and copyright for artistic works have been marked by appreciable fluctuations over the course of history. At the dawn of the 17th century, patents and copyrights were components of the feudal system in Western Europe.[2] Sovereigns awarded to selected individuals exclusive privileges to pursue a mechanical trade, publish books or music, and present theatrical performances—usually but not always those with close connections to the noble courts and often favorites of the court. The privilege system was attacked under the banner of the Enlightenment, first during the reign of James I in England (1603-25) and then during the 1779 French Revolution and the eastward spread of anti-feudal policies under Napoleon.[3] It was replaced by patents and copyrights made available to the middle classes through more transparent procedures, but limited in the time span over which exclusivity was applicable. In the New World, granting to authors and inventors exclusive rights to their writings and discoveries for limited times was enshrined in Article I, Section 8, of the U.S. Constitution.[4]

The period between the 1770s and 1840s, when patent and copyright laws spread rapidly, was followed, at least in Europe (but less so in the United States), by an "anti-patent" movement. In England, reforms following publication of Charles Dickens' spoof, *A Poor Man's Tale of a Patent*,[5] simplified the processes by which patents were issued, imposed stricter examination of patent applications, and allowed abrogation of exclusive rights in cases of demonstrated abuse.[6] The Swiss legislature repeatedly rejected proposals to enact patent laws, and in the Netherlands, existing patent laws were repealed in 1869, to be reenacted

---

2. For authoritative histories, see Fritz Machlup & Edith Tilton Penrose, *The Patent Controversy in the Nineteenth Century*, 10 J. ECON. HIST. 1 (1950); SUBCOMM. ON PATENTS, TRADEMARKS, AND COPYRIGHTS OF THE S. COMM. ON THE JUDICIARY, 85TH CONG., 2D SESS., AN ECONOMIC REVIEW OF THE PATENT SYSTEM (prepared by Fritz Machlup) (Study No. 15, Comm. Print. 1958), *available at* http://www.mises.org/etexts/patentsystem.pdf; putting copyright privileges in a more democratic light, HANSJÖRG POHLMANN, DIE FRÜHGESCHICHTE DES MUSIKALISCHEN URHEBERRECHTS (1962).

3. *See* Machlup & Penrose, *supra* note 2.

4. U.S. CONST. art. I, § 8

5. CHARLES DICKENS, A POOR MAN'S TALE OF A PATENT (1850), *reprinted in* D. VAVER. INTELLECTUAL PROPERTY RIGHTS: CRITICAL CONCEPTS IN LAW 37-42 (2006)

6. *See* DOMINIQUE GUELLEC & BRUNO VAN POTTELSBERGHE DE LA POTTERIE, THE ECONOMICS OF THE EUROPEAN PATENT SYSTEM 24 (2007).

only in 1910.[7] The severe recession of 1873 triggered more favorable attitudes toward patents, and in 1887, even conservative Switzerland found it prudent to pass a patent law.[8]

In the United States the patent system enjoyed widespread and persistent political support, among others, from Abraham Lincoln, who had personally patented an invention of his creation and who as an attorney in Illinois had litigated patent disputes. The public at large idolized inventors such as Thomas A. Edison and Alexander Graham Bell. Extensions over time of the Bell telephone monopoly and a cartel originally based upon the Edison electric lamp patents were sustained in a series of Supreme Court tests, reinforcing an earlier decision allowing a patent holder unilaterally to stipulate the minimum prices at which its licensees could sell its products and ignoring evidence that the patent-holder had pursued numerous parallel actions that in effect cartelized the relevant industry.[9] During the 1960s the Department of Justice sought to overturn the still-binding precedent, but was unsuccessful.[10]

In most respects, however, the tide turned again during the Great Depression of the 1930s. Growing hostility toward monopoly was precipitated by the belief that downward price rigidities enforced by monopolistic sellers (as well as by cartels authorized under President Franklin D. Roosevelt's National Recovery Administration) inhibited recovery from the depression. Threats to national security posed by patent-based cartels in tungsten carbide machine tools and synthetic rubber raised questions about the abuse of patent grants. Similar problems surfaced in the wide-ranging investigations of the Temporary National Economic Committee (TNEC), which showed inter alia how industries such as glass container-making had been thoroughly regimented through collusive control of patents by the Hartford-Empire Company. At an American Economic Association symposium reviewing the TNEC's findings, later Nobel Laureate George Stigler found the Hartford-Empire story "an eloquent example of an evil demanding correction" and concluded flatly that "[t]he case for limitation of restrictive [patent] licensing is surely irrefutable."[11]

---

7. *See* ERICH SCHIFF, INDUSTRIALIZATION WITHOUT NATIONAL PATENTS (1971).

8. ROLAND GROSSENBACHER, ANNUAL REPORT OF THE SWISS FEDERAL INSTITUTE OF INTELLECTUAL PROPERTY 8 (2007), *available at* http://www.ige.ch/e/institut/documents/i102jb07e.pdf.

9. *See* Bement & Son v. Nat'l Harrow Co., 186 U.S. 70 (1902); United States v. Gen. Elec. Co., 272 U.S. 476 (1926) (holding that since a valid patent allowed the patent holder to exclude others and hence to monopolize sale of the relevant products, licensing restraints that preserved the patent holder's monopoly reward were acceptable).

10. See, for example, United States v. Huck Mfg. Co., 382 U.S. 197 (1965), in which an attempt to overturn earlier *Bement* and *Gen. Elec.* precedents failed with a 4-4 division of Supreme Court justices.

11. George J. Stigler, *The Extent and Bases of Monopoly*, 32 AM. ECON. REV. 1, 14

Hartford-Empire was an early target of the reinvigorated antitrust enforcement paralleling the TNEC hearings. Its extensive patent agreements with other bottle-making technology providers and users were found to violate the antitrust laws. To remedy the situation, a federal district court judge ordered inter alia that Hartford-Empire and companies with which it had joined forces be required to license all their bottle-making machinery patents—after a Supreme Court intervention declaring royalty-free licensing to be confiscatory—at "reasonable" (i.e., modest) royalty rates.[12] After a subsequent Supreme Court decision stated that district court judges could exercise "judicial discretion" in formulating remedies for patent-based antitrust law violations, royalty-free licensing of General Electric's electric lamp patents was imposed.[13]

The Hartford-Empire and General Electric cases were followed by numerous antitrust settlements in which compulsory licensing of patents was ordered to remedy monopolistic situations where patents played a significant role. Between 1941 and the late 1950s, compulsory licensing decrees had been issued in settlement of more than 100 antitrust complaints, covering inter alia AT&T's transistor and other telecommunications apparatus patents, IBM's computer patents, and DuPont's nylon and other synthetic fiber patents.[14] The cumulative number of patents affected is estimated to have been between 40,000 and 50,000.[15] Although the pace abated after 1960, additional decrees covered the roughly one thousand patents in Xerox's plain-paper copying machine portfolio[16] and several pharmaceutical products. Many European nations had until recently laws allowing compulsory licensing of patents, notably, in cases where an invention was not actually produced within the patent-issuing nation. However, the cumulative number of compulsory licensing orders has seldom exceeded a dozen in the typical large European nation—a far cry from the tens of thousands of patents covered by U.S. antitrust decrees. Most of the U.S. compulsory licensing decrees were entered by mutual consent rather than as the result of fully contested litigation. Only the General Electric decree imposed royalty-free licensing through a contested court order,[17]

(Supp., June 1942). At the time, Stigler was teaching at the University of Minnesota.

12. United States v. Hartford-Empire Co., 46 F. Supp. 541 (1942), *aff'd in part, rev'd in part by* 323 U.S. 386 (1944), *aff'd by* 324 U.S. 570 (1944).

13. United States v. Gen. Elec. Co., 115 F. Supp. 835, 844 (1953).

14. STAFF OF SUBCOMM. ON PATENTS, TRADEMARKS, AND COPYRIGHTS OF THE S. COMM. ON THE JUDICIARY, 86TH CONG., REPORT ON COMPULSORY LICENSING UNDER ANTITRUST JUDGMENTS 2-5 (1960) (primarily authored by M. A. Hollabaugh & R. Wright).

15. *Id.*

16. Xerox Corp., *Decision & Order*, 86 F.T.C. 364 (1975).

17. STAFF OF SUBCOMM. ON PATENTS, TRADEMARKS, AND COPYRIGHTS OF THE S. COMM. ON THE JUDICIARY, *supra* note 14, at 5.

but several others, including the AT&T order of 1956, entailed royalty-free licensing by mutual consent.[18]

II. ECONOMIC IMPACT STUDIES

The 1956 decree ordering the compulsory licensing of roughly 8,600 AT&T patents and the nearly simultaneous decree affecting IBM patents inspired particularly intense public scrutiny. The *Wall Street Journal* observed in an editorial:

> So it may turn out that these are dangerous victories the Government boasts about. The settlements in these cases indicate a belief that everybody's patents should be everybody else's. But this is a philosophy that strikes at incentive; new ideas and new inventions may be lost. Such Government victories may turn out to be far more costly for the nation than for the companies. [19]

Shortly thereafter eight colleagues and I formed a group to meet the requirement for a "topic report" in a Harvard Business School course taught by Professor Georges F. Doriot, moonlighting president of the first modern American high-technology venture capital group, the American Research and Development Corporation. We decided to study the incentive effects of compulsory licensing decrees. We read widely in the relevant literature (aided by studies commissioned under an ongoing Senate Judiciary Committee investigation); fanned out to interview twenty-two American corporations, many of whom had entered compulsory licensing decrees; received mail questionnaires from sixty-nine companies holding 45,500 patents; and conducted an extensive statistical analysis of patenting trends. The results, privately published in two book editions,[20] were profoundly surprising to us. We discovered that with rare exceptions, whether or not well-established corporations could expect patent protection was typically unimportant in their decisions to invest in research and the development of new products and processes. "Of far greater everyday importance," we concluded, "are reward structures related to the necessity of retaining market positions, of attaining production more efficient than competitors', of securing the corporation through diversification against disastrous product obsolescence, and of gaining short-term advantages which can be

---

18. United States v. Western Elec. Co., 1956 Trade Cases (CCH) ¶ 68,246 (D.N.J. 1956).

19. *The Dangerous Victory*, WALL ST. J., Jan. 27, 1956, at 6.

20. F. M. SCHERER ET AL., PATENTS AND THE CORPORATION: A REPORT ON INDUSTRIAL TECHNOLOGY UNDER CHANGING PUBLIC POLICY (2d ed. 1959).

exploited by advertising and well-developed sales channels."[21] To be sure, there were exceptions—notably, situations in which firms were making risky investments into fields where they had little technical or marketing experience, and arguably (since our sample included no startup companies) for small new enterprises seeking a competitive foothold against well-entrenched rivals.[22] We found also from interviews, mail survey responses, and statistical analyses that prior compulsory licensing decrees had little or no unfavorable impact on research and development decisions, although they had led to less patenting of the inventions actually made and hence greater reliance on secrecy, especially on (concealable) process as distinguished from readily observed product inventions. This finding was supported in a later statistical study, conducted when company R&D spending data first became publicly available, which showed that the companies subjected to compulsory licensing decrees spent *more* on R&D relative to their sales on average than unimpacted companies of comparable size in the same fields of technology.[23]

Unaware of our study, economists at Cambridge and Oxford Universities undertook similar research on how the absence of patent protection would affect the R&D behavior of British companies. They found that across all industries covered, the weighted average reduction in R&D expenditures if all patents, anywhere in the world, were subjected to compulsory licensing with reasonable (i.e., modest) royalties, would be eight percent.[24] However, in pharmaceuticals, a negative impact of sixty-four percent was predicted.[25] Careful interviews with U.S. companies by Edwin Mansfield and colleagues revealed similar

---

21. *Id.* at 149.

22. The ambiguous situation of startup companies was characterized by the reaction of Professor Doriot when we told him about our contemplated research: "Hell, patents are simply instruments with which big companies bludgeon my startups" (conversation with author). See SPENCER E. ANTE, CREATIVE CAPITAL: GEORGES DORIOT AND THE BIRTH OF VENTURE CAPITAL (2008), for a biography of Doriot.

23. F. M. Scherer, *The Economic Effects of Compulsory Patent Licensing*, NEW YORK UNIVERSITY MONOGRAPH SERIES IN FINANCE AND ECONOMICS 67-75 (1977). However, in The Incentive Theory of Patents in Action: The Effects of Patent Relief on the Incentive to Invest and the Incentive to Disclose (September 2005) (unpublished S.J.D dissertation, Harvard Law School) (on file with the author), Ziv M. Preis examines the effects of Federal Trade Commission consent decrees involving patents—90 percent of which accompanied merger case settlements—between 1980 and 1999. The results vary widely, but in some analyses, high-impact compulsory licensing decrees are found significantly to reduce R&D/sales ratios in the few years following, after which a reversal is typically observed. The analysis makes no attempt to control for merger effects per se (i.e., a high R&D firm acquiring a low R&D firm), as contrasted to the effect of compulsory licensing in the decrees under which mergers were allowed to be consummated.

24. C. T. TAYLOR & Z. A. SILBERSTON, THE ECONOMIC IMPACT OF THE PATENT SYSTEM 199 (1973).

25. *Id.*

disparities between the incentive effect of patents in pharmaceuticals and other high-technology industries.[26]

Many surveys have shown that the expectation of patent protection is much more important to investment in pharmaceutical R&D than in most industries. Drug R&D comes closest to what economists call the generation of knowledge as a pure public good. Most of the expenditure is directed toward finding molecules that might have interesting therapeutic action in human beings and then, through costly clinical trials, ascertaining that the target molecule is really effective and safe.[27] Absent patents, once that evidence has been amassed, it might be available for any and all would-be generic imitators to exploit. All that may be needed for the free-rider (or more accurately, cheap rider) is to spend a sum on process engineering (tiny relative to the amounts spent on discovery and testing), whereupon a competing molecule can be marketed, if regulatory rules permit. However, further research added a caveat to this conclusion and clarified the role of what came to be known as "first mover" advantages as a barrier to rapid new product imitation and hence as a substitute for patent protection. Comparing side-by-side two pharmaceutical molecules, one unpatentable and one patented, Bond and Lean found that the erosion of the pioneer's price premium and market share was as slow for the unpatented product as for the patented product.[28] The reason, it became clear, was that being the first successfully to market a consumer product affixes in the mind of would-be purchasers an image of superiority and reliability that is hard for latecomers to surmount, whether the product is patented or not.[29] However, it should be noted that the Bond and Lean study focused on products developed during the late 1950s, when regulatory strictures were more lax and the research and testing costs required to market a successful new drug entailed only about $1 million. By the late 1990s, the comparable costs had mounted to hundreds of millions of dollars, while the costs of engineering imitative generic products rose much less.[30]

Four prominent economists at Yale University took a major step toward confirming the role hoped-for patent protection plays in R&D

26. Edwin Mansfield, *Patents and Innovation: An Empirical Study*, 32 MGMT. SCI. 173 (1986); Edwin Mansfield et al., *Imitation Costs and Patents: An Empirical Study*, 91 ECON. J. 908 (1981).

27. For a survey, see F. M. SCHERER, INDUSTRY STRUCTURE, STRATEGY, AND PUBLIC POLICY 357-62 (1996).

28. RONALD S. BOND & DAVID LEAN, F.T.C. STAFF REPORT, SALES, PROMOTION, AND PRODUCT DIFFERENTIATION IN TWO PRESCRIPTION DRUG MARKETS (1977)*; see also* William D. Robinson & Claes Fornell, *Sources of Market Pioneer Advantages in Consumer Goods Industries*, 22 J. MKTG. RES. 305 (1985).

29. *Id.*

30. *See* Joseph A. DiMasi et al., *The Price of Innovation: New Estimates of Drug Development Costs*, 22 DRUG INFO. J. 151, 151-85 (2003).

decisions.[31] They obtained elaborate survey responses from 650 U.S. R&D managers.[32] One set of questions, emulating earlier inquiries for a smaller sample by Mansfield, asked how much R&D, measured relative to the first mover's R&D, would be needed to duplicate the first mover's innovation. For major patented new products, the average fraction was roughly 85 percent (weighting category ranges by response rates); for major unpatented products, 65 percent.[33] Thus, patent protection raised imitation costs, but even without it, imitators could not simply "free-ride" on the innovator's work. The Yale group also asked respondents to rank on a scale of 1 ("not at all effective") to 7 ("very effective") the extent to which various instruments protected the competitive advantages from new and improved products and processes.[34] The average scores across 130 industrial lines on the effectiveness of various means to reap the economic benefits of new and improved products were as follows:[35]

Table 1: Average Scores

Method of Appropriation	Score
Secrecy	3.57
Patents to secure royalty income	3.75
Patents to prevent duplication	4.33
Moving quickly down learning curves	5.09
Being first with an innovation	5.41
Superior sales or service efforts	5.59

Having patent protection was found on average to be relatively unimportant compared to three other ways of gaining first mover advantages. For new and improved processes, it was even less important on average, while, not surprisingly, secrecy was ranked more highly than either of the patent measures. There were, to be sure, exceptions. Among seventy-seven industry groups with three or more responses, the pharmaceuticals industry ranked duplication-preventing patents as the most important means of holding off imitative competition, second in average score only to the agricultural chemicals field (with environmental effect test regulations similar to those imposed for pharmaceutical efficacy and safety).[36]

31. Richard C. Levin, Alvin Klevorick, Richard R. Nelson, & Sidney Winter, *Appropriating the Returns from Industrial Research and Development*, 18 BRKGS. PAPERS ON ECON. ACTIVITY 783, 790 (1987).
32. *Id.*
33. *Id.* at 819-20.
34. *Id.* at 792.
35. *Id.* at 800.
36. *Id.* at 816-17.

Generally similar responses were obtained in an even larger Carnegie-Mellon University survey administered in 1994, to which more than a thousand industrial laboratory managers responded.[37] Using a different scale than the Yale survey, respondents were asked on what percentage of their product innovations various means of appropriating inventions' profit potential were effective.[38] Patent protection had the second lowest average score of 34.83 percent, undercut only by "other legal" mechanisms.[39] Lead time was viewed as the most important means, with an average score of 52.76 percent.[40] Secrecy received much higher weight than in the Yale survey, with a 51 percent average, followed by complementary manufacturing capabilities (46 percent) and complementary sales and service efforts (43 percent). As in the Yale survey, patents received an unusually high score in pharmaceuticals, second only among 34 broad industry categories to medical equipment (ranging from catheters to imaging systems). Cohen et al. conclude that patents are only one piece of a broader strategy to protect inventions, cautioning, as other studies did, that situations exist, even in industries according only modest weight to patent protection, in which at the margin patents are decisive in inducing R&D investments.[41]

Important lessons emerge from these queries addressed to real-world managers. First, alternative barriers to rapid imitation—the substantial R&D costs imitators have to incur, lags in recognizing opportunities, image and cost advantages accruing to the first mover, and the like, leave a substantial class of cases in which would-be innovators can anticipate revenue gains exceeding their innovation and production costs even when patent protection is totally absent. Second, given that non-patent stimuli to innovation exist, established firms are driven to undertake their own innovation efforts for fear of being overtaken by more aggressive rivals. This is the Schumpeterian "creative destruction" effect.[42] Third, patent protection does substantially enhance profit expectations in some industries—e.g., much more so in industries with characteristics such as pharmaceuticals than in semiconductors or computers, with more complex, multifaceted products. Fourth, there may

37. Wesley M. Cohen, Richard R. Nelson, & John Walsh, *Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)* 4 (Nat'l Bureau of Econ. Research Working Paper No. 7552, Feb. 2000), *available at* http://www.nber.org/papers/w7552.pdf.

38. *Id.* at 5.

39. *Id.* at tbl.1.

40. *Id.*

41. *Id.; see also* Ashish Arora, Marco Ceccagnoli, & Wesley M. Cohen, *R&D and the Patent Premium* (Nat'l Bureau of Econ., Research Working Paper No. 9431, Jan. 2003), *available at* http://papers.nber.org/papers/w9431.

42. *See* JOSEPH A. SCHUMPETER, CAPITALISM, SOCIALISM, AND DEMOCRACY (1942), especially Chapter VII.

be feedback effects from patent protection to Schumpeterian creative destruction. Patent protection may help trigger an upstart firm's innovation that threatens established firms, but to the extent that it lessens the threat to established firms, it weakens their incentives to maintain a vigorous innovative pace.

These lessons appear to have trickled out at best slowly to the legal and policy-formulating communities. One might have expected the findings to have been especially relevant to legal scholars. However, a search of *Social Sciences Citation Index* for 1987 through May 2006 revealed that only 11 percent of the 496 citations received by the principal Levin et al. paper—the most acclaimed of the various patent survey reports, and with an appropriately high citation count—were in legal journals.[43]

The diffusion to economists also left something to be desired. Beginning in the early 1980s, there was an explosion of theoretical work on the economics of the patent system.[44] However, nearly all of the theoretical contributions assumed—contrary to the empirical evidence—that patent protection was the only or principal barrier to rapid imitation of an invention or innovation.[45] Clearly, economists were delinquent in providing an adequate theoretical basis for policy reforms.

III. THE IMPETUSES TO POLICY CHANGE

During the 1970s, new initiatives for patent policy change began accelerating in the United States. One might ascribe the changes to the cyclical character of patent policy change observed in the historical past, or to the increased susceptibility of the U.S. government to interest group lobbying. On the latter we shall have more to say later. There was, however, another impetus on the macroeconomic front.

In 1969, productivity—output per hour of labor input—in the nonfarm business sector of the U.S. economy stagnated and then entered a period of significantly diminished annual growth. By 1980, productivity

43. Social Science Citation Index, Citation Summary for Appropriating the Returns from Industrial Research and Development, http://www.thomsonreuters.com/products_services/scientific/Web_of_Science (search for author "Levin" and publication "Brookings*"; then follow hyperlink "APPROPRIATING THE RETURNS FROM INDUSTRIAL-RESEARCH AND DEVELOPMENT").

44. *See* F.M. Scherer, Patents: What Do We Know; What Must We Learn? (1996) (in the proceedings of a conference in Luxembourg on Appropriability and Patent Value: Econometric Aspects) (on file with the author), which shows that the number of articles covered by the ECONLIT bibliography with "patent" or some compound thereof in their titles rose from an average of four per year between 1969 and 1982 to 23 per year between 1984 and 1995.

45. An exception is Rufus Pollock, *Innovation and Imitation with and without Intellectual Property Rights*, (Cambridge University, MPRA Working Paper No. 5025, Sept. 2006), *available at* http://mpra.ub.uni-muenchen.de/5025/.

was 16 percent less than it would have been had it continued the 2.46 percent annual growth rate it experienced from 1947 through 1969.[46] By 1985, the shortfall was 20 percent.[47] Also, company-financed R&D expenditures by U.S. industry, adjusted for general inflation, experienced the first break from a rising trend since the collection of statistics was initiated beginning with the year 1950.[48] Further year-to-year declines occurred, and even in the good years growth was slower, so that by 1981, a 28 percent shortfall had accumulated.[49] Research by David Ravenscraft and myself tapping data from a small but unusually detailed sample of company business units revealed that the decline in R&D spending was probably attributable to a drop in the profitability of R&D investments, and when R&D was cut back, its profitability rose again, precipitating new growth.[50]

Two seminal papers published simultaneously in 1967 showed that, contrary to conventional wisdom among economists, the United States could attribute much of its comparative advantage in international trade to superior technological innovation.[51] As the industrial nations of Western Europe and especially Japan recovered fully from the devastation of World War II, however, they began aggressively to challenge U.S. corporations for technological leadership.[52] In 1975, U.S. exports of high-technology goods exceeded imports by a ratio of 2.4 to 1.[53] By 1980, the ratio had declined to 1.95 to 1 and by 1985 to 1.05 to 1.[54] The first reaction of U.S. industries to high-technology challenges from abroad was on average what the theory of arms races calls "submissive," i.e., a relative decline in R&D outlays. Some industries such as integrated steel, automobile tires, and television sets essentially

---

46. Computed from ECONOMIC REPORT OF THE PRESIDENT 328 (1995), with earlier data spliced from the same report for 1980. *See* ECONOMIC REPORT OF THE PRESIDENT 246 (1980).

47. ECONOMIC REPORT OF THE PRESIDENT 338 tbl.B46 (1991).

48. *See* F. M. Scherer, *R&D and Declining Productivity Growth*, 73 AMER. ECON. REV. Supp. 215 (1983).

49. *Id.*

50. David J. Ravenscraft & F. M. Scherer, *The Lag Structure of Returns to R&D*, 14 APPLIED ECON. 603 (1982). For similar results with the pharmaceutical industry, see F. M. Scherer, *The Link Between Gross Profitability and Pharmaceutical R&D Spending*, 20 HEALTH AFFAIRS 216 (2001).

51. William Gruber, Dileep Mehta & Raymond Vernon, *The R&D Factor in International Trade and International Investment of United States Industries*, 75 J. POL. ECON. 1, 20 (1967); Donald B. Keesing, *The Impact of Research and Development on United States Trade*, 75 J. POL. ECON. 1, 38 (1967).

52. For statistical analyses and eleven case studies, see F.M. SCHERER, INTERNATIONAL HIGH TECHNOLOGY COMPETITION (1992).

53. *Id.* at 4 fig.1.2.

54. U.S. NATIONAL SCIENCE BOARD, SCIENCE & ENGINEERING INDICATORS 379 (1989). Later editions of the same report suggest a more modest decline because of a redefinition of what constituted high technology industries.

gave up. But others such as the producers of integrated circuits, medical imaging apparatus, optical fiber cables, earth-moving equipment, and (less unambiguously) airliners responded aggressively and redoubled their R&D efforts to retain or regain their world market positions.

It was argued, among other fora in Congressional hearings, that patent policy reforms could help restore U.S. technological leadership. Perhaps, but the chains of causation were clearly more complex.[55] Reductions in corporate R&D spending were precipitated by a fall in profitability. If stronger patent protection could restore profitability, it might facilitate a resurgence. And it was true that the most formidable new rival to U.S. technological leadership, Japan, maintained a much weaker patent system, among other things requiring the licensing of most patents and limiting through foreign exchange controls the royalties Japanese firms could pay U.S. patent holders.[56] But the exercise of patent rights within the United States did blunt some Japanese competition, e.g., in optical fibers and integrated circuits.

Alternatively, however, the profits from innovation may have declined because the pool of attractive technological opportunities had been depleted following intensive "fishing" during the decades following World War II. In this sense, the productivity growth slump that began around 1969 was an extension of the so-called Kondratief cycles emphasized by Joseph A. Schumpeter in a 1939 classic.[57] Industrial research and development efforts were intensified in those industries that elected to fight back against tougher foreign competition.[58] But more importantly, growth was restored, sometimes with long lags, as a result of fundamental scientific and technological breakthroughs that underlay the information and biotechnology revolutions of the 1990s and the early

55. For similar arguments, see Richard Posner, *The Insignificance of Macroeconomics in Patent Antitrust Law: A Comment on Millstein*, 9 CARDOZO L. REV. 4, 1203 (1988). The paper on which Posner commented, by Ira Millstein, chief counsel at the time to the influential Business Roundtable, considered studies such as those by Levin et al., *supra* note 31, "inconclusive" and argued (fallaciously) that the effects of non-patent barriers "do not make the patent a less significant inducement." Ira Millstein, *The Role of Antitrust in an Age of Technology,* 9 CARDOZO L. REV. 1175, 1185 (1998).

56. *See* DANIEL OKIMOTO, BETWEEN MITI AND THE MARKET 27-28 (1989); Janusz Ordover, *A Patent System for Both Diffusion and Exclusion*, 5 J. ECON. PERSPECTIVES 43 (1991).

57. JOSEPH A. SCHUMPETER, BUSINESS CYCLES (1939). For the most persuasive empirical support, see ALFRED KLEINKNECHT, INNOVATION PATTERNS IN CRISIS AND PROSPERITY: SCHUMPETER'S LONG CYCLE RECONSIDERED (1987). For an analysis from the 1970s and 1980s skeptical of the general depletion hypothesis, see MARTIN N. BAILY & ALOK CHAKRABARTI, INNOVATION AND THE PRODUCTIVITY CRISIS (1988). For theoretical support rooted in the logic of highly skewed payoff distributions, see William D. Nordhaus, *Alternative Approaches to the Political Business Cycle*, 20 BROOKINGS PAPERS ON ECON. ACTIVITY 1 (1989).

58. SCHERER, *supra* note 52, at ch.5.

21st century—notably, the invention of integrated circuits around 1959 and microprocessors in the early 1970s and the steady cost declines that occurred through learning-by-doing and denser circuit-packing; the laser in the late 1950s and optical fiber data transmission during the 1970s; and gene splicing during the early 1970s. Patents played some role in all of these achievements, but given uncertainties, long lags, and the university origins of key breakthroughs, hardly a precipitating role. The Department of Defense insisted upon widespread licensing of integrated circuit patents, and several early developers of microprocessors cross-licensed their patents among one another and to other chip makers.[59] A small fortune was made through broad-based licensing of basic laser patents by the winner of a law suit claiming priority of invention, but only after litigation delays of more than two decades.[60] From a beginning in 1980, the Cohen-Boyer gene splicing patents were licensed at modest royalties to hundreds of entities by Stanford University and the University of California, yielding cumulative total royalties to the two universities of some $124 million by 1995.[61]

IV. HOW PATENT POLICY WAS CHANGED

We turn now to our analysis of the principal changes in U.S. patent policy, focusing mainly on events of the late 1970s and early 1980s.

A. Copyright Law

Changes in copyright law may have been precursors to what happened on the patent front, so a brief look is warranted. As of 1962, the life of a copyright was limited to 28 years, with one 28-year renewal to a maximum of 56 years allowed.[62] Then, in the four decades that followed, Congress extended copyright lives eleven times, so that by the turn of the century, works were copyrighted for 70 years beyond the life span of the copyrighted work's creator.[63] In 1976, copyright extensions

59. Texas Instruments later collected an estimated $1 billion in royalties on its integrated circuit patents until it lost key lawsuits in Japan and the United States. *See* Norm Alster, *New Profits from Patents*, FORTUNE, Apr. 25, 1988, at 185; *When Copying Gets Costly: Intellectual Property*, ECONOMIST, May 9, 1992, at 95; *Chip Patent Suit by Texas Instruments*, N.Y. TIMES, June 30, 1992, at D2; Edmund L. Andrews, *Texas Instruments Loses in Japanese Ruling*, N.Y. TIMES, Sept. 1, 1994, at D3.

60. *See* NICK TAYLOR, LASER: THE INVENTOR, THE NOBEL LAUREATE, AND THE THIRTY-YEAR PATENT WAR (2000).

61. Sheryl Winston Smith, The Cohen-Boyer Patent: A Case Study (1996) (unpublished manuscript, on file with author) (based on a phone conversation with Floyd Grolle, Manager, License Administration, Office of Technology Licensing, Stanford University; Dec. 15, 1995).

62. 17 U.S.C. § 24 (1962).

63. LAWRENCE LESSIG, FREE CULTURE 134 (2004); Kevin Kelly, *Scan This Book!*, N.Y. TIMES, May 14, 2006, § 6 (Magazine), at 48.

were made automatic, without the need to apply or register. According to Kevin Kelly, these changes occurred as an increasing number of creative works came to be owned not by individuals but by corporations able successfully to lobby Congress to prevent materials from returning to the public domain.[64] Or as Lawrence Lessig concludes, "The law speaks to ideals, but it is my view that our profession has become too attuned to the client. And in a world where the rich clients have one strong view, the unwillingness of the profession to question or counter that one strong view queers the law."[65]

B. *Patents from Government-Supported Research*

World War II and its aftermath, including the cultivation of basic science through the National Science Foundation and the development of radar and atomic energy, brought the U.S. federal government into extensive technological cooperation with private industry and universities. Who should have primary rights to patents resulting from government-financed R&D was a question settled in a diversity of inconsistent ways. Some clarity was brought through a policy statement issued by President John F. Kennedy in 1963,[66] but debate continued. In 1965 an inter-agency task force, the Committee on Government Patent Policy, operating under the auspices of the Federal Council for Science and Technology, undertook an ambitious empirical study of how the various patent policies were working.[67] It hired a consulting firm, Harbridge House, to compile data on 2,024 patents made under government contracts and several hundred more originating in government laboratories, and to conduct a series of historical case studies on attempts to bring inventions conceived with government financial support into private-sector utilization.[68] Harbridge House completed several interim volumes and, in May 1968, a four-volume compendium of research findings.[69] The Committee on Government Patent Policy published its own report and patent policy recommendations in the fall of 1968[70] and presented them at a briefing conference before the Federal

64. Kelly, *supra* note 63.

65. Lessig, *supra* note 63, at 304.

66. Memorandum for the Heads of Executive Departments and Agencies, 28 FED. REG. 10,943 (Oct. 12, 1963).

67. Memorandum about Government Patent Policy, 36 FED. REG. 16889 (Aug. 23, 1971), *reprinted in* BACKGROUND MATERIALS, *infra* note 69, vol. I at 11-23.

68. HARBRIDGE HOUSE, *infra* note 69, at I-17.

69. HARBRIDGE HOUSE, INC., GOVERNMENT PATENT POLICY STUDY, published in loose-leaf binder form, May 1968, *reprinted in* STAFF OF H.R. COMM. ON SCIENCE AND TECHNOLOGY, BACKGROUND MATERIALS ON GOVERNMENT PATENT POLICIES, Vol. II (August 1976), at 69-140.

70. It is reproduced in BACKGROUND MATERIALS, *supra* note 69, vol. II at 143-82. I

Bar Association in September 1969. The Committee's recommendations, which emphasized flexibility in allowing contractors to obtain exclusive patent rights mainly when there were prospects of commercial utilization or when granting exclusive rights broadened the government's potential contractor base, formed the basis for a new policy statement issued by President Nixon in August 1971.[71]

The Harbridge House research revealed that several variables affected the likelihood that government contract-originated inventions would be commercially utilized: (1) the intrinsic relevance of the technology to civilian needs; (2) whether the contractor had prior commercial experience in the relevant field; (3) how far the development had been carried under contract; (4) the magnitude of additional development outlays required in comparison to the market size and the risks attendant thereto; and (5) whether or not the contractor or another assignee had exclusive patent rights. For 1,720 patents on which complete data were available, commercial utilization rates varied over two key variables as follows:[72]

Table 2: Commercial Utilization Rates

	Contractor Had Prior Commercial Experience	Contractor Without Prior Commercial Experience
With exclusive rights	23.8%	6.6%
Without exclusive rights	13.3%	2.2%

Evidently, patent protection mattered, although the chain of causation remained ambiguous. In some cases, the qualitative studies showed, exclusive rights encouraged investments in commercial utilization; in others, contractors bargained more vigorously to obtain exclusive rights when commercial utilization was expected.

The pharmaceutical industry was found again to be an extreme case. One in-depth Harbridge House study revealed that, up to 1962, drug companies routinely screened new organic molecules synthesized under government grants by academic researchers.[73] However, when the Department of Health, Education, and Welfare (HEW) imposed new

served as principal economic adviser to the Committee throughout the Harbridge House study period.

71. Memorandum about Government Patent Policy, *supra* note 67.

72. This analysis is drawn from Scherer, *supra* note 23, at 78-84.

73. HARBRIDGE HOUSE, INC., REPORT ON EFFECTS OF GOVERNMENT PATENT POLICY ON DRUG RESEARCH AND NEW PRODUCT DEVELOPMENT secs.I & IV (1967).

reporting requirements that threatened the exclusivity of drug companies' rights to commercialize molecules found to be therapeutically interesting, such testing ceased abruptly. The moratorium ended in 1968 when HEW changed its policies to allow drug companies exclusive rights on grant-originated molecules they tested.[74]

A particularly controversial question at the time was whether, when a government agency allowed its contractors to obtain exclusive patent rights, the government should retain "march-in" rights to require wider licensing of the patent if there was a failure to commercialize or there were monopolistic abuses in commercialization. Cases of clear abuse were found to be rare, in all but one questionable instance, because adequate substitute products existed. Both the Committee on Government Patent Policy and the Nixon memorandum[75] recommended retention of march-in rights, to be used flexibly and presumably rarely under an implicit rule of reason, or in cases of jeopardy to public health or safety.

The U.S. Congress chose in due course to insert its own views into the debate. In 1965 S. 1809, embodying compromise policies, was approved by the Senate Judiciary Committee, but in 1967 its consideration by the full Senate was postponed indefinitely pending completion of the Harbridge House Study.[76] A draft bill was proposed to Congress by the White House in August 1976, supplanted by a bill drafted in the House of Representatives.[77] Hearings in 1976 before the House Committee on Science and Technology summoned as witnesses the executive secretary of the Committee on Government Patent Policy and others affiliated with it along with representatives of the principal government R&D contract-issuing agencies, industry, and an organization comprising university patent administrators.[78] The Harbridge House report summary and related documents were published as background materials. No legislation ensued at first, but in subsequent sessions of Congress, further hearings were held by the House Science Committee as well as the Monopolies subcommittee of the House Judiciary Committee. The latter hearing, in December 1977, added substantive balance, inviting as witnesses inter alia outspoken Admiral

74. *See* DAVID H. GUSTON AND DANIEL R. SAREWITZ, SHAPING SCIENCE AND TECHNOLOGY POLICY 62 (2006).

75. Memorandum about Government Patent Policy, *supra* note 67.

76. Howard Forman, *Retrospection and Introspection Concerning Patents and Government Patent Protection*, 49 J. PAT. OFF. SOC'Y 678 (1967).

77. They are reproduced in FEDERAL COUNCIL FOR SCIENCE AND TECHNOLOGY, REPORT ON GOVERNMENT PATENT POLICY, 88-119 (1976).

78. *Government Patent Policy: The Ownership of Inventions Resulting from Federally Funded Research and Development: Hearing Before the Subcomm. on Domestic and Int'l Scientific Planning and Analysis of the H. Comm. On Sci. and Tech.*, 94th Cong. 12 (1976).

Hyman Rickover (father of the Navy's nuclear submarine program), Walter Adams (an economist well-known for his anti-monopoly views), and the consumer activist chairman of the Federal Trade Commission.[79]

After characteristic delays, two major bills emerged from the effort: the Bayh-Dole Act, signed into law in December 1980;[80] and the Stevenson-Wydler Act, passed in October 1980.[81] The floor debates were brief, and both bills sailed through Congress (controlled in both houses by Democrats) on voice votes. Bayh-Dole reversed the prevailing but flexible presumption that the government would retain title to inventions made under R&D contracts. It articulated a presumption that government contracts or grants to academic researchers or small businesses would normally permit patent rights to be retained by the contractors, subject to march-in under imprecisely articulated conditions. A 1987 executive order extended the presumption to apply to all government R&D contract recipients, regardless of their size.[82] Stevenson-Wydler required the principal government agencies conducting R&D in-house to set up Research and Technology Applications offices. Since "the whole point of [the] bill [was] to stimulate the commercialization of industrial innovations," as one Congressional proponent observed in the final debate,[83] the offices were encouraged to negotiate exclusive patent licenses with industry for inventions resulting from agency research. In 1986, the Federal Technology Transfer Act extended Stevenson-Wydler to permit formation of cooperative research and development agreements (CRADAs) between government laboratories and industry, with the industrial partners retaining principal patent rights but paying royalties to cooperating agencies and their inventor employees.[84]

---

79. *Government Patent Policies: Hearings Before the Subcomm. on Monopoly and Anticompetitive Activities of the Select Comm. on Small Business*, 95th Cong. (Dec. 1977).

80. Bayh-Dole Act, Pub. L. No. 96-517, 94 Stat. 3019 (1980) (codified as amended at 35 U.S.C. §§ 200-212 (2006)).

81. Stevenson-Wydler Act, Pub. L. No. 96-480, 94 Stat. 2311 (1980) (codified as amended at 15 U.S.C. §§ 3701-3717 (2006)).

82. Exec. Order No. 12591, 52 FR 13414 (1987) (one purpose of which is to "promote the commercialization . . . of patentable results of federally funded research by granting to all contractors, regardless of size, the title to patents made in whole or in part with Federal funds, in exchange for royalty-free use by or on behalf of the government . . ..").

83. 126 CONG. REC. H. 24,565, 24,565-67 (daily ed. Sept. 8, 1980) (comments of Mr. Hollenbeck in support of S. 1250, the Stevenson-Wydler Technology Act of 1980).

84. Federal Technology Transfer Act, Pub. L. No. 99-502, 100 Stat. 1785 (1986) (codified as amended 15 U.S.C. 3710 (2006)) (amending Stevenson-Wydler Technology Innovation Act, Pub. L. No. 96-480, 94 Stat. 2311 (1980)). No explicit provisions were included on march-in rights. The FTAA is ambiguous on whether the waiver of federal rights exhausts the possibility of march-in for non-governmental uses, saying only that

> [A] Government-operated Federal laboratory may…waive, subject to reservation by the Government of a nonexclusive, irrevocable, paid-up license to practice the

These legislative patent policy changes had important implications. Academic institutions in particular changed their behavior. Many which had not done so already created technology licensing offices to encourage patenting of relevant inventions by faculty researchers. University patenting rose sharply—from an average of 332 patents received per year during the last three years of the 1970s to 952 per year in the last three years of the 1980s.[85] At least part of the increase appears to have been caused by the imposition of lower standards on the patents sought. There was a marked decline in the number of subsequent citations received by the average university patent following the law change.[86] Links between university researchers and their industry counterparts increased in number and intensity, with an undoubted positive impact on the commercialization of academic research, especially in the field of biotechnology. Whether academic research as a result has been diverted at least marginally from basic to more applied goals and whether discoveries are disclosed more slowly so as not to jeopardize patentability is less than certain. To the extent that such consequences have followed, their desirability continues to be debated.[87]

Especially in academic circles, but also on inventions made cooperatively with government laboratories, serious questions have arisen over the resulting product prices. As we have seen, patents are of special importance to pharmaceutical (and related biopharmaceutical) companies, in part because they provide strong protection from competitive imitation on products that often have relatively inelastic demands. This means that high prices can be commanded. AZT (azidothymidine), the first antiretroviral effective against AIDS, was synthesized by a medical institute researcher with federal research support.[88] After the unpatented molecule was offered to the National Institutes of Health by the private firm Burroughs-Wellcome, its therapeutic efficacy was demonstrated in clinical trials conducted initially

---

> invention or have the invention practiced throughout the world by or on behalf of the Government, in advance, in whole or in part, any right of ownership which the Federal Government may have to any subject invention made under the agreement by a collaborating party or employee of a collaborating party . . . .

Pub. L. No. 99-502, § 2(b)(3).

85. *See* Rebecca Henderson, Adam Jaffe, & Manuel Trajtenberg, *Universities as Sources of Commercial Technology: A Detailed Analysis of University Patenting*, *1965-1988*, *in* PATENTS, CITATIONS, AND INNOVATIONS 237, 254-55 (Adam B. Jaffe & Manuel Trajtenberg eds., 2005), *available at* http://www.mitpressjournals.org/doi/pdfplus/10.1162/003465398557221.

86. *See id.*

87. *See, e.g.*, DEREK BOK, UNIVERSITIES IN THE MARKETPLACE: THE COMMERCIALIZATION OF HIGHER EDUCATION 10-12, 140-43 (2003).

88. This discussion benefits from a case study. *See* Kris Thiessen, AZT: A Favored Orphan? (1998) (unpublished manuscript, on file with the John F. Kennedy School of Government, Harvard University).

at NIH and Duke University with significant support from federal government funds.[89] Burroughs-Wellcome was able to obtain "method of use" patents covering AZT along with exclusive marketing rights reflecting AZT's early "orphan drug" status.[90] It chose to sell AZT at annual costs per patient approximating $10,000[91] when production costs could not have been more than $2,000.[92] This pricing strategy provoked outrage among AIDS advocates and members of Congress and elicited demands that the National Institutes of Health exercise their march-in rights to require the issue of non-exclusive patent licenses. That was not done, but Burroughs-Wellcome eventually implemented substantial price reductions in response to the public pressure. Several other drugs conceived or developed with federal government support have had similar high-price histories.[93] What could have been the most egregious case was thwarted by a judicial finding of patent invalidity after the University of Rochester sought royalties it expected to reach $3 billion from its work, supported by National Institutes of Health grants, underlying the development of Cox-2 inhibitors.[94]

The National Institutes of Health directorate has declined to exercise its Bayh-Dole march-in rights on patents covering drugs sold at particularly high prices. Indeed, as of 2005, the march-in provision had never been invoked by a government agency.[95] There appear to be two main reasons. For one, the statutory text left ambiguities. The relevant march-in clause states in part that the granting agency has the right to

---

89. *Id.* at 5.

90. Treatment of Human Viral Infections, U.S. Patent No. 4,724,232 (filed Aug. 21, 1986).

91. KATHERINE FLOYD & CHARLES GILKS, COST AND FINANCING ASPECTS OF PROVIDING ANTI-RETROVIRAL THERAPY: A BACKGROUND PAPER (1989), http://www.worldbank.org/aidsecon/arv/floyd/whoarv-let.pdf (citing RENEE SABATIER, MARTIN FOREMAN, JON TINKER & MARTY RADLETT, AIDS AND THE THIRD WORLD (1989) for the statement that in the "late 1980s . . . annual per patient cost of AZT was reported to be US$ 10 000.").

92. Stephanie Lucchini et al., *Decrease in Prices of Antiretroviral Drugs for Developing Countries: from Political "Philanthropy" to Regulated Markets?*, *in* ECONOMICS OF AIDS AND ACCESS TO HIV/AIDS CARE IN DEVELOPING COUNTRIES: ISSUES AND CHALLENGES figs. 3-8 (Jean-Paul Moatti et al., eds. 2003).

93. *See id.*

94. *University's Patent for Celebrex Is Invalid*, N.Y. TIMES, Feb. 14, 2004, at C3. *See* Univ. of Rochester v. G.D. Searle Co., 358 F. 3d. 916 (Fed. Cir. 2004), *cert. denied* 543 U.S. 1015 (2004). See also Nicholas M. Ciarelli, *Jury Rules Company Infringed Drug Patent*, HARVARD CRIMSON, May 5, 2006, reporting on a Federal District Court finding in favor of royalties for a fundamental biological pathways discovery by Harvard University researchers licensed to a biotech company. The case was Ariad Pharm., Inc. v. Eli Lilly Co., 2005 U.S. Dist. LEXIS 10941 (D. Mass. 2005).

95. *See* John H. Raubitzchek & Norman J. Latker, *Reasonable Pricing – A New Twist For March-In Rights Under The Bayh-Dole Act*, 22 SANTA CLARA COMPUTER & HIGH TECH. L.J. 149, 155 (2005).

compel issuance of non-exclusive licenses when:

> (1) . . . [T]he contractor or assignee has not taken . . . within a reasonable time, effective steps to achieve practical application of the subject invention. . . [or]
> (2) [A]ction is necessary to alleviate health or safety needs which are not reasonably satisfied by the contractor, assignee[s], or their licenses.[96]

Debate centers on the meaning of the "reasonably satisfied" provision. In response to a critical article in *The Washington Post,*[97] the Bayh-Dole Act's co-sponsors insisted that the march-in rights are not contingent upon the pricing of a resulting product or the profitability of the commercializing company, but they can be invoked only "when the private industry collaborator has not successfully commercialized the invention as a product."[98] This seems an unreasonable interpretation of subparagraph (2) above even if not (1), but on such fuzzy constructs, reasonable people can disagree. Also, the National Institutes of Health, which have been a focal point of march-in rights conflict, have been reluctant to serve as a judge of whether product prices are reasonable, viewing such decisions as the province of Congress or the antitrust agencies.[99]

C. *A Special Court for Patent Appeals*

The status quo as the 1970s began was for patent case decisions at the Federal district court level to be appealed to any of the ten regional appellate courts, while appeals from decisions of the U.S. Patent and Trademark Office went to a special Court of Customs and Patent Appeals, sitting in Washington, D.C. There was considerable discontent over conditions in the appellate courts. Quite generally, an increased number of appeals with little expansion in the number of judges led to a perceived overload situation. Patent cases, which amounted to from one

---

96. 35 U.S.C. §§ 203(a)(1)-(2) (2006).

97. Peter Arno & Michael Davis, *Paying Twice for the Same Drugs*, WASH. POST, Mar. 27, 2002, at A21. For an extended analysis, see Peter Arno & Michael Davis, *Why Don't We Enforce Existing Drug Price Controls? The Unrecognized and Unenforced Reasonable Pricing Requirements Imposed Upon Patents Deriving in Whole or in Part From Federally Funded Research*, 75 TUL. L. R. 631 (2001).

98. Birch Bayh & Bob Dole, *Our Law Helps Patients Get New Drugs Sooner*, WASH. POST, Apr. 11, 2002, at A28.

99. For an explicit decision to this effect in a particularly egregious case—a fivefold increase in the price of an anti-AIDS drug that had already been marketed for seven years – see National Institutes of Health, Office of the Director, In the Case of NORVIR® Manufactured by Abbott Laboratories, Inc. (July 29, 2004), http://www.ott.nih.gov/policy/March-in-norvir.pdf.

to three percent of all decentralized appeals,[100] were only a small part of the problem, although it was said (without clear quantitative evidence) that patent appeals were more complex than the average appeal. Patent advocates were unhappy over what they claimed to be wide differences in the outcomes of their appeals, allegedly because some appellate courts took a tougher line toward the validity of challenged patents, and on whether patents passing the validity screen were actually infringed, than others. This was said to have led to "forum shopping"—patent owners sought venue in appellate courts friendly toward patent protection while alleged infringers sought more skeptical courts. Differences between courts in legal precedents were also an alleged problem, and inter-court differences were seldom carried to the Supreme Court for resolution. Patent advocates sought a unified appellate venue that would minimize forum-shopping and generate consistent precedents.

Appellate court reform questions were addressed repeatedly by diverse study groups. One of the most thorough was the so-called Hruska Commission, chaired by Senator Roman Hruska, which delivered its conclusions in 1975.[101] It favored creation of a new nationwide appellate court to which matters that posed important precedential questions (including patent cases) would be *transferred* at the behest of the normal appellate courts, which would retain jurisdiction over most patent appeals from federal district courts.[102] Or alternatively, cases could be referred to the court by the Supreme Court when the high court was reluctant to hear an appeal itself.[103] However, the proposal to create a separate court hearing *all* appeals on patents or other specialized subject matter was soundly rejected (a point largely neglected in subsequent Congressional reports and debate). The Commission warned that:

> [T]he quality of decision-making would suffer as the specialized judges become subject to "tunnel vision," seeing the cases in a narrow perspective without the insights stemming from broad exposure to legal problems in a variety of fields . . . . Judges of a specialized court, given their continued exposure to and greater expertise in a single field of law, might impose their own views of policy even where the scope of review under the applicable law is supposed to be more

100. *Compare* H.R. Rep. No. 97-312, at 147 (1981) (dissenting view of F. James Sensenbrenner Jr.) (one percent figure) *with* Commission on Revision of the Federal Court Appellate System Structure and Internal Procedures: Recommendations for Change (Hruska Report), 67 F.R.D. 195, 236 (1975) (demonstrating a slightly larger percentage).

101. Its report is reproduced as COMMISSION ON REVISION OF THE FEDERAL COURT APPELLATE SYSTEM STRUCTURE AND INTERNAL PROCEDURES: RECOMMENDATIONS FOR CHANGE, 67 F.R.D. 195 (1975) [hereinafter RECOMMENDATIONS FOR CHANGE].

102. *Id.* at 199.

103. *Id.*

> limited . . . . [I]ndeed the court as a whole may be "captured" by special interest groups.[104]

A consultant to the Commission found that among 90 identified conflicts on legal doctrines at the U.S. appellate court level, only three were in the patent field.[105]

The specific impetus for a unified court hearing patent appeals apparently coalesced when Attorney General Griffin Bell created within the Department of Justice an Office for Improvements in the Administration of Justice (OIAJ), headed by an assistant attorney general.[106] A proposal calling for a new centralized appellate court, merging the Court of Patent Appeals and the Court of Claims, was circulated in July 1978 to "every office, agency, organization, and individual likely to have any significant interest in the subject."[107] OIAJ's request for comments yielded 46 favoring the proposal, 29 opposed, and 15 that took no position.[108] Given this impetus, the U.S. Congress began considering bills (H.R. 3806, 2405, and S. 1477, and eventually H.R. 4482 and S. 1700) that would create a unified new Court of Appeals for the Federal Circuit with jurisdiction over all patent appeals as well as federal contract dispute claims, customs matters, and an array of other subject matter that was pruned back in Congressional committees.[109] To advance their proposal, OIAJ staff made a concerted effort to co-opt Senator Edward Kennedy, chairman of the Senate Judiciary Committee, who was expected to challenge President Jimmy Carter in the 1980 election and might oppose a Carter-backed bill, but who introduced the OIAJ bill along with his own, adding amendments, in 1979.[110] Bills were passed in both houses of Congress[111] but became bogged down through unrelated procedural complexities in late 1980. The proposal was called up again in the 97th Congressional session beginning in January 1981—a Congress in which Republicans had gained a Senate majority while Democrats retained control of the House. New hearings were held. Two witnesses at the principal House Judiciary Committee hearing were judges from existing courts who would be automatically promoted to the

---

104. *Id.* at 234-35; *see also* MARVER BERNSTEIN, REGULATING BUSINESS BY INDEPENDENT COMMISSION 116-17 (1955).

105. RECOMMENDATIONS FOR CHANGE, *supra* note 101, at 236.

106. Daniel J. Meador, *Origin of the Federal Circuit: A Personal Account*, 41 AM. UNIV. L. REV. 581 (1992). Meador headed OIAJ.

107. *Id.* at 591.

108. *Id.* at 593.

109. *See* S. REP. NO. 97-275 (1981) .

110. S. 677, 96th Cong. (1979); S. 678, 96th Cong. (1979).

111. *Court of Appeals for the Federal Circuit: Hearings on H.R. 2405 Before the Subcomm. on Courts, Civil Liberties, and the Administration of Justice of the H. Comm. on the Judiciary*, 97th Cong. (Apr. 1981) [hereinafter *Hearings on H.R. 2405*].

new court, and another was a company patent attorney who would later be appointed to the new court.[112] In addition to a former Commissioner of Patents, other witnesses represented the American Patent Law Association, the American Bar Association, the Industrial Research Institute (presumably reflecting the views of R&D-oriented corporations), and an independent committee opposing the new law, one member of which had testified in an earlier hearing on behalf of the American Bar Association.[113]

The Bar Association was split. Some of its patent law members, and especially those who practiced in Washington, D.C., favored the bill. Others were against it. The ABA had created committees to consider the proposal for a centralized patent appeals court. At its plenary meeting in February 1980, a majority of the members present voted against it.[114] The ABA representative at hearings in April 1981 reported "very, very substantial division in views among patent lawyers;" said that the forum shopping claim was overblown; and testified that:

> Uniformity, without more . . . is quite plainly not a desirable objective. . .[T]he legal system as a whole reaps the reward that various ideas are able, in the words of Mr. Justice Holmes, to "compete for acceptance in the marketplace" such that the law is refined and grows in a rational and just manner.[115]

A House committee report following the hearings recommended creation of the new court by merging the existing federal Court of Customs and Patent Appeals with the Court of Claims, with jurisdiction mainly for the subject matter of those lower courts but handling patent appeals from all federal circuits. It observed that the responsible Subcommittee had inquired "deeply into technological innovation as an element of productivity in the American marketplace"[116] and cited witness testimony arguing that the new court would be "one of the most far-reaching reforms that could be made to strengthen the United States patent system in such a way as to foster technological growth and industrial innovation."[117] There was in fact no focused testimony on the

112. *Id.*

113. *Id.*

114. *Id.* at 422 (statement of Benjamin L. Zelenko at the June 1980 hearings); *see also* Paul M. Janicke, *To Be or Not To Be: The Long Gestation of the U.S. Court of Appeals for the Federal Circuit (1887-1982),* 69 ANTITRUST L. JOUR. 645, 658 (2001).

115. *Hearings on H.R.* 2405, *supra* note 111, at 85 (statement of James W. Geriak).

116. STAFF OF H.R. COMM. ON THE JUDICIARY, REPORT ON COURT OF APPEALS FOR THE FEDERAL CIRCUIT ACT OF 1981 (accompanying H.R. 4482) 27 (Nov. 4, 1981).

117. *Id.* at 20. In *Origins of the Federal Circuit: The Role of Industry*, 11 FED. Cir. B.J. 541 (2001), one of the first appointees to the new appellate court, Judge Pauline Newman, recalls that judicial reform was recommended by a subcommittee to a Domestic Policy Review

causes of the productivity slump or on how changes in patent policy might be expected to remedy it.

During the most extended debate on the issues, Rep. Tom Railback (R-Ill.) submitted for the record a list indentifying selected individuals and organizations that had, usually through letters, supported passage of the bill.[118] Among 85 corporations favoring the bill, including two universities, 76 of the letters were signed by patent attorneys and only five by individuals whose titles suggested broader responsibilities.[119] Among the 20 organizations cited for their support (none with responsible individuals identified), six were patent law groups, two federal bar associations, six business interest groups, and two were American Indian tribes.[120] Since the call for comments in 1978 drew sharply divided opinions, mostly positive from corporate patent counsel and mostly negative from trial attorneys, one might ask why the letters listed in the 1981 debate were so overwhelmingly favorable. Selection bias could be one explanation, but another, according to OIAJ's head, is that "OIAJ staff had organized the corporate patent counsel into an effective support group for the Federal Circuit."[121]

One amendment made to the bill during its journey through Congress was a statement of the sense of Congress that the quality of the Federal judiciary is determined by the competence of its judges, and that the President should nominate as judges for the new court "from a broad range of qualified individuals"—a counterfoil to the charge that the court's judges would be narrow specialists.[122]

In the definitive House of Representatives roll call vote on the bill November 18, 1981, 321 voted in favor and 76 against.[123] Among Democratic congressmen, the vote in favor was 9.5 to 1; among

---

convened by President Jimmy Carter in 1978. In fact, dozens of such reviews tend to be ongoing at any given time. The broader Review group presumably included members of President Carter's Council of Economic Advisers, which at the time included William D. Nordhaus, who previously published seminal work on the theory of the patent system. However, although the Council's annual reports (published as part of the Economic Report of the President) dealt at length with the productivity slump and stagflation of the 1970s, there was no mention of the patent system as a significant cause.

118. 127 CONG. REC. H. 27,793, 27,793-94 (daily ed. Nov. 17, 1981).

119. *Id.*

120. *Id.*

121. Meador, *supra* note 106, at 610. Professor Meador asserts that "had it not been for OIAJ there would today be no Federal Circuit," because other organized sources of potential support failed to exercise leadership. *Id.* at 619.

122. Federal Courts Improvement Act of 1982, 28 U.S.C. § 45 (2006), *amended by* Pub. L. No. 97-164, 96 Stat. 51 §168(2) (1981)

123. The House vote count for H.R. 4482, the Federal Courts Improvement Act of 1982, Pub. L. No. 97-164, 96 Stat. 51 (1982), can be found at the Library of Congress, Thomas, http://thomas.loc.gov/cgi-bin/bdquery/z?d097:HR04482:@@@L.

Republicans (in the minority), 2.2 to 1.[124] My regression analysis of the vote division introduced three explanatory variables:

Table 3: Explanatory Variables

DEM	Dummy variable; 1 if Democrat, 0 if Republican.
RAND	Industrial research and development expenditures in 1981 (millions of dollars per million population), in a representative's home state.[125]
PROPAT	The percent of cases in which patents were found to be both valid and infringed on appeal in the representatives' home appellate circuits between 1953 and 1977.[126]

The resulting regression equation in ordinary least squares[127] was as follows, with VOTE scaled as 1 for a "yes" vote and 0 for a "nay" vote, and with t-ratios in subscripted brackets:

Equation 1: Regression

$$VOTE = \underset{[10.75]}{0.706} + \underset{[5.83]}{0.222}\,DEM + \underset{[2.31]}{0.00033}\,RAND - \underset{[2.04]}{0.0035}\,PROPAT;$$

$$R^2 = 0.112;\ N = 394.$$

The preponderance of Democratic support is verified, holding constant other variables. Representatives from states with relatively intensive R&D activity were more likely to support the bill, all else equal. Surprisingly, representatives from circuits with a high prior incidence of decisions in favor of patent holders were more likely to vote *against* the court's creation, all else equal.

The vote in the Republican-controlled Senate on December 8, 1981, was more one-sided, with 83 votes in favor and only six nays, three from each party.[128] And so the new Court of Appeals for the Federal Circuit (CAFC) was created, commencing its work on October 1, 1982.

---

124. *Id.*

125. National Science Foundation, Research & Development in Industry: 1987 (NSF 89-323), at 55-56, *data available at* http://www.nsf.gov/statistics/iris/tables.cfm?pub_year=nsf_89-323.

126. ADAM JAFFE & JOSH LERNER, INNOVATION AND ITS DISCONTENTS 100 (2005).

127. Logit regressions were quite similar; the coefficients in OLS regressions are more easily interpreted as the amount by which the vote fraction shifts with a unit change in an explanatory variable.

128. The Senate vote count for S. 1700, the Federal Courts Improvement Act of 1982, Pub. L. No. 97-164, 96 Stat. 51 (1982), can be found at The Library of Congress, Thomas, http://thomas.loc.gov/cgi-bin/bdquery/z?d097:SN01700:@@@L.

Its initial complement of judges was inherited from the prior Court of Customs and Patent Appeals and Court of Claims. As of early 1983, four of the eleven sitting judges had backgrounds in patent law; seven others were from alternative backgrounds.[129] The enabling statute urged the President to make new nominations "from a broad range of qualified individuals."[130] A committee appointed by President Reagan to explore the sources of declining productivity growth and identify improvements recommended to the contrary that the President appoint "experienced patent lawyers to vacancies that occur in the new Court of Appeals . . . ."[131] The recommendation does not appear to have had much impact. In May 2006, the court, whose membership had turned over completely, had five active judges with patent practice backgrounds and six without.[132] However, the court heard a spectrum of cases broader than merely patent matters. Although assignment to panels was in principle random, the choice of the judge who would report the panel's decision, and hence with the opportunity to set at least a precedential tone, was far from random. A study by John Allison and Mark Lemley revealed that in 143 patent validity decisions rendered by the Court between 1989 and 1996, 63 percent of the decisions were written by judges with prior patent practice experience, even though the judges with a patent background comprised only 38 percent of the total number of judges participating in panels hearing validity arguments.[133] Similarly, in a panel discussion among CAFC judges televised by C-SPAN3 on May 19, 2006, chief judge Paul Michel observed that the court did not want judges without patent law experience hearing patent cases and noted the importance of "cohesion" among the CAFC members.[134]

Senator Robert Dole was quoted in the floor debate as saying in Judiciary Committee deliberations preceding the passage of S-1700 that "the bill [will] not substantively affect current law."[135] However, affect it did. The changes were immediate and dramatic, but also subtle. Most significantly, the new CAFC proved to be much more generous than the decentralized appellate courts in ruling that patents whose validity was

129. *See* Federal Judicial Center, Courts of the Federal Judiciary, U.S. Court of Appeals for the Federal Circuit, http://www.fjc.gov/history/home.nsf/usca_fc_frm.

130. 28 U.S.C. § 45 (2006), *amended by* Pub. L. No. 97-164, 96 Stat. 51 § 168(2) (1981); *see supra* note 122.

131. WHITE HOUSE CONFERENCE ON PRODUCTIVITY, PRODUCTIVITY GROWTH: A BETTER LIFE FOR AMERICA 80 (1984).

132. Federal Judicial Center, *supra* note 129.

133. John R. Allison & Mark A. Lemley, *How Federal Circuit Judges Vote in Patent Validity Cases*, 27 FLA. ST. U. L. REV. 752 (2000).

134. *Panel Discussion among CAFC Judges* (CSPAN3 television broadcast, May 19, 2006).

135. 126 CONG. REC. S. 29,887 (daily ed. Dec. 8, 1981) (statement of Sen. Charles Grassley). I was told the same thing about the bill's intent by a member of the Senate Judiciary Committee staff at the time.

challenged on the basis of insufficient novelty or utility were in fact valid. The old courts rejected roughly two thirds of the patents on validity grounds; the new court upheld roughly two thirds.[136] This fed back to induce a higher acceptance rate at the district courts. With a validity ruling more likely, there were more attempts by patent holders to enforce patents, whose ultimate success depended then upon whether the courts ruled the relevant patents to have been infringed. The new appellate court's statistical record in infringement questions, on the other hand, was tougher on patent-holding claimants than in the previous decentralized courts.[137] In interpreting the so-called doctrine of equivalents, the CAFC tended to view the scope of litigated patents more narrowly than its predecessors.[138] But with a higher fraction of patents found to be valid, the percentage of tested patents found to be both valid and infringed rose during the first decade of the court's existence before declining, and the absolute number of patents found to be both valid and infringed per year more than doubled, with a generally rising trend.[139]

The new court also blazed a trail toward accepting new kinds of patents, e.g., on business methods[140] and computer software, on which the difficulties of showing that prior art would preclude patenting were particularly great, and (with Supreme Court encouragement)[141] an expanded array of life form inventions—much wider than the European Community chose to protect.[142] It proved more amenable to accepting jury findings, despite evidence that juries were more likely to be awed by claims of technical novelty than judges. The new court was more willing than the decentralized courts to grant preliminary and final injunctions

---

136. JAFFE & LERNER, *supra* note 126, at 100-06; John R. Allison & Mark A. Lemley, *Empirical Evidence on the Validity of Litigated Patents*, 26 AIPLA Q. JOUR. 185 (1998); Matthew D. Henry & John L. Turner, *The Court of Appeals for the Federal Circuit's Impact on Patent Litigation*, 35 J. LEGAL STUD. 85 (2006).

137. Glynn S. Lunney Jr., *Patent Law, the Federal Circuit, and the Supreme Court: A QuietRevolution*, 11 SUP. CT. ECON. REV. 1, fig.3 (2004).

138. *See* Henry & Turner, *supra* note 136; Lunney, *supra* note 137. A key case was *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 234 F. 3d 558 (Fed. Cir. 2000), *rev'd* 533 U.S. 915 (2001).

139. Lunney, *supra* note 137, at 80 (App. I).

140. However, in 2008, in the *In re Bilski* case, the appellate court invited outside comments on whether its earlier *State Street Bank* precedent (1998) allowing business methods patents should be overturned. *See America's Patent System: Methods and Madness*, ECONOMIST, May 10, 2008, at 75 (discussing *In re Bilski*, 545 F.3d 943 (Fed. Cir. 2008) and *State Street Bank & Trust v. Signature Financial*, 149 F.3d 1368 (Fed. Cir. 1998)).

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

142. For a survey of 1,770 DNA sequence patents issued between September 1998 and June 2000, see F. M. Scherer, *The Economics of Human Gene Patents*, 77 ACAD. MED. 1356, 1356-59 (2002). *See also* Kyle Johnson & Fiona Murray, *Intellectual Property Landscape of the Human Genome*, 310 SCI. 239 (2005).

eliminating infringers from a field—although on this, its exertions may be restrained by an important Supreme Court pronouncement in 2006 denying that there is a "general rule" supporting injunctions in patent infringement. Instead, the traditional four-factor test (including considerations of equity) should be applied.[143] And very significantly, the CAFC revised the principles for assessing damages in cases of proven infringement, making it more likely that estimates of profits lost by the patent holder would err on the generous side, favoring the "profits lost" standard over the milder "reasonable royalty" standard, and awarding damages under both standards even though the latter is logically subsumed within the former.[144] Under the new standards, courts imposed several damages awards running into the hundreds of millions of dollars.[145]

The Federal Circuit's new rulings on balance strengthened patent protection, made it likely that companies found to be infringing valid patents would pay substantial damages, and hence raised the perceived benefits to companies (and universities) from building strong patent portfolios. Patent applications and patent issues soared in the years following the creation of the CAFC (marked by a dotted vertical line), as shown in Figure 1. A regression analysis shows a distinct and statistically significant break in the series at the year 1983,[146] with the growth rate of applications (less subject than patent issues to Patent Office backlog fluctuations) averaging 1.4 percent per year between 1955 (after postwar adjustments were accomplished) and 1982, and 5.97 percent per year between 1983 and 2004. With many more patents being sought, more patent attorneys had to be hired. The number of patent attorneys per billion dollars of price level-adjusted industrial R&D expenditures rose from approximately 50 in the 1970s to 75 in the mid-1990s.[147] With many more patents being issued, specific areas of technology became more congested, leading to a higher likelihood that one firm's proprietary technology would conflict with another firm's.[148] In an analogue of an

---

143. eBay Inc. v. MercExchange, 547 U.S. 388 (2006).

144. *See* Cecil D. Quillen, Jr., *Innovation and the U.S. Patent System*, 1 VA. L. & BUS. REV. 207 (2006).

145. *See, e.g.*, Polaroid Corp. v. Eastman Kodak Co., 833 F.2d 930 (Fed. Cir. 1986); Gyromat Corp. v. Champion Spark Plug Co., 735 F.2d 549 (Fed. Cir. 1984); Bio-rad Lab., Inc. v. Nicolet Instrument Corp., 739 F.2d 604 (Fed. Cir. 1984).

146. The F-ratio in a test of differences is 8.54, which is highly significant statistically, with N = 20 and 81. The data, including only "utility" patents and not design or plant patents, were obtained from the Patent and Trademark Office web site. For a more detailed analysis, see Bronwyn Hall, *Exploring the Patent Explosion*, 30 J. TECH. TRANSFER 35 (2005).

147. John Barton, *Reforming the Patent System*, 287 SCI. 1933 (2000).

148. See, for example, JAMES BESSEN & MICHAEL J. MEURER, PATENT FAILURE: HOW JUDGES, BUREAUCRATS, & LAWYERS PUT INNOVATORS AT RISK (2008), which estimates that the combined litigation costs for plaintiffs and defendants exceed the estimated

arms race, companies strove all the more vigorously to expand their patent portfolios so they could use their patents in defensive counter-claims when accused of infringement. With many more patents and higher damages if one's technology were found to infringe another firm's patents, developing new products became like walking through a mine field, with dire consequences from a misstep.

While stronger patent protection per se should have increased the profitability of innovation and hence stimulated R&D expenditures, all else equal, the increased danger from infringing another firm's patents exerted an opposite negative influence. Figure 2 shows the long-run trend of U.S. industrial expenditures on research and development from 1953, the first year covered by consistent surveys, through 2000.[149] Outlays are measured in constant 1996 dollars. As in Figure 1, the plot is logarithmic, so that a straight line indicates a constant rate of growth. Factors other than the legal regime in which patents were administered—notably, macroeconomic shocks, the energy shocks of 1973-74, and the advent of wholly new technologies such as the World Wide Web—had an obvious impact. The most that can be said is that there is no noticeable acceleration of the growth rate in R&D following the creation of CAFC. In a statistical test comparing the periods 1956-82 and 1983-2000, the rates of growth are insignificantly different.[150]

I conclude that the CAFC did change patent policy when the legislators who supported it said it would not, that the record of debates on the enabling bill contains no solid evidence that the change would in fact stimulate R&D, and that there is no evidence of an acceleration in company-financed R&D between the 27 years before the bill was enacted and the 18 years thereafter.

D. *Pharmaceutical Patent Reforms*

As the 1980s dawned, pharmaceutical manufacturers had two major

---

benefits from patent protection. *See also* JAFFE & LERNER, *supra* note 126 (especially Chapter 2); Rochelle Dreyfuss, *Pathological Patenting: The PTO As Cause or Cure*, 104 MICH. L. REV. 1559 (2006) (a review of Jaffe and Lerner); Bronwyn Hall & Rosemarie Ziedonis, *The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry*, 32 RAND J. ECON. 101 (2001); Iain Cockburn & Megan MacGarvie, *Patents, Thickets, and the Financing of Early-Stage Firms: Evidence from the Software Industry* (Nat'l Bureau of Econ. Research, Working Paper No. 13644, 2007); Rosemarie Ziedonis, When the Giants' Shoulders Are Crowded: Fragmented Rights and Incentives To Patent (2002) (unpublished manuscript).

149. Figure 1 and Figure 2 are available in the Appendix of this article. NAT'L SCI. FOUND., SCIENCE & ENGINEERING INDICATORS: 2004, vol. 2, at A4-5, tbl.4-4, *available at* http://www.nsf.gov/statistics/seind04/pdf/volume2.pdf.

150. NAT'L SCI. FOUND., SCIENCE & ENGINEERING INDICATORS: 2004, vol. 2, at A4-6, *available at* http://www.nsf.gov/statistics/seind04/append/c4/at04-06.pdf. The F-ratio is only 1.33. Observations before 1956 are excluded because the National Science Foundation had not yet perfected its survey techniques.

complaints, leading eventually to the Hatch-Waxman Act of 1984.[151]

For the makers of relatively new, typically patented, drugs, the key problem was declining effective patent life. Responding to the record of adverse side effects found with the tranquilizer Thalidomide, the Kefauver-Harris Act of 1962[152] increased the Food and Drug Administration's power to ensure that new drugs were safe. It also required proof from well-controlled clinical trials of a new drug's efficacy as well as its safety. Clinical trial periods and FDA decision-making lengthened appreciably as a result—to an average of 7.5 years, with considerable variation—between the time when the FDA authorized testing in human beings to the date at which approval for marketing a new drug (a so-called NDA) was granted.[153] Typically, drug companies filed for patent protection when animal tests demonstrated possible therapeutic effects, about a year before human tests began. With an average lag between patent application and patent issuance of from two to four years and a patent life (since changed) of 17 years from issue to expiration, new drug marketers enjoyed on average only 10 to 13 years from the initiation of marketing to patent expiration, at which point, in principle, generic competition could begin. Both directly and through their trade association, the Pharmaceutical Manufacturers' Association (PMA), the research-oriented drug companies sought relief from Congress in the form of patent life extension.

The generic drug manufacturers also had a problem. Because of restrictive FDA rules approved by the Supreme Court,[154] the obstacles to generic competition were substantial even after relevant patents expired. Generic producers were not able simply to "free ride" on the test results of the original drug producers, which, the pioneers claimed, generated data that were their exclusive property. Would-be generic producers were required to conduct their own clinical trials nearly as extensive as those of the pioneers. This barrier to imitation significantly discouraged generic entry.[155] Generic drug companies sought from Congress eased testing requirements taking advantage of an original drug's evident safety and efficacy, proved in both FDA-required tests and the marketplace.

Extensive hearings were conducted by several Congressional committees.[156] The witnesses included not only top officials of the

---

151. Federal Food, Drug and Cosmetic Act, 21 U.S.C. §351 (2006), *amended by* Hatch-Waxman Act, PUB. L. NO. 98-417, 98 Stat. 1585 (1984) [*hereinafter* Hatch-Waxman Act.]

152. Federal Food, Drug and Cosmetic Act, 21 U.S.C. §351 (2006), *amended by* Kefauver-Harris Act, Pub. L. 87-781, 76 Stat. 780 (1962).

153. *Id.*

154. United States v. Generix Drug Corp., 460 U.S. 453 (1983).

155. *See* E. W. Kitch, *The Patent System and the New Drug Application*, *in* REGULATING NEW DRUGS 81-108 (R. L. Landau ed., 1973).

156. *E.g.*, *Patent Term Restoration Act of 1981: Hearings on H.R. 1937, H.R. 6444, and S.*

principal interested parties—the PMA, the Generic Pharmaceutical Industry Association, the Food and Drug Administration, and various drug companies—but also the government's Office of Technology Assessment, which had made a study of the various proposals; a leading economic researcher on the economics of pharmaceutical innovation; a university-based physician who had done important research on drug testing; consumer advocate Ralph Nader; and a representative of the AARP, among others. The relevant issues were thoroughly aired.

In the end, compromise language was negotiated by the two principal outside parties—the PMA and the Generic Industry Association. It had two main parts. First, an extension on the life of one patent, chosen by the drug firm, would be allowed to compensate for regulation-mandated test and decision delays. The maximum extension, however, could not be more than five years or enough only to allow an effective patent life of 14 years from the time of FDA approval.[157] Second, once patents expired, generic producers would be allowed to enter the market immediately on the basis of chemical analysis and abbreviated clinical tests—typically involving 24 subjects—showing that the generic version was chemically identical (i.e., bioequivalent) to, and was absorbed into a patient's bloodstream at approximately the same rate as the originally patented and FDA-approved drug.[158] The most controversial part of the compromise, Section 202, the so-called Bolar amendment,[159] allowed generic drug makers to produce experimental quantities of a still-patented product "solely for uses reasonably related to the . . . submission of information under a Federal law which regulates . . . drugs"—i.e., to conduct the trials demonstrating bioequivalence. In this way, the generic drug maker could submit its application to the FDA and, with luck, hit the ground running with its marketable product the day the original drug's blocking patent expired. The Bolar amendment established a new principle—that experimental

255 Before the Subcomm. on Courts, Civil Liberties, and the Administration of Justice of the H. Comm. on the Judiciary, 97th Cong. (July-Nov. 1981); *Patent Term Extension & Pharmaceutical Innovation Before the Subcomm. on Investigations and Oversight of the H. Comm. on Science and Technology*, 97th Cong. (Feb. 1982); *Patent Term Restoration Act of 1983: Hearings on S. 1306 Before the Subcomm. on of Patents, Copyright and Trademarks of the S. Comm. on the Judiciary*, 98th Cong. (June-Aug. 1983); *Drug Legislation: Hearings on H.R. 1554 and H.R. 3605 Before the Subcomm. on Health & Env. of the H. Comm. On Energy and Commerce*, 98th Cong. (July-Oct. 1983).

157. Hatch-Waxman Act, *supra* note 151.

158. *Id.*

159. The name comes from a decision by the new Court of Appeals for the Federal Circuit in *Roche Prod. Inc. v. Bolar Pharm. Co.*, 733 F.2d 858 (Fed. Cir. 1984), *superseded by* 35 U.S.C. § 271(e) (2006), preventing generic manufacturers from producing test quantities of a drug while the drug was still under patent.

uses of a product might not be blocked by patent protection.[160]

After the more controversial provisions were accepted, the compromise law was passed unanimously in the House of Representatives and by voice vote in the Senate.[161] Within the pharmaceutical industry, however, controversy persisted. A cabal led by the Swiss-based company Hoffmann-LaRoche was displeased and saw to it that the president of the Pharmaceutical Manufacturers Association, Lewis Engman, who had played a key role in brokering the compromise that eventually reached Congress,[162] was fired from his position.

The Hatch-Waxman Act had important effects. The share of all drug prescriptions dispensed in the United States and filled generically rose steadily from 19 percent in 1984, when the new law was passed, to 47 percent in 2000, with further increases expected.[163] Generic competition clearly became more vigorous.[164] Significant patent life extensions were also achieved, partly under the main terms of the Act and partly through strategic manipulation of provisions defining the various parties' rights in patent disputes.[165] The extension in patent lives should have increased industry profits, but more rapid and extensive generic competition worked in the opposite direction. Industry profitably did increase markedly after passage of the Act,[166] but the rising trend began three years earlier and had two other plausible causes—the advent of so-called "rational drug design" in which scientific knowledge played a larger role, and the rapid spread of health insurance plans with drug expenditure reimbursement, which reduced the elasticity of demand and hence supported increased prices for patented drugs sold under monopolistic conditions.

A plausible argument can be advanced that the Act shaped an ideal compromise in terms of stimulating pharmaceutical innovation. Longer

---

160. For an extension reversing the CAFC's narrow reading of the Bolar amendment and allowing use in investigating novel drugs at preclinical stages as well as for generics, see *Merck KGAA v. Integra LifeSciences*, 545 U.S. 193 (2005).

161. *See* Mary K. Olson, *Political Influence and Regulatory Policy: The 1984 Drug Legislation*, 32 ECON. INQUIRY 363, 376-80 (1994).

162. *See* Milt Freudenheim, *Lewis Engman, 59, U.S. Official And Drug Industry Spokesman*, N.Y. TIMES, July 13, 1995, at B12.

163. PHARMACEUTICAL RESEARCH & MANUFACTURERS ASSN. OF AMERICA, PHARMACEUTICAL INDUSTRY PROFILE 62 (2003).

164. One consequence is little recognized. By reducing the front-end testing costs incurred for generic entry, the Act's provisions not only encourage early generic competition, but make it possible for *more* generic firms to squeeze into a given market, intensifying price competition. The existence of Hatch-Waxman plus the large size of the U.S. market explains why U.S. generic drug prices tend to be the lowest in the industrialized world.

165. Many of the manipulations were found to be illegal. *See* FEDERAL TRADE COMM., GENERIC DRUG ENTRY PRIOR TO PATENT EXPIRATION: AN FTC STUDY (2002); Joe Nocera, *Generic Drugs: The Window Has Loopholes*, N.Y. TIMES, July 1, 2006, at C1.

166. *See* Scherer, *The Link*, *supra* note 50.

patent protection had at the margin its desired effect in increasing the profitability of a given efficacious new drug. Less widely recognized, but equally true, the acceleration of generic competition forced pharmaceutical makers to intensify their efforts to discover and test improved replacement products, for without them, the sales and profits from a patented drug can be expected to plummet shortly after patent expiration.[167] Thus, the Act provided both a carrot and a stick to encourage innovation.

E. Changes in Administration of the Patent-Antitrust Interface

There were other Congressional and judicial decisions altering patent policy in the 1980s and 1990s. Here we note briefly one other line of development—the presumptions applied by the U.S. antitrust agencies when the exploitation of patent positions was alleged to conflict with antitrust prohibitions.

During the 1970s the Antitrust Division of the Department of Justice articulated a list of nine so-called "no-no's," most of which delineated what a patent holder could do in licensing to other firms before running afoul of the antitrust laws.[168] The approach in effect asked whether restrictions written into patent licenses were necessary and whether less restrictive measures could have achieved the same objectives. Agreements to set minimum prices at which licensees could sell licensed products and to restrict licensing of third parties, mandatory package licensing, and requirements that the licensee buy unpatented products from the licensor (i.e., ties) were viewed with special skepticism.[169]

Partly because of Supreme Court decisions taking a more benign view of certain vertical restraints (such as exclusive franchising) and the installation of relatively pro-business Reagan appointees, a more tolerant view emerged on how patents and antitrust interacted. An early statement by an Antitrust Division official said that the nine no-no's "contain more error than accuracy" as statements of rational economic policy.[170] Five years later a deputy assistant attorney general criticized the "history of antagonism toward patent licensing" and urged that patent

---

167. *See Interview with Sidney Taurel* (CEO of Eli Lilly) (CSPAN3 television broadcast May 8, 2006).

168. *See* AMERICAN BAR ASS'N, SECTION ON ANTITRUST LAW, THE FEDERAL ANTITRUST GUIDELINES FOR THE LICENSING OF INTELLECTUAL PROPERTY: ORIGINS & APPLICATIONS 8-10 (2d. ed. 2002) [hereinafter ABA ANTITRUST GUIDELINES]. The document provides a comprehensive overview of the issues and reproduces Guidelines published by the antitrust agencies.

169. *Id.*

170. Statement of Abbott B. Lipsky Jr. before the American Bar Association Nov. 5, 1981, *reprinted in* 4 Trade Reg. Rep. (CCH) para. 13,129.

licensing could have numerous pro-competitive benefits.[171] On this he was clearly correct. Some deeper premises, however, were debatable. Ignoring the emerging literature on alternative first-mover advantages, he singled out patents as instruments for preventing free-riding on investments in technology, arguing that "patents create property rights without which technology would not exist—or certainly not in its current abundance."[172] As the work of Taylor and Silberston and of Mansfield, already available at the time, made clear, this could be true for *some* new technologies, but by no means for all.[173] The DoJ spokesman's further premise, therefore, is also questionable:

> Efforts to appropriate as much as possible of the surplus—the social value in excess of marginal cost—lying under the demand curve for the patented technology do not harm competition. Indeed, the potential for appropriating those rents is *the* engine that drives the technology market.[174]

In effect, the implication was that almost anything done unilaterally to increase an innovator's profits was beneficial for competition—and given the way antitrust had come to be interpreted, beneficial for consumers. Such a view goes too far. In 1995, after substantial interaction with the legal and scholarly communities, the Department of Justice and Federal Trade Commission jointly issued new *Guidelines for the Licensing of Intellectual Property* (*Guidelines*).[175] In effect, the *Guidelines* stated that the antitrust agencies would analyze questionable patent/antitrust interactions on a "rule of reason" basis, asking whether a restraint "is reasonably necessary to achieve procompetitive benefits [e.g., superior or more extensive innovation] that outweigh . . . anticompetitive effects."[176] Given the complex repercussions of the practices addressed, a careful "rule of reason" approach seems eminently reasonable. One might hope, however, that antitrust agency staff charged with enforcing the guidelines and the courts interpreting them possess a broad understanding of what economic analysis—on both the theoretical and empirical sides—reveals about the limited and conflicting roles patents play.

Pessimism on this last point is in order, since two more recent authoritative reports on the intersection between antitrust policy and

---

171. Statement of Charles F. Rule before the World Trade Association and the Cincinnati Patent Law Association, Oct. 21, 1986, *reprinted in* Trade Reg. Rep. (CCH) para. 13,131.

172. *Id.*

173. *See* Mansfield, *supra* note 26; TAYLOR & SILBERSTON, *supra* note 24.

174. *Reprinted in* Trade Reg. Rep. (CCH) ¶ 13,131 (emphasis added).

175. ABA ANTITRUST GUIDELINES, *supra* note 168, at 116.

176. *Id.*

patent policy essentially mimic the assumptions accepted by Department of Justice staff in the 1980s. Ignoring decades of empirical evidence accumulated by economists on the role of patents, both imply that the expectation of patent protection is a principal basis for investment in new technology. The Department of Justice/Federal Trade Commission report, for example, opens by asserting that "Intellectual property laws create exclusive rights that provide incentives for innovation . . . [and] prevent others from appropriating much [sic] of the value derived from . . . inventions or original expressions."[177] It goes on to assert that "intellectual property laws protect the ability to earn a return on the investments necessary to innovate."[178] Initiating its analysis with an approving citation to the 1981 Department of Justice statement quoted above, the Antitrust Modernization Commission states that "the courts and the antitrust agencies in recent decades have evidenced a greater appreciation of the importance of intellectual property" and suggests that "[i]ntellectual property may be critical to future innovation in an industry."[179] If the enforcement agencies and courts are led to believe that the expectation of patent rights is the principal inducement to innovation and ignore the important role of other first-mover advantages, they will be wrong more often than right in balancing antitrust objectives against intellectual property considerations in rule of reason cases.[180]

F. Extension of U.S. Patent Standards to Other Nations

Undoubtedly more important than reforms in domestic patent law were U.S. efforts to influence the patent laws of other nations, especially less-developed nations. Piracy of copyrighted music, motion pictures, and computer programs—matters not addressed in this paper—was one provocation.[181] On patents, a key problem was the fact that the Paris

---

177. U.S. DEP'T OF JUSTICE & FED. TRADE COMM'N, ANTITRUST ENFORCEMENT AND INTELLECTUAL PROPERTY RIGHTS: PROMOTING INNOVATION AND COMPETITION 1 (2007).

178. *Id.* at 2.

179. ANTITRUST MODERNIZATION COMM'N, REPORT AND RECOMMENDATIONS at 39-40 (2007). A more nuanced view—"the premise that greater protection of intellectual property necessarily fosters more innovation turns out to be false"—is advanced by the Commission's only economist member. *See* Dennis W. Carlton, *Does Antitrust Need To Be Modernized?*, 21 J. ECON. PERSP. 155, 165 (2007).

180. This bias is not evident in the National Academy of Engineering report on the patent system. In Chapter II of the report, the equivocal role of patents as incentives for innovation is clearly acknowledged. *See* NAT'L RES. COUNCIL, PATENT SYSTEM FOR THE 21ST CENTURY 32-33 (2004).

181. The term "piracy" was already used to denote cribbing of musical compositions in the 18th Century. *See* F. M. SCHERER, QUARTER NOTES AND BANK NOTES: THE ECONOMICS OF MUSIC COMPOSITION IN THE EIGHTEENTH AND NINETEENTH CENTURIES 167, 176 (2004).

Convention governing international patent relations, inaugurated in 1883, allowed member nations to determine the coverage of their patent laws, requiring mainly that they not discriminate between domestic and foreign patent applicants.[182] Many nations had patent systems providing much less protection for inventions than the United States did. Among 33 sizeable developing and high-income nations in 1990, for example, 14 offered no patent protection for pharmaceutical products, 15 none for food products, and 11 none for chemical products.[183] Eight of the 33, including Switzerland, home to three of the world's leading pharmaceutical companies, had joined the list of nations allowing patents for pharmaceutical products only between the years 1975 and 1989.[184]

For pharmaceuticals, in which patents are accorded such importance, Italy was an early bete noire and focus of action. A patent law passed in 1939 and still applicable in the 1970s excluded pharmaceutical products from patentability.[185] As a consequence, Italy became a world leader in producing and exporting generic pharmaceuticals to other nations—immediately for importing nations without product patent protection, otherwise as soon as national patent laws allowed. Among other things, during the late 1960s Italy was a major supplier of early "wonder drugs" (broad-spectrum antibiotics) to the U.S. military purchasing authorities. This was stopped through an amendment to a foreign assistance bill, offered by a Congressman from Indianapolis on the floor of the House of Representatives in 1961 and passed by a vote of 87 to 65 (less than a quorum) after cursory debate.[186] A 1963 attempt to change the Italian law, led by large Italian pharmaceutical companies, was blocked in the Italian Parliament owing to small-firm opposition.[187] During the 1970s, a group of multinational pharmaceutical companies from the U.S.A., Germany, Japan, and Switzerland, joined by some larger Italian firms, challenged the constitutionality of Italy's law. In March 1978, Italy's Corte Constitutionale found the exclusion of pharmaceutical products to be unconstitutional and ordered the prompt acceptance of drug patent applications.[188] In the decade that followed, pharmaceutical R&D and

182. Paris Convention for the Protection of Industrial Property, art. 2, Mar. 20, 1883, 21 U.S.T. 1583, TIAS No. 6932.

183. Edson K. Kondo, Patent Laws and Foreign Direct Investment: An Empirical Investigation 62 (May 1994) (unpublished Ph.D. dissertation, Harvard University) (on file with Harvard University).

184. *Id.* at 64-65.

185. *See* Weisburst, *infra* note 188.

186. 107 CONG. REC. H 16,283, 16,283-85 (daily ed. Aug. 18, 1961).

187. Herbert Koshetz, *Italian Sees Rise in Drug Research*, N.Y. TIMES, Sept. 26, 1963, (Business and Finance), at 47.

188. Sandy Weisburst, Strengthening Patent Protection in Italy (Mar. 1995) (unpublished senior thesis, Harvard University) (on file with Harvard University). The results are

new product launches did not rise relative to world trends, while Italy's balance of trade in pharmaceuticals dropped from positive to negative.[189] India took Italy's place as the world's leading supplier of generic drugs to nations without product patents and, given its first-mover advantage, as an early generic supplier in the United States.

Beginning in the late 1970s a concerted effort began to bring the full array of laggard nations up to U.S. patent law standards. Among the prime movers were the U.S. pharmaceutical companies. Unlike the other legislative developments covered by this paper, the lobbying efforts that followed are richly documented.[190] Between 1981 and 1987, Edmund Pratt, CEO of Pfizer Inc., was chairman of the U.S. President's Advisory Committee on Trade and Negotiations (ACPTN). Its subcommittee on intellectual property was chaired by IBM CEO John Opel. In their role as advisors to the U.S. Trade Representative (USTR), coordinating international trade matters for the Executive Branch, and also in their communications with Congress, they pushed hard to bring patent and copyright issues to the forefront of U.S. trade dealings with other nations and international agencies. At the time, USTR had, with the exception of one overburdened staff member, virtually no independent economic analysis capability. Pratt and Opel reached out to organize lobbying efforts by other industry groups such as the Pharmaceutical Manufacturers Association, the Business Roundtable, and a panoply of organizations seeking copyright protection.

These lobbying efforts led initially to the passage of two amendments to Section 301 of the Trade Act of 1974,[191] which defines unfair trade practices against which the United States might retaliate. The first, in 1984, authorized the U.S. government to impose unilateral sanctions against nations that failed to provide adequate intellectual property protection.[192] Section 301 was strengthened into what was called "Special 301" in 1988, requiring the USTR to prepare an annual report identifying foreign nations with the most objectionable patent and copyright policies, placing those nations on a priority list, and commencing an investigation to determine whether the subject nations'

---

summarized in F.M. Scherer & Sandy Weisburst, *Economic Effects of Strengthening Pharmaceutical Patent Protection in Italy*, *in* F. M. SCHERER, PATENTS: ECONOMICS, POLICY, AND MEASUREMENT 67 (2005).

189. *Id.* at 67-82.

190. *See* MICHAEL P. RYAN, KNOWLEDGE DIPLOMACY: GLOBAL COMPETITION AND THE POLITICS OF INTELLECTUAL PROPERTY (1998); MICHAEL SANTORO, PFIZER: GLOBAL PROTECTION OF INTELLECTUAL PROPERTY (1992). Much of what follows is drawn from those publications.

191. U.S. Trade Act of 1974, 19 U.S.C. §§ 2101, 2102 (2006).

192. 19 U.S.C. §§ 2101, 2102, *amended by* Pub. L. No. 98-573 (1984).

"IP" policies merited retaliatory measures.[193] The USTR proceeded cautiously, establishing in 1989 only a "priority watch list" that included Brazil, India, Mexico, the Peoples Republic of China, South Korea, Saudi Arabia, and Thailand.[194] In May 1989 the United States imposed 100 percent tariffs on $39 million of imports from Brazil as punishment for its deficient pharmaceutical patent policies.[195] Threats were levied against Mexico, South Korea, China, and Thailand, among others. In 1991 the first actual priority list was issued, naming Thailand, India, and China as prime targets. Thailand's government had been dissolved in a no-confidence vote as a direct consequence of a patent bill introduced into the National Assembly in 1988 in response to early U.S. pressure.

The business advisors to the U.S. government and their industry allies also worked on a broader international front. Both directly and through U.S. representatives, they sought to have the Paris Convention modified to require uniformly high patent law standards for member nations. Attempts to reach this goal through the World Intellectual Property Organisation (WIPO), a branch of the United Nations, and at the Nairobi round of Paris Convention negotiations were a failure. Efforts with WIPO were "a disaster," a Pfizer executive said, because "WIPO works by majority, and simply put, there were more of them than us."[196] Nairobi Round initiatives during the late 1970s failed because United States, European, and Japanese delegates were unable to agree on a united front.[197] Absorbing the lessons from these failures, Pratt and Opel organized a combined lobbying campaign by U.S. patent and copyright-sensitive industries, who in turn recruited their counterparts in Europe, e.g., the Dolder Group of pharmaceutical companies,[198] and the Keidanren in Japan. All put pressure on their governments to make stronger intellectual property rights a priority issue in international trade deliberations.

The opportunity arose with the start of a new round of international trade policy negotiations—the Uruguay Round—in September 1986. The United States' component of the effort was organized through an "Intellectual Property Committee" comprising the chief executives of 13

---

193. 19 U.S.C. §§ 2101, 2102, *amended by* Omnibus Trade and Competitiveness Act of 1988, Pub. L. No. 100-418 (1988).

194. RYAN, *supra* note 190, at 80, 85-86.

195. James Brooke, *Brazil Cites Large Debt In Its Defense on Trade*, N.Y. TIMES, May 27, 1989, at A31.

196. Santoro, *supra* note 190, at 7 (quoting Lou Clemente, Pfizer general counsel and chair of the intellectual property committee of the U.S. Council for International Business).

197. *See* Fenton Hay, *Canada's Role in International Negotiation Concerning the Patent Laws*, *in* 8 RES. IN LAW & ECON 239-63 (John Palmer ed., 1986).

198. So-called because their chief executives met each year at the Dolder Grand Hotel in Zürich.

major companies.[199] Working with their counterparts from Europe and Japan, the IPC members distributed in June 1988 a 100-page "Basic Framework" setting goals for the inclusion of intellectual property issues in whatever treaty resulted from Uruguay Round negotiations.[200] A key to the agreed-upon strategy was "linkage." Most less-developed nations opposed their inclusion, but United States negotiators, supported inter alia by individuals seconded to their team from the Patent and Trademark Office, made it clear that the United States would not ratify any treaty unless it included IP standards, and there would be no cherry-picking—all provisions had to be accepted by a ratifying nation. If less-developed nations were eventually to secure relief from the Multi-Fibre Agreement, which limited the textile exports on which they had comparative advantage, and developed-nation barriers to agricultural product imports, they would have to go along with the intellectual property provisions. And perhaps even more important, having intellectual property questions covered by the ratified Uruguay Round Treaty removed most possibilities that the United States could brandish its Section 301 sword unilaterally. Tough bargaining yielded a compromise draft of what came to be called the "TRIPS" (Trade-Related Aspects of Intellectual Property Rights) agreement, which was included in the final draft treaty compiled by the GAAT Secretary-General and in the ultimate Treaty of Marrakech that replaced GAAT with the World Trade Organization.[201]

U.S. advocates of TRIPS argued among other things that less-developed nations should welcome strengthened patent laws because they would encourage domestic innovation, which among other things flourished in the early history of the United States, and because it would induce more inward technology transfer through foreign direct investment by multinational enterprises. There is an element of paradox in this argument, since most less-developed nations with weak patent policies were opposed to the changes, which suggests that the LDCs did not know what was good for them. The argument also overlooks the fact that during the first 47 years of its existence, the United States provided strong patent protection to domestic residents, but denied patents to foreigners, whereas LDCs were being asked under TRIPS to increase the scope of their patent protection to both domestics and foreigners.

199. Pharmaceutical makers Pfizer, Merck, du Pont, Bristol-Myers, and Johnson & Johnson, plus General Electric, Warner Communications, Hewlett-Packard, FMC Corporation, General Motors, and Rockwell International.

200. INTELLECTUAL PROPERTY COMMITTEE, BASIC FRAMEWORK OF GATT PROVISIONS ON INTELLECTUAL PROPERTY (1988).

201. Agreement on Trade-Related Aspects of Intellectual Property Rights, Apr. 15, 1994, 869 U.N.T.S. 299, 33 I.L.M. 1197 [hereinafter TRIPS].

Economic theory provided at best ambiguous guidance on the alleged benefits to poor nations of strong and open patent systems.[202] Some econometric studies suggested that strong patent systems encouraged inward foreign direct investment, but the most positive early findings were based on subjective measures of patent system strength that could have reflected the evaluators' broader views on the desirability of nations for investing. The only early study using more objective measures reported negative or inconclusive results.[203]

The opposition of LDC negotiators to uniform U.S.-grade patent protection led to compromises in the TRIPS version ultimately accepted. For one, full implementation of TRIPS by nations categorized as least-developed could be delayed until 2005.[204] Provision was made in Article 40 for non-exclusive compulsory licensing of patents in cases of monopolistic abuse and also, in Article 31:

> [Such] use may . . . be permitted if, prior to such use, the proposed user has made efforts to obtain authorization from the rights holder on reasonable commercial terms and conditions and that such efforts have not been successful within a reasonable period of time. This requirement may be waived by a Member in case of a national emergency or other circumstances of extreme urgency or in cases of public noncommercial use.[205]

Curiously, most references to this provision in the U.S. press have stressed the "national emergency" part and ignored the language allowing compulsory licenses when negotiations have failed to converge on "reasonable commercial terms." How that misconception was propagated is unclear.

Article 31, subparagraph (f), also stipulated that compulsory licenses be authorized "predominantly for the supply of the domestic market of the Member authorizing such use."[206] For most of the world's least-developed nations, this provision posed a special difficulty in such areas as pharmaceuticals, since those nations typically had neither the technical

---

202. *See, e.g.*, Alan Deardorff, *Should Patent Protection Be Extended to All Developing Countries,* 13 THE WORLD ECON. 497 (1990); Alan Deardorff, *Welfare Effects of Global Patent Protection*, 59 ECONOMICA 35 (1992); F. M. Scherer, *A Note on Global Welfare in Pharmaceutical Patenting*, 27 THE WORLD ECON. 1127 (2004).

203. *Compare* Richard Rapp & R. Rozek, *Benefits and Costs of Intellectual Property Protection in Developing Countries*, 24 J. WORLD TRADE 75 (1990) *and* Jeon-Yeon Lee & Edwin Mansfield, *Intellectual Property Protection and U.S. Foreign Direct Investment*, 78 Rev. Econ. & Stat. 181 (1996) *with* Edson Kondo, *The Effect of Patent Protection on Foreign Direct Investment*, 29 J. WORLD TRADE 97 (1995) *and* Kondo, *supra* note 183. *See also* KEITH E. MASKUS, INTELLECTUAL PROPERTY RIGHTS IN THE GLOBAL ECONOMY 171-97 (2000).

204. TRIPS, *supra* note 201.

205. *See* TRIPS art. 31, *supra* note 201.

206. *See* TRIPS art. 31 (f), *supra* note 201.

capabilities nor sufficient demand to support efficient domestic drug production under license. The problem was singled out as critical at the start of the Doha Round of trade negotiations in 2002, and in 2003, agreement was reached on amendments allowing waivers from subparagraph (f) for least-developed nations and for other nations showing that they lack the capacity to manufacture particular pharmaceutical products.[207]

Thus far, the compulsory licensing provisions of the TRIPS agreement have been implemented sparingly—most notably, by Thailand for seven pharmaceutical patents and Brazil for one.[208] But their use has been threatened frequently to induce, especially from multinational pharmaceutical companies, substantial product price concessions or, e.g. in Brazil, voluntary licensing to domestic suppliers at modest royalties.[209] Indeed, even the United States (along with Canada) threatened compulsory licensing in 2001 to elicit substantial price reductions from Bayer AG of Germany on the drug Cipro when terrorist activity threatened an epidemic of otherwise untreatable anthrax.

V. PROPAGANDA

In many contemporary discussions of patent policy, and even in this paper, the term "intellectual property" trips off the tongue as if it were implanted in the human brain's genetically inherited grammar. It is certainly a magical phrase. "Patents" and "copyrights" are words with little or no appeal to the moral sensibilities. But "intellectual property!" What right-thinking person could be against property? And who among the scribbling professions could not be all the more entranced when the property is intellectual?

What strikes a scholar who has been studying patent questions for more than a half century is that the phrase "intellectual property" was almost never heard during the 1950s and 1960s. None of the O'Mahoney Committee's 28 commissioned titles exploring the history,

---

207. World Trade Organization, Decision of the WTO General Counsel of 30 Aug. 2003 (IP/C/W/405). At the December 2005 WTO Ministerial Conference in Hong Kong, the compromise was adopted as a permanent amendment to the TRIPS agreement.

208. *See* Sauwakon Ratanawijitrasin, Pharmaceutical Policy in Thailand, Address at a conference on Pharmaceuticals in the Asia-Pacific (Mar. 2008) (paper available at Stanford University). After three licenses were issued, Thailand in 2007 was put on the United States "Priority Watch List" for alleged TRIPS violations. In 2008, after the issue of four more compulsory licenses, a new Thai government was reconsidering its policies. Rwanda is believed to have imported an AIDS drug from Canada under compulsory license.

209. *See, e.g.*, Paulo Prada, *Brazil Again Seeks to Cut Cost of AIDS Drug*, N.Y. TIMES, Aug. 19, 2005, at C5; *A Conflict of Goals: Brazil's AIDS Programme*, ECONOMIST, May 12, 2007; U.K. COMM. ON INTELLECTUAL PROPERTY RIGHTS, INTEGRATING INTELLECTUAL PROPERTY RIGHTS AND DEVELOPMENT POLICY 43 (Sept. 2002).

implementation, and economic consequences of the patent system during the late 1950s contains the term. A search of the two most comprehensively bibliographic of the O'Mahoney Committee studies and a later Joint Economic Committee study reveals very few cited works, mostly ancient, using the term.[210] It repays effort therefore to investigate how the phrase achieved common currency.

At first, "property" appears to have entered the literature without its "intellectual" modifier. Patent-like privileges were given out by sovereigns in the period of late feudalism, and in the revolutions against feudalism and royal fiat, some acceptable substitute for "privilege" had to be invented. The U.S. Constitution refers to "exclusive rights," but in Europe at the end of the 18th Century, it was de rigueur to refer to a creator's rights in inventions and artistic creations as "property." The usage was not without controversy. In their survey of French antecedents, Machlup and Penrose observe that "those who started using the word property in connection with invention had a very definite purpose in mind: they wanted to substitute a word with a respectable connotation, 'property,' for a word that had an unpleasant ring, 'privilege.' This was a very deliberate choice on the part of politicians working for the adoption of a patent law in the French Constitutional Assembly."[211] The property construction was rejected by Thomas Jefferson, who wrote flatly that "Inventions then cannot, in nature, be a subject of property."[212] Nevertheless, the property concept proved to be durable, and the first world-wide patent treaty, in 1883, was called the Paris Convention for the Protection of Industrial Property.

"Intellectual" was added to "property" much later. The earliest known printed use of the term is in an obscure Massachusetts federal circuit court ruling.[213] Polymath Lysander Spooner used the term in the

---

210. Machlup & Penrose, *supra* note 2; SUBCOMM. ON PATENTS, TRADEMARKS, AND COPYRIGHTS OF THE S. COMM. ON THE JUDICIARY, 85TH CONG., 2D SESS., (Study No. 14, Comm. Print. 1958) *in* JULIUS W. ALLEN , ECONOMIC ASPECTS OF PATENTS AND THE AMERICAN PATENT SYSTEM: A BIBLIOGRAPHY (1958); S.C. GILFILLAN, INVENTION & THE PATENT SYSTEM, Joint Economic Committee of the U.S. Congress (1964).

211. Fritz Machlup & Edith Penrose, *The Patent Controversy in the Nineteenth Century*, 10 J.ECON. HIST. 1, 16 (1950); *see also* Machlup & Penrose, *supra* note 2, at 22.

212. Letter from Thomas Jefferson to Issac McPherson (1813), *in* THE JEFFERSONIAN CYCLOPEDIA 728 (John P. Foley, ed., 1900). A consistent but more extended discussion is found in what appears to have been an earlier letter to McPherson reproduced at 433.

213. Davoll v. Brown, 1 Wood. & M. 53, 7 F. Cas. 197 (Cir. Ct. D. Mass. 1845). (Following his mention of the term, Judge Woodbury cites a Supreme Court decision, Grant v. Raymond, 31 U.S. 218 (1832), but nowhere in that decision is the phrase "intellectual property" found). In a German-language paper available in English only on the world-wide web, Harvard Law School professor William W. Fisher reports a search uncovering one use of the term by the U.S. federal courts during the 19th Century, no uses between 1900 and 1930, two in the 1930s, six in the 1940s, ten in the 1960s, and 41 in the 1980s. William W. Fisher

title of a monograph left incomplete and unpublished around 1855.[214] The term appears four times in French and German titles from the 1860s cited in Machlup's bibliography, mostly addressed to the attack on patent systems being waged in Europe at the time.[215] Its next recorded appearance in American literature titles, gleaned from a search of three major research library catalogs, was in a collection of essays by N.S. Shale in 1878.[216] It then reappears, according to the compendium by Julius Allen,[217] in the titles of three articles published between 1944 and 1952 in the house organ of the U.S. Patent Office, J. OF THE PATENT OFFICE SOC. A published lecture by Sir Arnold Plant titled *The New Commerce in Ideas and Intellectual Property* followed in 1953.[218]

The phrase's takeoff into widespread use may have been associated with the creation of the Geneva-based World Intellectual Property Association (WIPO) in 1966 and its predecessor, United International Bureaux for the Protection of Intellectual Property, founded in 1963. Few intervening references could be found in bibliographies and library catalogs. A seminal role in establishing those organizations was played by Arpad Bogsch, who before their formation was a legal counselor at the U.S. Copyright Office. Obituaries at the time of his death in 2004 called him "the founding father of modern Intellectual Property"[219] and "the creator of the modern intellectual property system."[220] None of the six books, all on copyright, written by Bogsch before 1966 and listed in the Harvard University catalog, included the words "intellectual property" in their title, but he appears to have been an important contributor to their acceptance in popular discourse. He plainly did not create the modern system of granting exclusive rights in inventions and other creative works.

III, The Growth of Intellectual Property (1999), http://cyber.law.harvard.edu/property99/history.html.

214. *See* CHARLES SHIVELY & LYSANDER SPOONER, THE COLLECTED WORKS OF LYSANDER SPOONER chap. VII (1971).

215. Machlup & Penrose, *supra* note 2, at 85-86 (citing works by Molinari, Paillotet, Rentzsch, & Vermeire). The University of Pennsylvania library catalog lists an additional 1859 book by Frederic Passy.

216. *See* NATHANIEL S. SHALER, THOUGHTS ON THE NATURE OF INTELLECTUAL PROPERTY AND ITS IMPORTANCE TO THE STATE (1878).

217. Ratanawijitrasin, *supra* note 208, at 15, 29.

218. Arnold Plant, *The Economic Theory Concerning Patents for Inventions*, 1 ECONOMICA 30 (1934) (Plant's earlier and more famous work which does not use the phrase and contains a remarkably prescient view of first mover advantages as a substitute for patenting).

219. Francois Curchad, *Obituary of Dr. Arpad Bogsch*, AIPPI Update, Jan. 2005, at 2; Press Release, World Intellectual Property Organization, WIPO Director General Expresses Condolences on Passing of Dr. Arpad Bogsch (Sept. 21, 2004), http://wipo.int/edocs/prdocs/en/2004/wipo_pr_2004_389.html*; see also* RYAN, *supra* note 190, at 126.

220. Press Release, World Intellectual Property Association, *supra* note 219.

Other organizations followed suit during the period when the U.S. patent policy reform movement was at its peak. The American Patent Law Association changed its name to American Intellectual Property Law Association and made a corresponding change in the name of its journal (now AIPLA Q. JOUR.) in 1983 or 1984. The relevant section of the American Bar Association was still named the Section of Patent, Trademark & Copyright Law in 1987, but it then changed its name to Section on Intellectual Property Law and in 1993 renamed its quarterly newsletter the *IPL Newsletter* in place of *PTC*[221] *Newsletter*. The ABA sponsored a conference on "Industrial and Intellectual Property: The Antitrust Interface," in October 1984. *The Intellectual Property Journal* was initiated in 1984. During the early 1980s the office of the U.S. President's Special Trade Representative created a new position, Assistant USTR for International Investment and Intellectual Property.[222] The industry lobbying group formed in 1986 to influence deliberations under the Uruguay Round was called the Intellectual Property Committee. In 1989 a revived subcommittee of the U.S. House of Representatives Committee on the Judiciary was named the Subcommittee on Courts, Intellectual Property, and the Administration of Justice. In 1994 the U.S. Senate still had a Subcommittee on Patents, Copyrights, and Trademarks. It was dissolved in 1995 and reborn in 2005 as the Intellectual Property Subcommittee.

Semantics are not policy. But they undoubtedly influence policy-making as well as being influenced by it. The growing use of the term "intellectual property" to describe patent and trademark matters probably contributed to the emergence of a favorable mind set that in turn set the stage for the patent policy reforms of the 1980s.

CONCLUSION

Legislative, administrative, and judicial actions altered U.S. patent policy in significant ways during the 1970s and 1980s. Some of the legislative changes were well-grounded in objective analyses of the problems at hand and what could be accomplished; others, and in particular the centralization of patent appeals in a Court of Appeals for the Federal Circuit, were not. In most cases, the parties with the strongest vested interest in new legislation got what they wanted—most generally, with the exception of the generic drug provisions in the Hatch-Waxman Act, a strengthening of the role patents play in American industrial life. The patent law profession thrived. But the changes brought negative consequences along with the positive. In

221. PTC stands for Patent, Trademark, and Copyright.
222. SANTORO, *supra* note 190, at 9.

particular, by encouraging the proliferation of patents covering inventions of dubious novelty and increasing the statistical probability that knowing or inadvertent infringement of patents leads to dire consequences, it increased the risks as well as the rewards from inventive activity. It is far from clear that the positive effects outweigh the negatives. Fortunately, as economic studies have shown repeatedly, patents do not play a particularly important role in most fields of industrial innovation, and equally fortunately, those who advise industrial leaders in their journeys through the patent minefield are adept at negotiating solutions that in most instances avoid serious impediments to the pace of technological progress. It is nevertheless useful to assess the negatives and attempt to correct them through legislative or judicial action. In this, we would be emulating the example of one of the world's most famous inventors, James Watt, who observed "I have been trying experiments on the reciprocating engine, and have made some alterations for the better and some for the worse, which latter must return to their former form."[223]

On the assumption that the Appellate Court for the Federal Circuit will not be disbanded, one key to improvement is seating judges with a broad perspective on how technological progress is actually induced. Over the long run, this can be achieved if the President and Senate, in exercising their powers of appointment and consent, insist that nominees be persons of broad experience and wisdom and shun nominees representing a narrow interest group—e.g., the patent bar. In practice, to be sure, judges with the capabilities of an Oliver Wendell Holmes or a Learned Hand are rare and best-suited for higher responsibilities. At minimum, therefore, nominees to the court should be subjected to a searching examination on their knowledge about how innovation takes place in the real world. Appropriate preparatory readings can be suggested.[224] Effecting a transformation in the composition of the Court

---

223. Letter from James Watt to Dr. William Small (Jan. 28, 1769), *in* JAMES. P. MUIRHEAD, THE ORIGIN AND PROGRESS OF THE MECHANICAL INVENTIONS OF JAMES WATT 36 (vol. 1 1854).

224. For example, in addition to the standard legal texts: in law, ROCHELLE C. DREYFUSS, HARRY FIRST & DIANE L. ZIMMERMAN, EXPANDING THE BOUNDARIES OF INTELLECTUAL PROPERTY: INNOVATION POLICY FOR THE KNOWLEDGE SOCIETY (2001); Rebecca Eisenberg & Michael Heller, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCIENCE 698 (1998); Robert Merges & Richard Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM L. REV. 839 (1990). In sociology, BERNARD BARBER & WALTER HIRSH, THE SOCIOLOGY OF SCIENCE (1972). In economics: WILLIAM J. BAUMOL, THE FREE-MARKET INNOVATION MACHINE: ANALYZING THE GROWTH MIRACLE OF CAPITALISM (2003); LEWIS M. BRANSCOMB & PHILIP E. AUERSWALD, TAKING TECHNICAL RISKS: HOW INNOVATORS, MANAGERS, AND INVESTORS MANAGE RISK IN HIGH-TECH INNOVATIONS (2003); BURTON H. KLEIN, DYNAMIC ECONOMICS (1977); Levin, *supra* note 31; Mansfield, *supra* note 26; NATHAN

is likely, however, to take at least a decade. In the interim, it would be desirable for the highest judicial authorities to encourage attendance of ACFC judges at broad-ranging seminars on the science, sociology, and economics of technological innovation. These should be quite different from the outings organized at posh spas by special interest groups. They should be planned and operated by a reputable university faculty and staffed by scholars with a diverse range of interests and biases.

Absent such remedies, the ACFC's worst abuses can be checked by active Supreme Court rejection of the Federal Circuit's decisions. This happened in May 2006 when the Supreme Court articulated more stringent guidelines for the issuance of injunctions in the *eBay* case.[225] It happened again in April 2007 when the Court demanded more careful scrutiny of inventions claiming novel ways of applying well-known concepts.[226] Among other things, the Court exhibited commendable social science insight into the dynamics of invention:

> We build and create by bringing to the tangible and palpable reality around us new works . . . . These advances, once part of our shared knowledge, define a new threshold from which innovation starts once more. And . . . the results of ordinary innovation are not the subject of exclusive rights under the patent laws. Were it otherwise patents might stifle, rather than promote, the progress of useful arts.[227]

Congressional clarification of key concepts might also help. It remains uncertain whether reforms being considered in 2008, but stalemated at the time this article was revised, will be sufficient to do the job.[228]

With the Bayh-Dole Act, the key open challenge is balancing the interest in exclusive rights against the broader public interest in securing maximum public benefit from the government's investments in basic science. As a general principle recognized by the law's drafters, exclusivity helps to stimulate investment in development and commercialization. But there was recognition, at least at the time the law was enacted, that abuses might occasionally require the exercise of the law's march-in rights. Congress should reiterate that it intended a

---

ROSENBERG, INSIDE THE BLACK BOX: TECHNOLOGY AND ECONOMICS (1982).

225. eBay Inc., 547 U.S. 388.

226. KSR Int'l Co. v. Teleflex Inc., 550 U.S. 398 (2007). In *KSR* the Court appears to have been returning to the more critical stance adopted by Thomas Jefferson, who drafted the first U.S. patent law and served implicitly as the first federal patent examiner. *See* Letters from Thomas Jefferson to Oliver Evans (1814) and Isaac McPherson (1813), *in* THE JEFFERSONIAN CYCLOPEDIA, *supra* note 212, at 680.

227. *KSR Int'l,* 550 U.S. at 727.

228. For an analysis of the pending measures, see JOHN R. THOMAS & WENDY H. SCHACHT, Cong. Res. Serv., PATENT REFORM IN THE 110TH CONGRESS (2008).

balance to be struck. It should create a special panel with the difficult task of determining when exclusive rights on government-supported inventions have been abused and the extent of licensing required to set matters right.

The Hatch-Waxman Act was in many respects an ideal compromise, trading longer periods of patent protection to compensate for regulatory lags with speedier entry of generic drugs into production once blocking patents have expired. The threat of generic entry in turn spurs pharmaceutical firms to redouble their R&D efforts in order to replenish their new product pipelines. The main problem with Hatch-Waxman is that drug developers have exhibited great ingenuity in finding ways to extend their periods of patent protection by accumulating patents on minor variants of the originally proven molecule and paying the first-moving generic entrant not to enter, using a loophole in the law to block the entry of other would-be generic producers. Congress should clarify the law, remaining faithful to the Constitutional requisite that exclusive rights be "for *limited* Times"[229] and insisting that drug production be opened up for generic competition once basic patents have expired, leaving however the right to produce validly patented improvement molecules exclusively in the hands of the original drug developer (or any other firm that patents and tests improved variants).

For the federal antitrust agencies, the extension of patent monopolies in time through profuse improvement patenting and their extension in scope through restrictive cross-licensing agreements pose important enforcement problems. Here too, the problem is in part one of education. Those who manage the antitrust agencies need to learn that there are important barriers to rapid imitation, enhancing incentives for innovation, other than the patent system, so maximization of monopoly rewards associated with patent holdings is unlikely to maximize economic welfare. These agencies need to learn that extension of patent monopolies over time and in scope is more likely to suppress than stimulate innovation.[230] They need to learn, as my colleagues and I did a half century ago, that compulsory licensing of patents is not likely to decimate firms' incentives for investment in innovation. Knowing this, they may come to appreciate that carefully considered intervention in cases of protracted and abusive monopoly through patenting can on balance be beneficial. The emphasis, to be sure, is on careful consideration, clear precedents, and appropriate timing.

Without doubt the most important of the issues addressed in this

---

229. U.S. CONST. art. I, §8 (emphasis added).

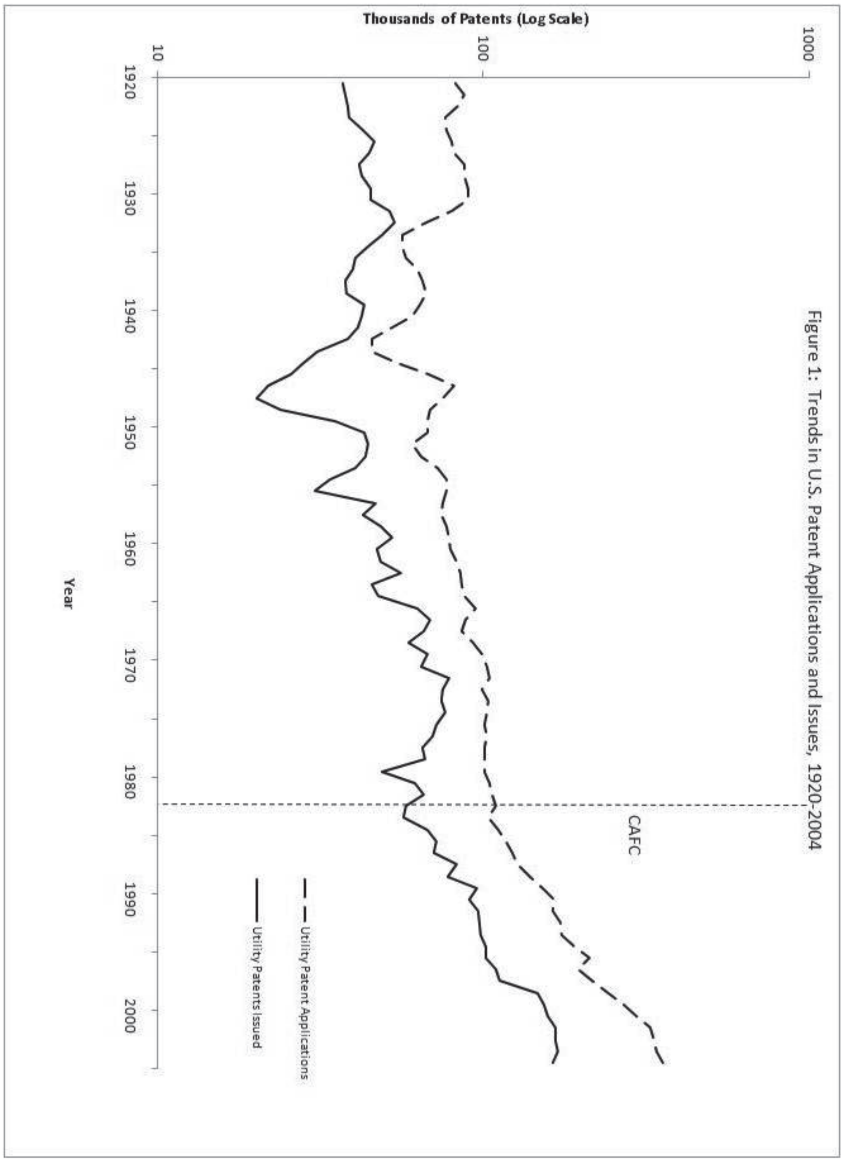
230. *See* Merges & Nelson, *supra* note 224; for seven case studies, F. M. Scherer, *Technological Innovation and Monopolization*, *in* ISSUES IN COMPETITION LAW & POLICY (Wayne Dale Collins ed., 2008).

paper is the extension of first-world patent standards to third-world nations under the Treaty of Marrakech. At their present stage of development, having to confer patents on first-world products is likely to reduce, not enhance, the welfare of hard-pressed low-income nations. The United States and other rich nations should not undertake retaliatory measures against less-developed nations that exercise their clear right under the Treaty to order compulsory licensing, import patented drugs from other low-price nations, or limit the scope of patent protection on borderline products.[231] Even when they allow patent rights to be exercised, their demand and the monopoly profits that can be derived from it are unlikely to be sufficient to stimulate greatly increased inventive activity in the first world.[232] Among other things, their demand is too weak to stimulate much development of new drugs targeted toward tropical diseases, i.e., those prevalent only in the third world. But it is at least arguable, even if not universally accepted, that the rich nations have an obligation to help their fellow humans in this regard. This means that in rich nations, public and philanthropic funds should be generously allocated to foster the development and distribution of drugs and vaccines whose main use will be to lessen the burden of disease in the third world. This will be a step back from the Machiavellian logic that underlay negotiation of the Marrakech Treaty, but it will be a step forward for humanity.

231. *Cf.* Deardorff, *Should Patent Protection Be Extended to All Developing Countries?*, *supra* note 202; Deardorff, *Welfare Effects of Global Patent Protection*, *supra* note 202.

232. *See* Scherer, *supra* note 202.

APPENDIX: FIGURES†



†Figure 1 Data: U.S. Patent Office, U.S. Patent Activity Calendar Years 1790 to the Present: Table of Annual U.S. Patent Activity Since 1790 (Mar. 2009), http://www.uspto.gov/go/taf/h_counts.pdf; Figure 2 Data: SCIENCE & ENGINEERING INDICATORS, *supra* note 149.

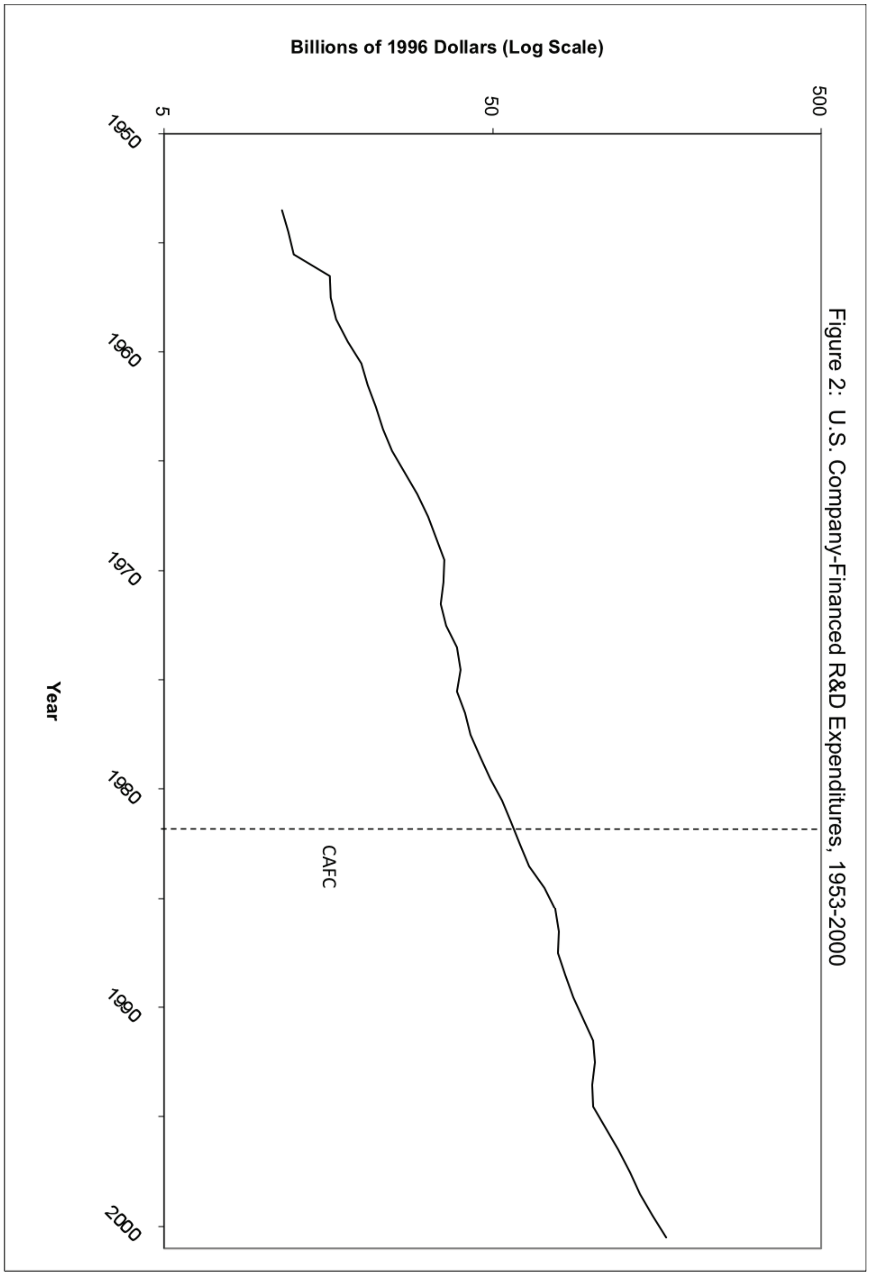


Figure 2: U.S. Company-Financed R&D Expenditures, 1953-2000

THE NEW "EMERGENCE ECONOMICS" OF INNOVATION AND GROWTH, AND WHAT IT MEANS FOR COMMUNICATIONS POLICY

RICHARD S. WHITT AND STEPHEN J. SCHULTZE*

It is not the strongest of the species that survives, nor the most intelligent, but the one most responsive to change.[1]

* Mr. Whitt currently is Washington Telecom and Media Counsel at Google Inc. Mr. Schultze served as a policy intern with Google, and now works as a fellow at Harvard Law School's Berkman Center for Internet and Society. This paper reflects personal views held by Mr. Whitt and Mr. Schultze, and not necessarily those of their current employers. The authors would like to thank Rob Atkinson, Susan Crawford, George Ford, Michael Pelcovits, Hal Varian, Phil Weiser, Kevin Werbach, and Tim Wu for their thoughtful comments and suggestions on various versions of this paper. While the final product benefited enormously from their contributions, any remaining infirmities large or small obviously are the sole responsibility of the authors for not listening well enough.

1. Although this aphorism summarizes well a strand of evolutionary thinking and frequently is attributed to Charles Darwin, its initial author is unknown. *See, e.g.*, John van Whye, *It Ain't Necessarily So*, GUARDIAN, Feb. 9, 2008 (Features & Extracts), at 10.

INTRODUCTION

As the financial crises of 2008-09 amply illustrate, your typical member of Congress or White House staffer probably knows less about economics than ordinarily is assumed. And what they think they know has been superseded by newer, more robust, and more accurate forms of thought.

The hoary economics presented to us in public policy debates maintains, for example, that the market is linear and always seeks equilibrium, that economic actors are perfectly rational, with perfect knowledge of themselves and the marketplace, that production is generated only by capital markets or government subsidy, that growth is exogenous, and the whole of the economic system is always equal to the sum of its parts. It turns out that every one of these key assumptions is either overstated, or plain wrong. Recently, the U.S. economy has been paying a heavy price for some of these flawed assumptions.

This paper will introduce the rough formula for what we call here "Emergence Economics"—namely, that individual agents, acting through interconnected networks, engage in the evolutionary market processes of differentiating, selecting, and amplifying certain business plans and technologies, which in turn generates a host of emergent economic phenomena. This formula is fueled by the latest findings from physics, biology, psychology, cognitive neuroscience, and plain common sense. The Internet then will be discussed as a notable and perhaps unique product of market and non-market forces, as modular infrastructure, and as a platform for broad-based innovation. Next, the paper will turn to some key emergent phenomena, including ideas, innovation, economic growth, and what we call "Net effects." Finally, we will bring these economic and technological elements to bear in the world of communications policy, where a proposed new framework separates out the virtues of "tinkering" with market gaps and inputs, versus the vices of "tampering" with evolutionary processes and outcomes.

Back to School

Today's discussions about national communications policy often seem to be rooted to the past, in the form of economic and technology assumptions that more or less ended in the 1960s. As it turns out, the rise of new economic thinking, along with new technology platforms culminating in the Internet, directly challenge many of those chief assumptions. Now is the time to articulate the fundamental economic and technology tenets that should inform our nation's communications and information policies, and to begin suggesting some ways those policies should be recast in their light.

We don't deny that today's economics of academia currently incorporates much of the schools of thought that collectively we refer to here as "Emergence Economics," or that certain strands of this thought have surfaced repeatedly throughout the history of economics. It is certainly not our intention to create, or settle, a dispute over whether and how the various facets of Emergence Economics are being incorporated into presumed policy economics, here called "Old School Economics." Our only point is that the prior version represents the story of economics as told to, and accepted by, most of our nation's policymakers and thus has become a bulwark of official thinking about public policy issues. By contrast, the various schools of Emergence Economics simply are not familiar to most policymakers. It is our hope that, little by little, this reality can begin to change.

A New Economy

Twenty years ago, no one would have anticipated the Internet as we know it today. Consumer-grade computer modems were still in their infancy, and the few dial-up online communities that existed were a far cry from the globally connected "network of networks" that now pervades so much of what we do. In many ways, the Internet was a happy accident. What started out as isolated islands—universities, bulletin board systems, commercial services—linked together and grew as the result of the actions of millions of unaffiliated people. The underlying software protocols opened up the ability to interact and speak freely across thousands of interconnected networks. The growth of the Internet, and the online and offline economy it facilitates, was beyond even the wildest predictions. In many ways, the Internet is what happened while we were busy planning something else.

Even with the benefit of hindsight, we do not fully understand what led to this success. Advances in computer technology, including digitization of information, dramatic increases in computing power, and concomitant declines in the cost of computer storage, were necessary conditions, but they also could have facilitated a variety of other outcomes. To comprehend how the Internet developed into a thriving marketplace of innovation and economic growth, many turn understandably to the field of economics. Unfortunately, Old School Economics has little to say about this leading test case.

A New Framework

Two fundamental observations drawn from Emergence Economics lead us toward a more helpful analytic framework.

First, we live in a *networked* economy, formed bottom-up by interactions between people in a highly connected marketplace. Some basic rules govern these interactions, but for the most part the system emerges freely and unpredictably. Economic actors become nodes, and the structure of the market evolves based on their collective practices. Whereas Old School Economics more often than not attempts to statically engineer these relationships, Emergence Economics recognizes that such an abstraction fails to accurately represent a far messier reality. Instead, the network economy thrives when there is space for experimental evolution, in which new ideas emerge and technology constantly is refined.

Second, we live in a *growth* economy in which the chief currency is ideas, and the mechanism for growth is innovation. While Old School Economics tells us that productivity comes simply from adding more capital, or generating greater efficiency, Emergence Economics emphasizes ways in which technologies, broadly defined, transform the means of production. The advent of new technologies can create better recipes for economic growth. What's more, these technologies do not emerge inexplicably from outside the system, but instead are the product of economic actors working within (indeed, comprising) the system according to diverse motives. In addition, growth is not limited by physical goods, but is enabled by the transmission, reproduction, and improvement of ideas. An economic framework that actually recognizes this dynamic can establish a foundation for even greater growth in the future.

This networked growth economy differs greatly from some traditional economic models that posit static, linear forms of growth. Economists often use the phrase "virtuous circle" to describe systems that contain positive feedback loops in which their outputs cycle back into their inputs. Rather than moving toward equilibrium, these economies are self-reinforcing and have the potential to multiply their effects in unexpected ways—generating both positive and negative feedback. The virtuous circle often appears in the form of technological innovation that facilitates future technological innovation and drives "network effects" where each new user adds benefit to the rest of the system. Emergent economies combine these dynamics, harnessing the discoveries of others in such a way that the system as a whole grows far more effectively than if it were a disconnected set of actors. The interdependent virtuous circles can become part of a complex *virtuous feedback* network.

And so, what of the proper role of the government policymaker—the legislator, the regulator, the reviewing judge—in the face of an innovation-fueled, network-connected, emergent economy? There is little doubt that the policy environment needs to catch up to newer economic thinking. As we shall see, while the lessons here are many, at bottom they point to caution, and even outright skepticism, about becoming a more active force in the market. Such caution should be tempered by optimism. The tools of government, when employed sparingly, carefully, and in the right context, can improve the environment for new ideas, economic growth, and human freedom. However, this bottom-up regulatory approach requires an appreciation for, and understanding of, esoteric-seeming topics like network-based dynamics, and the conditions that are most likely to foster productive tipping points.

Some Caveats

We have a few important caveats to relay at the outset. First, the literature in this field is broad and deep in some places, scanty and unfinished in others. Nonetheless, as Michael Shermer rightly declares, economics has been undergoing "the most dynamic revolution since Adam Smith," because "rich transdisciplinary hybrids are emerging to breathe new life into an old science . . . ."[2] This paper necessarily presents an overview of what so far is known, or surmised, or even guessed at, but it does not purport to replace the foundational work of many others. Nor do we seek to discredit the world of modern economic thinking; indeed, as previously noted, many of the intellectual trends we discuss here slowly but surely are being incorporated into more mainstream schools. In particular, we look to Joseph Schumpeter and Friedrich Hayek as two towering transitional figures linking together various strands of Old School Economics and Emergence Economics. The chief concern is that, while mainstream economic thinking today appears to be incorporating new intellectual frontiers, many policymakers still cling to the basic assumptions underpinning older forms of economic theory, as if those assumptions remain received wisdom. As a result, our objective here is comparatively humble, yet practical: to condense what currently comprises this sprawling body of advancing work in a way that can aid policymakers and others in analyzing public policy issues, particularly in the communications realm.

In addition, the coming discussion of "markets" and "systems" and "properties" should not cause us to lose sight of the common human element. Whatever happens in any agent-constructed space, such as economics, should be taken to reflect humanity in all its breadth and depth. Words and concepts are not the things they describe; they merely serve as organizing principles to try to make sense of a seemingly disorganized world. Ironically, economics, technology, and law—

2. MICHAEL SHERMER, THE MIND OF THE MARKET xix (2008).

together which form the three-way intersection for this paper—often are perceived as sterile, artificial, and even tedious fields. We would argue instead that they should be seen as the flesh-and-blood instantiation of ordinary humans participating actively in the world.

Moreover, it must be stressed that all economic models inherently are wrong to one degree or another. As mere abstractions of reality, they inevitably miss at least some of the nuance and sinew of the world that each of us inhabits. While the concepts explored here should not be confused with exacting corollaries to the real world, nonetheless, they do provide an important corrective to what has gone before.

Finally, we must point out that neither author of this paper is an economist, at least (yet) by training or trade. Perhaps some will dismiss what follows solely for that reason; after all, the saying goes, "the only thing more dangerous than an economist is an amateur economist."[3] Still, the world of economic theory should not only be available to, and articulated by, a cloistered few.[4] As Old School Economics gradually, and we think inevitably, gives way in the minds of policymakers to new forms of economic thinking, we can only hope that more of us are able to join in the evolving conversations.

I. EMERGENT ECONOMY: OVERTHROWING THE OLD REGIME

A. Introducing Emergence Economics

So what exactly is this supposedly new form of economics? Analyst Eric Beinhocker adopts the term "Complexity Economics" to describe a variety of different analytic and empirical approaches to the economy, with a more exacting faithfulness to the real world. In this paper we combine Beinhocker's impressive synthesis with emerging work in other areas, most notably network science, new growth theory, and behavioral economics. Other useful contributions include the origins of innovation, competition theory, "Net effects" like spillovers and social production, and economic sociology. As a result, we thought it best to avoid confusion by using a broader umbrella phrase, "Emergence Economics," to denote this more wide-ranging set of emerging viewpoints that have found, or are finding, their way into mainstream economics.

In reality, this "new" economics draws from several long existent but submerged themes in neoclassical economics. Phenomena once thought

3. Economist Jokes, http://netec.mcc.ac.uk/JokEc.html.

4. We also take heart from Hayek's statement that "an economist who is nothing but an economist cannot be a good economist." F.A. HAYEK, THE FORTUNES OF LIBERALISM 196 (Peter G. Klein, ed., 1992).

to be minimally important, peripherally relevant, or outside the scope of proper economic thinking altogether are being brought to the forefront. In the case of communications-based technology sectors, for example, economist Hal Varian has noted that:

> [T]here are some forces that are particularly important in high-tech . . . [T]hese forces are not "new"; indeed, the forces at work in network industries in the 1990s are very similar to those that confronted the telephone and wireless industries in the 1890s.[5]

Varian undoubtedly is correct—many of the various elements at play in today's economy have been there for many decades. Perhaps the best way to understand Emergence Economics is to compare it to the earlier, more traditional form of economics, which was handed down to us by the neoclassical theorists, synthesized by post-war economists, and still remains the intellectual grounding for many of today's political theories. Below, we will look briefly at four aspects of what we call Old School Economics—markets, competition, people, and analysis—and contrast them with the different perspectives brought by Emergence Economics.

One gating question first must be addressed, however: if Emergence Economics actually presents a more truthful version of the economic landscape, and has been or is being incorporated into more mainstream economic theory, why have these ideas taken so long to be absorbed into the social and political mainstream? Economist Paul Ormerod cites as likely reasons sheer intellectual inertia and, until recently, a lack of sophisticated analytic tools.[6] Unfortunately, as we shall see, public policy in the United States still tends to hew closely to the dictates of Old School Economics. In a small way, this paper seeks to offer a much-needed corrective to that situation.

1. The Nature of Markets: Complex Cascades

Neoclassical theory states that the economy is a static equilibrium system, existing at rest, and moving from one equilibrium point to another as it seeks balance. Under this view, the economy literally is a

---

5. H. VARIAN, ET AL., THE ECONOMICS OF INFORMATION TECHNOLOGY 3 (2004).

6. PAUL ORMEROD, WHY MOST THINGS FAIL ix (2005). One author goes so far as to claim that traditional economist theories employ an explicitly religious defense of the marketplace. *See* ROBERT NELSON, ECONOMICS AS RELIGION (2001). Another author notes that some supporters of traditional economics want so badly to believe in their models that, "paradoxically, these economic formulas and models were symptoms of the very desires and emotions they were designed to eliminate." MARK TAYLOR, CONFIDENCE GAMES 276 (2004).

closed equilibrium system.[7] Neoclassical theory also sees an "invisible hand" at work in competitive markets. In a free market economy, the thinking goes, a natural resting point is reached, where supply equals demand, resources are put to their most efficient use, and the welfare of society is optimal. Such a market is deemed optimally efficient. The business cycle is just that: a regular, periodic, and predictable movement between boom and recession. Economic processes are dominated by dampening, negative feedback that keeps things contained. Any indeterminacy typically is assumed away by econometric models.[8] The business cycle is determined exogenously, by occasional shocks originating from outside the system itself.[9]

By contrast, a central tenet of Emergence Economics is that the economy is a "complex adaptive system," which is a subcategory of open systems.[10] In a complex adaptive system, micro-level interactions lead to the emergence of macro-level patterns of behavior. A key aspect of any complex adaptive system is the inherent lack of predictability in its future operations, because they do not add up in a simple, linear way. As one example, financial markets are far more turbulent, deceptive, and risky than previously thought, with prices leaping up and down in a more or less concentrated fashion.[11] Oftentimes, small, innocuous events can set off avalanches of change that are inexplicable in Old School Economics, while large disturbances ultimately may have no lasting impact. The direction of the stock market relies on the actions of millions of individual agents, motivated by innumerable and interrelated concerns. Dips and peaks in the economy refuse to recur in a predictable manner.[12] To quote 2002 Nobel Prize winning economist Vernon Smith, "We do not understand why markets work as they do."[13]

Further, the economy is not an equilibrium system at all, and will never reach a resting place, for such a state would equal death. Ormerod

---

7. *See* TAYLOR, *supra* note 6, at 239-40.

8. ORMEROD, *supra* note 6, at 79-80.

9. *Id.* at 191. Léon Walras, champion of economic equilibrium theory, famously noted, "For, just as a lake is, at times, stirred to its very depths by a storm, so also the market is sometimes thrown into violent confusion by crises, which are sudden and general disturbances of equilibrium." LEON WALRAS, ELEMENTS OF PURE ECONOMICS 381 (William Jaffe trans., George Allen and Unwin Ltd. 1954) (1874).

10. W. BRIAN ARTHUR, ET AL., THE ECONOMY AS AN EVOLVING COMPLEX SYSTEM II (1997). Or, as Michael Shermer puts it, economies are complex systems "that emerge out of the simple actions of people just trying to make a living and provide for their children." SHERMER, *supra* note 2, at 5.

11. BENOIT MANDELBROIT, THE (MIS)BEHAVIOR OF MARKETS 225-52 (2004).

12. PHILIP BALL, CRITICAL MASS 189 (2004).

13. Vernon Smith, Nobel Prize Lecture at the Interdisciplinary Center for Economic Science at George Mason University: Constructivist and Ecological Rationality in Economics, (Dec. 8, 2002), at 506 n.14, *transcript available at* http://nobelprize.org/nobel_prizes/economics/laureates/2002/smith-lecture.pdf).

points out that there are numerous empirical and theoretical bases to criticize the concept of general equilibrium.[14] The real world also exhibits positive feedback, or increasing returns; there are always new sources of positive feedback, so there is no "long run" in the real world. Time is an important element of economic phenomena, one missing in general equilibrium theory.

The idea that the market is optimally efficient also runs headlong into reality. Lawrence Summers has noted that careful analysis "call[s] into question the theoretical as well as empirical underpinnings of the Efficient Market Hypothesis,"[15] while Yale's Robert Shiller claims that the "efficient market hypothesis is the most remarkable error in the history of economic theory."[16] Mark Taylor, a noted expert on complexity theory, agrees that the efficient market hypothesis and related theories "were wrong on virtually every count," so that "the more carefully one ponders the markets, the more suspect the whole notion of efficiency becomes."[17]

Old School Economics is riddled with these basic flaws for a good reason. It is widely understood that neoclassical economic theory has viewed the economy through the prism of the physical sciences of the late 19th Century, particularly atomistic statistical mechanics.[18] The metaphor appropriated from the Industrial Revolution is the economy as a human-made machine (like a steam engine), whose behavior is fixed, stable, predictable, and controllable.[19] This perspective emphasizes static rules executed by top-down hierarchies of relatively expert and impartial officials who prize efficiency and consistency.[20] By contrast, Emergence Economics views the economy through the prism of modern day physics and biology, with metaphors better suited to the Information Revolution. Under that perspective, human society, of which the economy is a subset, is more like a living organism—a complex system characterized by constant change, evolution, and disequilibrium that percolate from the bottom up.[21]

---

14. ORMEROD, *supra* note 6, at 50-51.

15. Lawrence H. Summers, *Does the Stock Market Rationally Reflect Fundamental Values?*, 41 J. OF FIN. 591, 592 (1986).

16. KEVIN PHILLIPS, BAD MONEY-RECKLESS FINANCE, FAILED POLITICS, AND THE GLOBAL CRISIS OF AMERICAN CAPITALISM 78 (2008). As Warren Buffet put it, "I'd be a bum on the street with a tin cup if markets were always efficient." *Id.*

17. TAYLOR, *supra* note 6, at 269, 273.

18. *See* BALL, *supra* note 12, at 204-06; TAYLOR, *supra* note 6, at 273; ORMEROD, *supra* note 6, at 17-35.

19. *See generally* ROBERT AXELROD & MICHAEL D. COHEN, HARNESSING COMPLEXITY 28-31 (1999); PAUL ORMEROD, BUTTERFLY ECONOMICS (2000).

20. AXELROD & COHEN, *supra* note 19, at 28-31.

21. *See* ORMEROD, *supra* note 19. Crucially, the neoclassical economists failed to include in their thinking the Second Law of Thermodynamics, which describes entropy and the notion of an open system.

One illustrative example of the different approaches is the prices for goods and services. Traditionally one of the central tensions in economics is between producers (supply) and consumers (demand). Supply and demand supposedly balance out precisely; there is no waste, and goods are distributed in a "Pareto optimum" way. Further, to maximize a firm's profits, the price should always be set equal to marginal cost (or where additional revenues will exceed additional costs).[22] Relatedly, stock prices accurately reflect all available information at all times, and so should follow a "random walk," with no patterns or clues for future prices.[23]

Emergence Economics challenges each of these assumptions.[24] Supply rarely equals demand. Empirically, we often see a wide divergence in the prices of individual goods and services. Demand and cost curves are extremely difficult to know with any clarity.[25] Uncertainty and lack of information shroud the future in doubt, which makes setting prices that much more difficult.

2. The Nature of Competition: Imperfect Incentives

The traditional view of economic competition is that "perfect competition" compels free markets to allocate scarce resources in a manner so efficient that all our conflicting wants and needs are resolved in the most satisfactory manner possible.[26] The watch phrase is "consumer sovereignty."[27] Free markets are presumed to be both efficient and fair. The efficiency principle says that under perfect competition, and with no market failures, free markets will squeeze as many useful goods and services as possible out of the available resources (maximal output at minimal prices), and that anything that interferes with the price system's ability to do so is a detriment to social well-being.[28] The concept that the supply of every traded good or service is precisely equal to the demand for it at prevailing prices led to the related concept that the economy rests in perfect equilibrium.[29]

Now, however, the study of competition slowly is turning from a

---

22. ERIC D. BEINHOCKER, THE ORIGIN OF WEALTH, 60-62 (2006).

23. TAYLOR, *supra* note 6, at 244-46.

24. Robert Nelson calls these theories "an economic tautology." NELSON, *supra* note 6, at 58-69.

25. ORMEROD, *supra* note 6, at 23-35; BALL, *supra* note 12, at 254-55.

26. Ormerod observes that the phrase "perfect competition" is "yet another example of the terrifying ability of economists to brand their central concepts so effectively." ORMEROD, *supra* note 6, at 83-84.

27. This concept was first formulated by Ludwig van Mises. *See* G. Stolyarov II, *The Concept of Consumer Sovereignty in Economic Theory*, HELIUM, http://www.helium.com/items/112764-the-concept-of-consumer-sovereignty-in-economic-theory.

28. JAMES CASE, COMPETITION 160-63 (2007).

29. *Id.* at 183-84; BALL, *supra* note 12, at 191.

theoretical to an experimental science. And the "free market rarely, if ever, operates under conditions of perfect competition . . . ."[30] Further, careful analysis shows that most firms are not pure maximizers of profit or utility, but also seek to attain other primary objectives such as attracting and retaining productive workers.[31] Common diversions to raising prices include predatory pricing, discriminatory pricing, price-fixing, attempts to monopolize, and unfair and deceptive advertising claims. In addition, the demonstrated phenomenon of "sticky prices" leads to higher than necessary prices that refuse to descend to their theoretically sanctioned levels.[32]

Old School Economics also has failed to convey the reality of modern-day competition in another fundamental way. The classical world view is founded on scarce material objects and their efficient allocation—or as Paul Romer puts it, "a finite quantity of things with which we can work—basically, the matter in the earth's crust."[33] Value comes from rearranging that matter into a more valuable form. The physical world is characterized by diminishing returns, and increasing cost per additional unit produced. Firms compete via prices in existing goods, and laws were built around establishing property rights and ensuring no monopoly control. In the Information Age, however, we rely increasingly on ideas as "the recipes we use to rearrange [matter] to create more value and wealth."[34] Because ideas are not scarce, the process of discovering them does not suffer from diminishing returns. Indeed, increasing returns come from both the "shoulders-of-giants" process (new ideas build on existing ideas, and then beget more new ideas), and falling costs per unit (such as producing a software CD). Competition is facilitated not by firms trying to drive prices, but by firms seeking to capture market share through new products.

Joseph Schumpeter was an original prophet in this area. He saw claims about "perfect competition" as relatively unimportant; instead, what counts is "competition from the new commodity, the new technology, the new source of supply, the new type of organization. . . competition which. . . strikes not at the margins of the profits and the outputs of the existing firms but at their foundations and their very lives."[35] Schumpeter also argued that some degree of monopoly is preferable to perfect competition, because supernormal profits are the

---

30. BALL, *supra* note 12, at 255.

31. *Id.* at 268-69.

32. CASE, *supra* note 28, at 196.

33. Joel Kurtzman, *An Interview with Paul Romer*, STRATEGY+BUSINESS, First Quarter 1997, http://www.strategy-business.com/press/16635507/9472.

34. *Id.*

35. JOSEPH SCHUMPETER, CAPITALISM, SOCIALISM, & DEMOCRACY 84 (Harper Perennial 1976) (1942).

temporary fruits of innovation. Persistent competition from innovations creates a threat that "disciplines before it attacks," so that as soon as innovators cease to innovate, pricing power will desert them.[36]

3. The Nature of People: Behavioral Beings

Under Old School Economics, the pursuit of self-interest is a rational activity of individuals, based on utility (a measure of pleasure and pain). Paul Samuelson, a towering figure in 20th Century economics, assumed that people are "representative agents," both logical and consistent in their behaviors, and have perfect knowledge of all probabilities. Old School Economics also includes assumptions about *homo economicus* possessing "rational and errorless choice, presupposing perfect foresight," and "foreknowledge free from uncertainty."[37]

Unlike the academic equations and assumptions that undergird Old School Economics, behavioral economics is based on the actual discernible rules by which humans make everyday decisions.[38] In the real world, barriers to decision making almost always exist. Information is costly, incomplete, and rapidly changing. At best, we employ "bounded rationality," by making decisions in the face of obvious external and internal constraints.[39] Ormerod observes that every individual decision involves massive complexity and defies the orderly application of the rational calculations of economic theory. Indeed, "in the new economics, we not only address a specific problem, we try to start from the outset with rules of behaviour which have empirical support rather than with rules which we believe *a priori* a rational agent ought to follow."[40] We will discuss this critical point in further detail in Part II.

Further, there is no such thing as self-interested individuals acting in isolation. Hayek showed us that desirable outcomes are the joint product of both individual actions and the institutional framework in which individuals operate.[41] Social change is both volatile and often inexplicable, as agents engage in "clustering" and "herding" behavior. Ample evidence demonstrates that, in Beinhocker's words, "the

---

36. *Id.* at 85. Some modern-day students of "imperfect competition" do not necessarily agree with this point, as some oligopolies may cling to pricing power indefinitely. One salient example mentioned in the literature is Microsoft, which could have sold its Windows operating system for $49, but instead chose a profit-maximizing price of $89. *See* CASE, *supra* note 28, at 206. Indeed, these same economists observe that free markets long have tolerated all manner of supernormal profits, and such markets tend to evolve into tight oligopolies over time. *Id.*

37. CASE, *supra* note 28, at 199; BALL, *supra* note 12, at 209-11.

38. BALL, *supra* note 12, at 213-14.

39. *Id.* at 211-12.

40. ORMEROD, *supra* note 6, at 125.

41. *Id.* at 224.

interactions of millions of people, making decisions, engaging in strong reciprocal behavior, acting out their cultural norms, cooperating, competing, and going about their daily lives, creates an emergent phenomenon that we call society—a phenomenon as real as the emergent pattern of a whirlpool."[42]

4. The Nature of Analysis: Mismatched Models

Traditional economic theory has proven inadequate in terms of the two standard criteria for a scientific theory: prediction and explanation. Models in Old School Economics often use simplifying and highly restrictive assumptions. Famously, Milton Friedman insisted that unrealistic assumptions in economic theory do not matter so long as the theories make correct predictions.[43] Such optimism would seem misplaced. Indeed, all mathematical statements are conditional in nature. Assumptions must be appropriate for the purpose of the model, and must not affect the answers the model provides for that purpose.[44] In econometrics, statistical correlations do not provide a causal explanation of the phenomena, and data often is not readily available or is problematic. Paul Ormerod believes that, "to be of any value, theories must be confronted with reality."[45] Philip Ball explains further that:

> Economic models have been augmented, refined, garlanded, and decorated with baroque accoutrements. Some of these models now rival those constructed by physicists in their mathematical sophistication. Yet they still lack their "Newtonian" first principles: basic laws on which everyone agrees.[46]

There also is an uncomfortable feeling that economic models oftentimes lose the human element in their too-neat equations. We want to believe that economic theory does not regard us as "automata" and

---

42. BEINHOCKER, *supra* note 22 at 450.

43. *See, e.g.*, Milton Friedman, *The Methodology of Positive Economics, in* ESSAYS IN POSITIVE ECONOMICS 3, 30-31 (1953) ("Perennial criticism of 'orthodox' economic theory as 'unrealistic' . . . because it assumes man to be selfish and money-grubbing . . . ready to change prices and/or pricing rules whenever their sensitive intuitions . . . detect a change in demand and supply conditions" and "assumes markets to be perfect, competition to be pure, and commodities, labor, and capital to be homogeneous . . . is largely beside the point unless supplemented by evidence that a hypothesis differing in one or another of these respects from the theory being criticized yields better predictions for as wide a range of phenomena.").

44. Robert Atkinson notes, for example, that "innovation changes the quality of capital. If all you can measure is quantity, you're going to miss the real story . . . . In short, we need to look at the real economy as it plays itself out over time in the millions of workplaces in the nation." ROBERT D. ATKINSON, THE PAST AND FUTURE OF AMERICA'S ECONOMY 147 (2004).

45. ORMEROD, *supra* note 6, at xiii.

46. BALL, *supra* note 12, at 181.

"preprogrammed, omniscient computers" to make their mathematical models work, but instead takes seriously "the thousand and one parts of our daily lives that cannot be reduced to numbers but that make our lives worth living."[47]

B. Presenting a Rough Formula For Emergence

Economic activity fundamentally is about order creation. We organize our world by transforming energy, matter, and information into the goods and services we want. By cooperating, specializing, and trading, we can create even more order than otherwise we could on our own.[48]

The complex interactions that make up our networked innovation economy are not simple to model, and in turn do not lend themselves to simplistic policymaking. This kind of market operates as an open, dynamic, and nonlinear system. Emergence Economics is our suggested umbrella phrase for a rapidly developing field that incorporates a broad set of tools to understand this type of activity. Eric Beinhocker outlines some of the principles of this approach in his recent book, *The Origin of Wealth*.[49] We have assembled these and other elements into a rough formula that captures the essence of the economic activity:

$$\text{Agents} + \text{Networks} + \text{Evolution} = \text{Emergence}$$

Agents in this case consist of the full spectrum of economic actors—large and small businesses, noncommercial organizations, ordinary consumers, individuals with varying motivations, universities that generate foundational research, government officials, and others. These agents form ad hoc relationships that change over time and interconnect into larger social *networks*. Through this process, individual agents build on others' innovations, and the system overall *evolves* toward greater productivity. When these dynamics arise, the system develops an *emergent* structure, generating spontaneous and nonlinear growth (or in some cases, decay). Emergence, then, is what results from a complex interplay of agents, networks, and evolutionary forces.[50]

This is not meant to suggest a straightforward linear equation. Each element is its own complex adaptive system, which greatly expands the

---

47. *Id.* at 208.
48. BEINHOCKER, *supra* note 22, at 5-20.
49. *Id.*
50. It must be noted here that the rough formula for emergence is a generic calculation that can work equally well for humans acting in other complex adaptive systems, such as political or social roles. For a far more in-depth treatment of the mathematical equations of evolutionary dynamics, see MARTIN NOWAK, EVOLUTIONARY DYNAMICS (2006).

scale and scope of the resulting emergent behavior. Instead of straightforward emergence (if there is such a thing), we have emergence layered on top of emergence, many times over. The "rough formula," thus, is an over-simplified approximation intended to illuminate these properties and should not be misconstrued as a faithful representation of reality. By isolating, we both illustrate and distort.

Importantly, what we exchange in this system are not just finished goods and services—per traditional economic theory—but raw ideas, applied technologies, and new means of productivity. Communications networks—increasingly converging to the Internet—constitute a core physical infrastructure that supports such growth across many sectors of our economy. Innovation and technology are key elements, because they propagate through the network as components of new recipes for economic growth. We will explore more fully in later sections the role of the Internet, and the emergence of ideas and innovation, economic growth, and "Net effects." We will also show how policymakers generally should not attempt to engineer or intrude into these market-based relationships, but still can help keep the system open to productive dynamism. For now, though, we will sketch out the four interrelated components of a "rough formula for emergence."

1. Agents

Any theory of economics must begin with a sound theory of human nature. After all, "[e]conomies are ultimately made up of people."[51] With that overarching premise in mind, we will briefly examine how Old School Economics is built on the flimsy and ultimately unsupportable premise that human beings are perfect economic agents.

In this paper, we will use the word "agent" generically to describe humans acting in their environment. To be an agent is to have several different meanings and connotations: as a self-possessed entity, as acting on behalf of others, and in the chemical sense, providing catalytic change. The term is preferred to either consumer or user, both of which tend to reduce humans to a one-way relationship of purchasing access to goods, services, or other resources.

Agents are economic actors, and individual nodes in a network. Whether acting as consumers or investors, CEOs or government officials, all of us play this interactive role in the economy. The central insight of economics is that agents respond to incentives.[52] Beyond that observation, traditional economic theory assumes that agents have

51. BEINHOCKER, *supra* note 22, at 115; *see also* SHERMER, *supra* note 2, at 190 ("Any theory of economics must begin with a sound theory of human nature.").

52. WILLIAM EASTERLY, THE ELUSIVE QUEST FOR GROWTH 143 (2001); ORMEROD, *supra* note 6, at 63.

definite characteristics. These assumptions include: agents are modeled collectively, use complex deductive calculations to make decisions, have complete information available for free to gather and process, account for all relevant factors, face no transaction costs, have perfect freedom to act, make no errors and have no biases and—being perfect—have no need for learning or adaptation.[53] Under the standard model of human behavior, each of us displays perfect rationality, by pursuing our economic self-interest in carefully calculated ways.[54] Economic actors only interact through market prices. As Leijonhufvud has put it, the usual economic model of human behavior posits incredibly smart people in unbelievably simple worlds.[55]

Each of these assumptions is misplaced. Much well-grounded thinking about agents and what they do comes from the latest teachings of evolutionary psychology, neuroscience, and game theory. In fact, a new form of economics has emerged—behavioral economics—with the actual human being at its core. Behavioral economics seeks to right some of the false assumptions that lie at the heart of Old School Economics. For example, agents do not possess perfect rationality; instead, at best they live with bounded rationality.[56] Imperfect and asymmetric information is the rule, rather than the exception, in most high-stakes competition.[57] Nor are we "homogeneous billiard balls or gas molecules" but creatures with different interests, intentions, and biases, all of which inevitably color whether and how we make economic decisions.[58] These aspects of our behavior stem from the fact that our senses, thoughts, and memory are attuned to the embodied, evolved environment of early *homo sapiens*. Survival and procreation, not "truth," are the governing realities that have shaped us.[59] In summary, then, recent research in the area yields several observations:

- we prefer stories to statistics (relying on anecdotal evidence);

53. BALL, *supra* note 12, at 204-25; BEINHOCKER, *supra* note 22, at 97, 115-39.

54. BEINHOCKER, *supra* note 22, at 51; ORMEROD, *supra* note 6, at 64.

55. Axel Leijonhufvud, *Towards a Not-Too-Rational Macroeconomics*, *in* BEYOND MICROFOUNDATIONS 39, 39-55 (David Colander ed. 1996).

56. Herbert Simon first introduced this concept in the 1950s, but only recently has it begun to influence everyday economic thought. BALL, *supra* note 12, at 211-12; ARIEL RUBINSTEIN, MODELING BOUNDED RATIONALITY 3 (1998). Joseph Stiglitz and George Akerlof, 2001 Nobel Prize winners, have helped further the concept of bounded rationality in economics. As Daniel Kahnmen puts it, "The failure of the rational model is not in its logic but in the human brain it requires." PETER L. BERNSTEIN, AGAINST THE GODS 284 (1996).

57. CASE, *supra* note 28, at 49.

58. TAYLOR, *supra* note 6, at 273.

59. To put it more colloquially, one can apply neuroscience metaphorically to the motion picture A FEW GOOD MEN (Castle Rock Entertainment 1992), where Tom Cruise (as The Brain) asserts "I want the truth," and Jack Nicholson (as The World) responds, "You can't handle the truth."

- we seek to confirm (remembering the hits and forgetting the misses);
- we crave causality (underestimating the role of chance and coincidence in life);
- we misperceive aspects of our world (senses can be deceived);
- we oversimplify (avoiding analysis paralysis);
- we have faulty memories (memory is constructive);
- we hold beliefs based on many external influences (parental, sibling, peer, educational, social, and cultural);
- we have framing biases;
- we rely only on available evidence;
- we utilize linear processing;
- we have difficulty accurately calculating risk and probabilities;
- we can be confused and even paralyzed by having too many options;
- we compartmentalize our economic behavior; and
- we have individually varying skills, perspectives, and intuitions.[60]

In particular, Stanovich notes that while our problem-solving strategies lead us to select regular, deterministic, indication-dependent, functional, and linear processes, the world itself exhibits irregular, indeterminate, and independent processes.[61] We fall for the "decision illusions" our minds show us, because "we are limited to the tools nature has given us, and the natural way in which we make decisions is limited by the quality and accuracy of these tools."[62]

These varying aspects of our behavior obviously directly affect

---

60. The literature literally teems with excellent treatments of all these well-grounded scientific findings. *See, e.g.*, DAN ARIELY, PREDICTABLY IRRATIONAL (2008); PAUL BLOOM, DESCARTES' BABY (2005); MARK S. BLUMBERG, BASIC INSTINCT (2006); GILLES FAUCONNIER & MARK TURNER, THE WAY WE THINK (2003); CORDELIA FINE, A MIND OF ITS OWN (2006); ROBERT FOGELIN, WALKING THE TIGHTROPE OF REASON (2005); CHRIS FRITH, MAKING UP THE MIND (2007); MARC D. HAUSER, MORAL MINDS (2006); THOMAS KIDA, DON'T BELIEVE EVERYTHING YOU THINK (2006); MELVIN KONNOR, THE TANGLED WING (2003); GEORGE LAKOFF & MARK JOHNSON, PHILOSOPHY IN THE FLESH (1999); DAVID J. LINDEN, THE ACCIDENTAL MIND (2007); HUMBERTO R. MATURANA & FRANCISCO J. VARELA, THE TREE OF KNOWLEDGE (1992); READ MONTAGUE, YOUR BRAIN IS (ALMOST) PERFECT (2007); ANDREW B. NEWBERG & MARK ROBERT WALDMAN, WHY WE BELIEVE WHAT WE BELIEVE (2006); TOR NORRETRANDERS, THE USER ILLUSION (1998); DANIEL L. SCHACTER, THE SEVEN SINS OF MEMORY (2002); BARRY SCHWARTZ, THE PARADOX OF CHOICE (2005); KEITH E. STANOVICH, THE ROBOT'S REBELLION (2005); NASSIM NICHOLAS TALEB, FOOLED BY RANDOMNESS (2005); RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE (2008); TIMOTHY D. WILSON, STRANGERS TO OURSELVES (2004).

61. STANOVICH, *supra* note 60, at 63-69.

62. ARIELY, *supra* note 60, at 243.

whether, why, and how we make decisions in the marketplace.[63] As just one example, profit-seeking entities can actively prey on those constraints as part of the process of selling goods and services, which only exacerbates the shaky foundations for an agent's market decisions. A broker in a particular transaction might obscure relevant information in the interest of gaining higher commissions, or the exclusive provider of services might package them in such a way that consumers do not realize that they could get a better deal with alternative combinations. Others have pointed out the political implications for democratic societies as well.[64]

Of course, firms are just collectives of individuals, and often act as if possessed of a single mind. Firms share similar individual characteristics of agents, in particular routinely lacking relevant information and possessing inherent uncertainty.[65] Ormerod states that agents of all types, including firms and governments, "have very limited capacities to acquire knowledge about the true impact either of their strategies on others or of others on them."[66] Agents face massive inherent uncertainty about the effects of their actions.

Some see the constraints inherent in our human information processing systems as signs of significant and inherent weakness—the proverbial glass half empty. However, other research points in the opposite direction: human beings are more capable, multi-faceted, and flexible than heretofore has been recognized. Again, these fundamental characteristics are not reflected in Old School Economics. Among the key findings:

- we have a variety of motivations, including non-economic ones;
- we utilize not just reason but imagination, intuition, and insight as the foundations for creative thinking;
- we often know more than the official producers;
- we can understand and even transcend our constraints;
- we are wired to engage in market-exchange calculations;
- we are altruistic, cooperative, and sharing creatures;

63. *See, e.g.,* PAUL W. GLIMCHER, DECISIONS, UNCERTAINTY, AND THE BRIAN (2003); RICHARD RESTAK, THE NAKED BRAIN (2006); TALEB, *supra* note 60 (humans tend to overestimate causality and underestimate luck). Recent books have also begun to apply lessons from biology and complexity science to the management of large organizations. *See, e.g.*, THE BIOLOGY OF BUSINESS (John Henry Clippinger III, ed., 1999) (collection of essays explaining "the Complex Adaptive System of management"); AXELROD & COHEN, *supra* note 19.

64. BRYAN CAPLAN, THE MYTH OF THE RATIONAL VOTER (2007); DREW WESTEN, THE POLITICAL BRAIN (2007).

65. ORMEROD, *supra* note 6, at 21-35.

66. *Id.* at 221.

- we can use intelligent action to "tip" the world in certain directions;
- we use induction (pattern recognition), as well as deduction (the scientific method); and
- we possess important traits as autonomous entities interacting to carry out particular tasks.[67]

Most importantly, human beings have an inherent ability to learn, to adapt, to change, and to grow. We "evolved the adaptation of adaptability."[68] Our brains have been created with built-in plasticity, so that they are malleable and open to conscious change from new experiences and new learning.[69] We are adaptive agents (or more precisely, agents capable of adaptation) in an ever-evolving landscape. Nor are we the selfish automatons that Old School Economics presupposes.[70] Stanovich further insists that we can use our rational self-determination to gain control over our mismatched genetic and cultural programming,[71] while Donald surmises that we can take advantage of our hybrid brain/cultural mind to break free from our evolutionary heritage.[72] As populations of agents, we can learn from each other, share new ideas and innovations, and serve as a fertile environment for growth.[73] We also have recourse to a vast array of culturally and socially embedded "idea-spaces" that populate our extended minds.[74]

Another point is worth stressing here: the traditional focus on the single individual, standing alone in her perfect wisdom and forethought, ignores growing evidence that large groups of people often can be better at solving problems, reaching decisions, predicting the future, and

67. *See, e.g.*, HOWARD GARDNER, CHANGING MINDS (2006); MALCOLM GLADWELL, THE TIPPING POINT 259 (2002); RICHARD OGLE, SMART WORLD (2007); SHERMER, *supra* note 2; NASSIM NICHOLAS TALEB, THE BLACK SWAN (2007); ERIC VON HIPPEL, DEMOCRATIZING INNOVATION (2005); James Odell, *Agents and Complex Systems*, vol. 1, no. 2 J. OF OBJECT TECH. 35 (2002), *available at* http://www.jot.fm/issues/issue_2002_07/column3. In Benkler's memorable words, "it turns out that we are not intellectual lemmings." YOCHAI BENKLER, THE WEALTH OF NETWORKS 466 (2007).

68. SHERMER, *supra* note 2, at 190 (emphasis in original removed).

69. JEFFREY M. SCHWARTZ & SHARON BEGLEY, THE MIND AND THE BRAIN (2002).

70. In an often-overlooked work, Adam Smith declares: "How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortunes of others, and render their happiness necessary to him, though he derives nothing from it, except the pleasure of seeing it." ADAM SMITH, THE THEORY OF MORAL SENTIMENTS 1 (J.J. TOURNEISEN 1793) (1759).

71. STANOVICH, *supra* note 60, at 95-171.

72. MERLIN DONALD, A MIND SO RARE (2001).

73. AXELROD & COHEN, *supra* note 19, at 5.

74. OGLE, *supra* note 67, at 13-17 (2007).

fostering innovation.[75] Philip Ball puts it well:

> One of the features of collective behavior arising from local interactions is that it becomes impossible to deduce the global state of a system purely by inspecting the characteristics of its individual components. This is physical science's most important message to social science: do not be tempted too readily into extrapolating from the psychology of the individual to the behavior of the group.[76]

While we will return to this concept at a later point, for now the crucial takeaway is that Old School Economics has little to say about collective intelligence operating in the marketplace.

In short, evolutionary psychology, neuroscience, and game theory studies together show that humans are both more limited, and more limitless, than Old School Economics has assumed. We are eminently fallible, yet highly adaptable, agents. Any well-grounded economic theory must take nuanced account of both the half-empty and half-full views of economic agents. Neoclassical economic theories fail this fundamental test.

2. Networks

Of course, constrained yet adaptable agents (normal human beings) do not exist in a vacuum. The full productive potential of agents comes from their interactions with each other, which facilitate sharing of information and effort. Any particular agent may have a link to several other agents, which in turn link to others through lines of communication, common tasks, market agreements, or any number of other relationships.

In Old School Economics, agents only interact indirectly, through static and closed market mechanisms.[77] As a result, many of the connections within the economy are downplayed, or even ignored.[78] Reality is a bit more complex than that. In a dynamic system, relationships are bound to change over time. The true value of an agent is affected, and often greatly enhanced, by links to other agents. It is the structure of the connections between the component parts that gives systems of people their distinctive and characteristic features.[79] When viewed as a whole, human systems show themselves to be complex sets of

75. HOWARD RHEINGOLD, SMART MOBS: THE NEXT SOCIAL REVOLUTION (2003); CLAY SHIRKY, HERE COMES EVERYONE (2008); JAMES SUROWIECKI, THE WISDOM OF CROWDS (2004).
76. BALL, *supra* note 12, at 297-98.
77. BEINHOCKER, *supra* note 22 , at 97, 141-59.
78. ORMEROD, *supra* note 6, at 146.
79. *Id.* at 173.

relationships that can be fully understood neither from the perspective of the individual agents, nor the system as a whole.

The economy is best conceptualized and analyzed as a connected system, a network of individual agents. In the language of network science, the agents are "nodes" and the links are "edges." Thus, networks are interactions of nodes and groups of edges between and among nodes. In particular, the interactions of agents (whether individuals, firms, or governments) include inherent elements of unpredictability and help create complexity in the overall system. Networks both define, and are defined by, these interactions.[80]

The characteristics of networks sometimes can be counterintuitive, but once understood, can be extremely powerful. The field of network science explores how networks form and attempts to explain why certain dynamics arise in networks that do not appear in more static, linear systems.[81] In fact, there is a growing consensus that "common structures, growth patterns, and collective behaviors will arise in networks composed of very different kinds of elements and linkages."[82]

As mentioned previously, physicists and biologists for decades have been studying complex adaptive systems, which are open systems of interacting agents that adapt to each other and the environment. Examples of complex adaptive systems include neurons in the brain, immune systems, biological ecosystems, and the Internet. Economies too are a type of complex adaptive system. Such complex systems may be understood as energy flow structures organized by thermodynamic principles.[83] Economic and social systems are essentially dynamic, and not static.[84] Some have termed it the "econosphere"—the economy as a dynamic, evolving system.[85]

In many systems, individual actors end up having indirect positive effect on others. Economists call these effects "positive externalities", and often discuss the benefits that accrue to others as "spillovers."[86] For

80. "The network is the dominant pattern of the new digital economy." W. Brian Arthur, Myths and Realities of the High-Tech Economy 1 (2000), http://www.santafe.edu/~wbarthur/Papers/Pdf_files/Credit_Suisse_Web.pdf.

81. ALBERT-LASZLO BARABASI, LINKED (2003); STEVEN JOHNSON, EMERGENCE (2002); DUNCAN J. WATTS, SIX DEGREES (2003).

82. Katherine J. Strandburg et al., *Law and the Science of Networks: An Overview and an Application to the "Patent Explosion"*, 21 BERKELEY TECH.L.J. 1293, 1301 (2006).

83. ERIC D. SCHNEIDER & DORION SAGAN, INTO THE COOL 293 (2005).

84. Ormerod, supra note 6, at 18-21, 50-51.

85. Science writer Philip Ball criticizes the practice of using the term "complexity science" to explain aspects of human behavior. He relies instead on the concept of a science of "collective behavior," and sees the market laws emerging from the ordinary (but still unpredictable) push and pull of trade. BALL, *supra* note 12, at 5-6, 179-80.

86. Some prefer to call these externalities "demand side economies of scale." *See* MARK COOPER, FROM WIFI TO WIKIS AND OPEN SOURCE 133 (2006), *available at* http://cyberlaw.stanford.edu/node/5522.

example, I may invent a new method for scanning bar codes that yields me great profit, but you might adopt or adapt this technology to your own benefit (provided that the law allows). Furthermore, to the extent that different agents share this standard—say, a manufacturer using bar codes for inventory management and a retailer using the same codes to automate checkout—the system benefits exceed the sum of the parts. In complex networks, these benefits flow more freely than in disconnected islands. The type of externalities referred to as "network effects" arise only in networks. In this case, each new node added to the network creates added value for the existing nodes. One classic case is the telephone network, in which a globally interconnected system is substantially more valuable to all than a regional or locally delimited system.

A more recent example is the digital network Ethernet standard. The more individuals that owned Ethernet equipment, the more useful the network that connects them together—which eventually helped catalyze the explosion of consumer Internet use. The presence of externalities means that a great deal of what happens in a network, and the value that is created, comes from and flows to other nodes. It also means that the total value created is greater than what each node can create or capture in isolation. In other words, a network becomes more valuable to its users as it grows.[87] We use the term "Net effects" later in this paper to refer to a diversity of presumed externalities that in fact arise internally from the complex network itself.

Network formation theory looks at networks as endogenous constructs that both produce and are produced by a collection of interactions.[88] There are two broad classes of how networks form: random formation, from graph theory (as formulated by Spulber and Yoo),[89] and strategic formation of individual, self-interested agents, from game theory (as formulated by Werbach).[90]

To begin with, networks have a tendency to expand slowly and then exhibit explosive growth as individual networks interconnect. Positive externalities accelerate this activity, because these highly interconnected networks represent considerably more value, and the effects of each new

87. Here is another instance where a basic tenet of Old School Economics—most markets are characterized by declining and eventually negative returns to scale—does not necessarily comport with reality.

88. Kevin Werbach, *The Centripetal Network: How the Internet Holds Itself Together, and the Forces Tearing it Apart*, 42 U.C. DAVIS L. REV. 343, 386 (2008).

89. Daniel F. Spulber & Christopher S. Yoo, *Network Regulation: The Many Faces of Access* 6-7 (Vanderbilt Pub. Law Research Paper No. 05-19; Vanderbilt Law & Econ. Research Paper No. 05-15; Northwestern Law & Econ. Research Paper No. 05-16, 2005), *available at* http://ssrn.com/abstract=740297.

90. Werbach, *supra* note 88, at 21-24.

node feeds back into the system. However, these types of "phase transitions"—abrupt jumps from one state of connectedness to another—can also work in the reverse direction. If those who control particularly central nodes, edges, or clusters see benefit in restricting use of those assets, they can exponentially dampen the growth of the network as a whole.[91]

Another feature of networks is that they can help reduce "transaction costs" of finding and negotiating interactions with partners. This is true both of literal networks, like the Internet, and figurative networks, like social or market relationships. An isolated node would have to generate its own value or negotiate with others to obtain what it needed. Traditionally, the presence of these transaction costs has been used to explain why "firms" are created.[92] By bringing many entities together under a single umbrella, an organization can limit the transaction costs required. In complex networks, these units need not be limited to literal "firms," and the multitude of links can reduce transaction costs in more dynamic fashion.

Complex real world networks exhibit three other kinds of behavior worth noting here. *Small world behavior* states that the diameter of a network (the average number of links between any two nodes) tends to grow much more slowly than the number of nodes.[93] This means that a relatively small number of "hops" is necessary to connect any two nodes in the network. In other words, a small worlds network is relatively tightly connected.[94] *Scale-free dynamics* states that some nodes are vastly more connected than others, so that additional links are more likely to connect to nodes that are already well connected. This behavior explains the so-called "rich get richer effect," where preferential attachment by new users is a real element of networks.[95] Finally, *self-organized criticality* and *critical points* refer to a network's state of precarious stability, where one of several paths is imminently possible.[96] Taken together, these three characteristics provide important insights on how and why complex networks like economies behave the way they do.

91. *Id.* at 28. While such restrictions can appear to make rational sense from the perspective of one agent, another agent with better understanding of the greater dynamics at work likely will find a way to avoid such counter-productive behavior, while also capturing more value than an isolationist approach would yield.

92. R. H. Coase, *The Nature of the Firm*, 4 ECONOMICA 386 (1937).

93. *See, e.g.,* MARK BUCHANAN, NEXUS (2002); DUNCAN J. WATTS, SMALL WORLDS (1999).

94. Strandburg et al., *supra* note 82, at 1305.

95. *Id.* at 1308-09.

96. BALL, *supra* note 12, at 227-41.

3. Evolution

Old School Economics has no explicit mechanisms for explaining endogenous novelty (within the system), agents who learn and adapt, or sudden growth in complexity.[97] By contrast, Emergence Economics uses the universal algorithm of evolution as the basis for much of its analysis.

"Ultimately, economics is a biological science. It is the study of how humans choose. That choice is inescapably a biological process."[98] Ilya Prigogine explains that "[w]e live in an evolutionary universe . . . [where] the laws of nature . . . no longer deal with certitudes but possibilities . . . [and] irregular, chaotic motions . . . constitute[] the very foundation of macroscopic systems."[99] The economy is one such macroscopic system, and, as we have seen, specifically a complex adaptive system. As such, evolution becomes the ideal algorithm for creating value within that system, an iterative process of experimentation by agents that includes first differentiation, then selection, and finally amplification of things that work. To Schumpeter, "[t]he essential point to grasp is that in dealing with capitalism we are dealing with an evolutionary process."[100] Hayek, another transitional figure in 20th Century economics, also gained the insight that markets involve "the evolutionary formation of such highly complex self-maintaining orders."[101]

a. *The Three Stage Process*

Evolution is the universal algorithm for change in biological systems, and now has been identified as operating within economic systems as well.[102] Natural selection is simply a description of certain evolutionary processes initiated by agents.[103] In economic systems, one can usefully think of the process of "natural" selection as comprised of three interrelated stages: differentiation, selection, and amplification. The first step of evolution is *differentiation*, in which intelligent agents identify and propose various possible approaches. Next, through observation and action, these agents sort through the variation to find

97. BEINHOCKER, *supra* note 22, at 97, 187-217.
98. Glimcher, *supra* note 63, at 336.
99. ILYA PRIGOGINE, THE END OF CERTAINTY 155 (1997).
100. SCHUMPETER, *supra* note 35, at 82-83.
101. F. A. HAYEK, THE FATAL CONCEIT: THE ERRORS OF SOCIALISM 9 (1988).
102. GEERAT VERMEIJ, NATURE: AN ECONOMIC HISTORY, at 43-58.
103. SHERMER, *supra* note 2, at 41. Corning clarifies that "natural selection does not in fact *do* anything"—it is not a mechanism or causal agency. In reality, "the differential 'selection' of a trait, or an adaptation, is a consequence of the functional effects it produces in relation to the survival and reproductive success of an organism in a given environment. It is these functional effects that are ultimately responsible for the trans-generational continuities and changes in nature." Peter A. Corning, *The Re-Emergence Of "Emergence": A Venerable Concept In Search Of A Theory,* COMPLEXITY, July/Aug. 2002, at 18, 27.

what works and what does not, and *select* the most fit solutions. Finally, the agents share and iterate on the most successful approaches, throwing out the others and *amplifying* the effects.[104] In other words, natural selection both "weeds out" what fails to work and "weeds in" what does.[105]

So adaptation is the formation and continual testing of hypotheses about the environment.[106] This same evolutionary formula lies at the heart of the market process. Agents, acting as "selectors," pick and choose which products, services, and other transactions they want to engage in, and other agents respond accordingly. The different routines that each firm develops are analogous to the genes, or "genotypes," of biological organisms; in turn, these routines influence the specific characteristics of the output (the "phenotypes," or physical organisms themselves) produced by the different firms. These firms then use the infrastructure of the network as the environment to evolve both their practices and the structure of the network itself. Out of this astonishingly complex series of moves, an ordered market system evolves.[107]

Evolution allows for experimentation with a variety of solutions to problems, means of innovation, and shared experience. Many problems we encounter are complex and lack clear ideal paths to a solution. Scientific discovery has long exhibited this hit-or-miss characteristic,[108] and technological breakthroughs similarly can come from unexpected directions. As Daniel Dennett puts it, "evolution is a search algorithm that 'finds needles of good design in haystacks of possibility.'"[109] Evolution discovers design, through trial and error, acting as "The Great Tinkerer." Chance and accident also play a significant role in

---

104. For a more complete overview of the basics of evolution in a networked economy, see BEINHOCKER, *supra* note 22 , at 213-16.

105. TALEB, *supra* note 67, at 17. Other analysts employ a somewhat different schema for the evolutionary process. For example, authors Axelrod and Cohen divide up the evolutionary algorithm into Variation (the raw material of adaptation), Interaction (between agents and populations of agents) and Selection (to promote adaptation). AXELROD & COHEN, *supra* note 19, at 32-151. They explain that "harnessing complexity" refers to changing the structure of a complex system to increase some measure of performance. *Id.* William Wallace talks about technology creating disruptions to the economy that trigger the "FROCA" process (Frontier, Release, Overexploited, Crash, Adaptation). WILLIAM WALLACE, TECHNO-CULTURAL EVOLUTION 7 (2006). By an interesting reverse analysis, Geerat Vermeij shows how processes common to all economic systems—competition, cooperation, adaptation, and feedback—in turn also govern evolution. VERMEIJ, *supra* note 102.

106. VERMEIJ, *supra* note 102, at 55.

107. *See, e.g.*, RICHARD NELSON & SIDNEY WINTER, AN EVOLUTIONARY THEORY OF ECONOMIC CHANGE (1982).

108. *E.g.* THOMAS S. KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS (1970); KARL POPPER, CONJECTURES AND REFUTATIONS (1963).

109. BEINHOCKER, *supra* note 22, at 14 (citing DANIEL DENNETT, DARWIN'S DANGEROUS IDEA (1996)).

evolution,[110] as well as simple luck.[111] If we assume a particular design space in which agents experiment, they can adapt successful designs by continuing to iterate on what proves useful, and eventually converge on one or more "fitness functions."[112] The environment is the design space of evolution; the market—the "econosphere" or "marketspace"—is the design space of economics. People use this design space to purchase, sell, and barter the goods and services best suited to meet their unique needs and desires.[113]

Fitness is the measure of the potential for value creation; it is a contingent concept, premised on the challenges and opportunities of a particular environment. By one account, fitness is simply an entity's capacity to satisfy customer concerns.[114] Viewed functionally, "fitness is measured by the capacity to connect and interrelate effectively and creatively."[115] In networked systems, fitness is an emergent property, arising as an interplay of dynamic elements within the system as a whole.[116] Ogle argues that increasing fitness triggers tipping points by balancing an agent's reach and reciprocity (its weak and strong ties to other entities) within a dynamically linked network of idea-spaces.[117] Ormerod explains that if we increase the fitness threshold at which agents become extinct in the design space, we are making it more difficult for them to survive, and if we reduce it, we are making it easier. "We can readily think of this as corresponding to more and less competitive environments, respectively."[118] Because this "solution space" of fitness is complex and often changing, this is not a linear process. Instead, it is continuous innovation that takes place in parallel and works by building shared knowledge that feeds back into the system.

Although the three-stage formulation of the market's evolutionary process sounds simple, there are several challenges to successfully employing it. One problem is the sheer number of possible formulas. To

110. STEPHEN JAY GOULD, WONDERFUL LIFE 285, 288 (1989) ("Our own evolution is a joy and a wonder because such a curious chain of events would probably never happen again, but having occurred, makes eminent sense . . . . The modern order is largely a product of contingency.").

111. "The reason free markets work is because they allow people to be lucky, thanks to aggressive trial and error, not by giving rewards or 'incentives' for skill. The strategy is, then, to tinker as much as possible . . . ." TALEB, *supra* note 67, at xxi.

112. BEINHOCKER, *supra* note 22, at 195-206.

113. SHERMER, *supra* note 2, at 8.

114. OGLE, *supra* note 67, at 104-05.

115. TAYLOR, THE MOMENT OF COMPLEXITY 197 (2001).

116. OGLE, *supra* note 67, at 111.

117. *Id.* at 109.

118. ORMEROD, *supra* note 6, at 230. Others note that strong selection pressure amplifies the success of the "best" agent while diminishing overall variety in the system, while weaker selection pressure provides more variety but sacrifices some agent fitness. AXELROD & COHEN, *supra* note 19, at 129-30.

try each one can prove to be impractically complex, or the partnerships required can introduce insurmountable transaction costs. To complicate further the process, it can be difficult to discern whether an idea will prove fruitful until several iterations are made or until a complementary approach is developed. In this case, it helps to encourage a plethora of experimentation with minimal barriers to cross-pollination. Participants in the fitness environment inevitably are blind at any given moment to the higher level patterns that are emerging.[119] Finally, successful evolution can only take place when experimenters overcome the social tendency of "path dependence," in which agents simply do things "as they have always been done."

Ormerod's "Iron Law of Failure" further explains that it is failure, not success, which is the distinguishing evolutionary feature of corporate life.[120] Most firms fail. On average, more than 10 percent of all economically active firms in the United States become extinct each year, with roughly the same number of new firms added back to the market.[121] As part of this process, weaker firms are replaced by firms with higher levels of fitness to the existing environment.[122] Traditional economic theory simply ignores this widespread existence of corporate failure.[123] As biological evolution relies on accident—mutation—as the basis for potential change, so do entities in the economic environment often prosper, or fail, due to the exigencies of a particular environment—in other words, fickle fortune.[124]

119. OGLE, *supra* note 67, at 112.

120. ORMEROD, *supra* note 6, at 12.

121. *Id.* at 180; *see also* ATKINSON, *supra* note 44, at 115-16 (the underlying churning of business is a central feature of the New Economy).

122. Taleb claims that the concept of evolutionary fitness is overstated, and that evolution ultimately is a series of flukes, some good, some bad. "The fools, the Casanovas, and the blind risk takers are often the ones who win in the short term." TALEB, *supra* note 67, at 116-17. Some companies survive simply because they were "the lucky ones." Taleb insists we should love free markets because "operators in them can be as incompetent as they wish." *Id.* at 181.

123. ORMEROD, *supra* note 6, at 17-35. Ball agrees that many economic theories of the firm fail to acknowledge that "most firms are ephemeral." BALL, *supra* note 12, at 267. The larger lesson is that, for selection to occur, the system needs "superfecundity"—more designs than the environment can support—which thus creates competition. In biology, there are more potential organisms than any ecosystem can support. The same undoubtedly is true for the market, which helps explain Ormerod's "Iron Law of Failure."

124. Again, Taleb finds luck to be the grand equalizer in a free market, because almost everyone can benefit from it, and it is far more egalitarian than even intelligence. "Randomness reshuffles society's cards, sometimes knocking down the big guy." TALEB, *supra* note 67, at 222. Ormerod observes as one example the success of Microsoft's Windows operating system, which "was far more the result of a series of accidents than of a far-sighted, planned strategy." ORMEROD, *supra* note 6, at 122-24.

b. Two Types of Technology

Writ large, technologies can be thought of as "knowledge of everything—products, processes, and forms of organization—that can create economic value."[125] Evolution operates on two broad types of technologies, which Richard Nelson refers to as "Physical Technologies" ("PTs") and "Social Technologies" ("STs").[126] Physical Technologies are means or recipes for producing objects or ideas; they consist of specifications, instructions, shareable practices, and other ways of transforming materials to serve a goal. These technologies have a modular, building-block character of components plus architecture, and instill order in the physical realm. Social Technologies, on the other hand, are methods and designs for organizing people in service of a goal, and instilling order in the social realm. This might consist of a particular team structure or collaborative relationship.[127] The modern day corporation is seen by some as an enabling technology in its own right and crucial to economic development.[128]

In reality, the two types of technologies evolve in relation to each other, [129] and with concrete business designs (referred to in Beinhocker's work as "Business Plans") that incorporate one or both. A software company might find that one specific software development toolkit makes its work easier, and that small working groups of engineers further improves productivity. Physical Technologies can enable Social Technologies and vice-versa. Each type of technology constitutes an evolution of modular ideas that has the potential to be plugged into other scenarios. As with firms, technologies are subject to their own "law of failure" in the market.[130] The long-term power of these successful

125. RICHARD G. LIPSEY ET AL., ECONOMIC TRANSFORMATIONS 10 (2005). While new technologies cause economic growth by increasing the output that can be produced from a given set of resources, they also enable new products, new processes, and new forms of organization. *Id*.

126. Richard Nelson, *Physical and Social Technologies, and Their Evolution* (LEM Working Paper Series 2003). Others perceive the proper unit of selection in the market as occupations, or "making a living," rather than technology. VERMEIJ, *supra* note 102, at 44.

127. BEINHOCKER, *supra* note 22 , at 241-77.

128. JOHN MICKLETHWAIT & ADRIAN WOOLDRIDGE, THE COMPANY xxi (2003). The company has flourished in modern markets because capital can be pooled for investment, investor risk is spread, transaction costs are reduced, and effective management structures are imposed on large organizations. *Id.; see also* BALL, *supra* note 12, at 250-54.

129. Vermeij observes that "in organisms, technology is part of the body; in people, it is an extension—mechanical, intellectual, and cultural—that we design and that, at least figuratively speaking, takes on a life of its own. In both cases, technology evolves; in organisms it does so largely through natural selection, in humans by engineering and market forces." VERMEIJ, *supra* note 102, at 47. Kurzweil has commented that "technology is the continuation of evolution by other means." TAYLOR, *supra* note 115, at 221.

130. As Romer has remarked, "there are many more dead ends out there than there are useful things to discover." Ronald Bailey, *Post-Scarcity Prophet*, REASON, Dec. 2001,

technologies lies in their capacity to be shared and re-used.

As we have seen, biological ecosystems provide a powerful analogy and insight to the functioning of business networks. Under one model, companies work to connect a large and distributed network of companies to their customers, providing "platforms" that other firms can leverage to increase productivity, enhance stability, and spur innovation.[131] The "keystone" is a pattern of behavior that improves the performance of an ecosystem and, in so doing, improves individual performance.[132] Just as "keystone species" in nature play central roles in their ecosystems, companies such as Wal-Mart, Microsoft, and Li & Fung deploy "keystone strategies," using effective collaboration to actively shape and regulate the workings of their business ecosystems.[133]

c. Losing One's Balance

All of this flies in the face of traditional economic notions of linear progression and natural equilibrium. Old School Economists project a single optimal balance for a particular market, and see growth as a smooth trajectory of improved efficiency and increased output. Our more complex view of the process acknowledges that there are several possible "peaks" of high productivity that operate in different ways, and that it is possible to arrive at those peaks via different "fitness functions." Indeed, just when one peak has reached its maximum utility (say, bamboo-based light bulb filament), an entirely different approach might offer a far better fit (such as tungsten-based light bulb filament).

The notion of fitness implies that combined Physical Technologies and Social Technologies are used by agents to navigate a market landscape of possible growth trajectories— like a map of mountains. In these fitness landscapes,[134] agents combine PTs and STs into a Business Plan ("BP"), according to various strategies. As one approach reaches its limit or a peak, one might say that an equilibrium of sorts has been reached—but only until it is upset inevitably by a different approach making use of a different combination. This leads to a "punctuated equilibrium" that is disrupted by "keystone" technologies.

---

http://www.reason.com/news/show/28243.html.

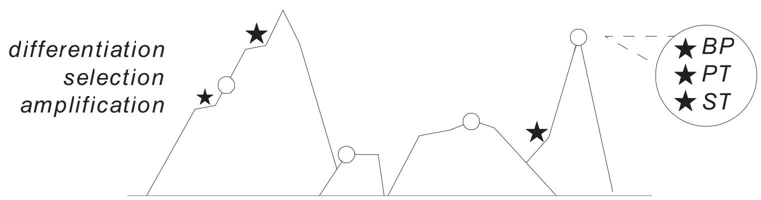
131. *See* MARCO IANSITI & ROY LEVIEN, THE KEYSTONE ADVANTAGE (2004).

132. *Id* at.113-15.

133. *Id.* at 145-67.

134. For further discussion of the creation and development of fitness landscapes in evolutionary biology, see DENNETT, *supra* note 109, at 190-95.

Fig. 1: Fitness Landscapes



Ultimately, no one company can hope to out-innovate the market. An ecosystem tends to beat a product (perhaps even something as innovative as the iPod) because its collective of competitors can explore and innovate and invest in many more ideas than any single company can muster.[135] Beinhocker observes that "in evolutionary systems, *sustainable* competitive advantage does not exist; there is only a never-ending race to create new sources of temporary advantage."[136] The bottom line is, "evolution is cleverer than you are."[137]

4. Emergence of Networks and Growth

Decades of research show that the economic system is a complex adaptive system, where micro interactions of agents lead to macro structures and patterns.[138] In other words, "more is different."[139]

Emergence is not some mystical force that magically comes into being when agents collaborate.[140] Emergent properties are physical aspects of a system not otherwise exhibited by the component parts. They are macro-level features of a system arising from interactions among the system's micro-level components, bringing forth novel

---

135. *See* John J. Sviokla, *In Praise of Ecosystems*, FASTCOMPANY.COM, Aug. 2005, http://www.fastcompany.com/magazine/97/open_essay.html.

136. BEINHOCKER, *supra* note 22, at 332 (emphasis in original).

137. DENNETT, *supra* note 109, at 74 (citing Francis Crick's version of Orgel's Second Rule).

138. *See, e.g.*, TERRY BOSSOMAIER & DAVID G. GREEN, PATTERNS IN THE SAND (1998); MARK BUCHANAN, UBIQUITY (2001); SCOTT CAMAZINE ET AL., SELF-ORGANIZATION IN BIOLOGICAL SYSTEMS (2001); JOHN HOLLAND, EMERGENCE (1998); JOHN HOLLAND, HIDDEN ORDER (1996); ROGER LEWIN, COMPLEXITY (1992); KLAUS MAINZER, THINKING IN COMPLEXITY (4th ed. 2004).

139. *See, e.g.*, BOSSOMAIER & GREEN, *supra* note 138; BUCHANAN, *supra* note 138; CAMAZINE ET AL., *supra* note 138; HOLLAND, EMERGENCE, *supra* note 138; HOLLAND, HIDDEN ORDER, *supra* note 138; STEVEN JOHNSON, EMERGENCE 78 (2001); LEWIN, *supra* note 138; KLAUS MAINZER, *supra* note 138.

140. JOHNSON, *supra* note 139, at 116.

behavior.[141] The brain is an example: the single neuron has no consciousness, but a network of neurons brings forth, say, the smell of a rose. Similarly, when agents interact through networks, they evolve their ways of doing work and discover new techniques. Out of this combined activity, a spontaneous structure emerges. Without any centralized control, emergent properties take shape based on agent relationships and the conditions in the overall environment. Thus, emergence stems from behavior of agents, system structures, and exogenous inputs.[142]

Emergent systems are often described as being "organism-like" in the sense that they are constantly growing and adapting. Each agent follows localized rules and motivations, but the end result is additive and interdependent. Analogies drawn from biology include the ant colony. Ants follow basic rules for seeking food, emitting pheromones to leave a trail to the food they find, and following other ants' pheromone trails to make their way to food and back to the colony. These characteristics appear in many human systems. James Odell notes that, "[w]ith the stock market, thousands of agents act independently to buy and sell shares of particular stocks and bonds. Yet from this independent behavior, an organism-like product called the stock market emerges."[143] Much of the development of cities similarly derives from the bottom up.[144]

Emergent systems have no single ideal structure. They exist in an ever-changing environment and consist of complex interactions that continuously reshape their internal relationships. Brian Arthur notes that our subjective beliefs constitute the very DNA of the market, and so "co-evolve, arise, decay, change, mutually reinforce, and mutually negate."[145] The market "emerges from subjectivity and falls back into subjectivity."[146] The many independent actions of agents unify, but they do not necessarily work toward one particular structure or equilibrium. For example, emergent systems can be robust to change, and they can be far better at evolving toward efficiency than top-down systems. On the other

---

141. Tom De Wolf & Tom Holvoet, *Emergence Versus Self-Organisation* 3464 LECTURE NOTES IN COMPUTER SCI. 1 (2005). Characteristics of emergent systems include micro-macro effects, radial novelty, coherence, interacting parts, dynamical, decentralized control, bi-directional links between the macro- and micro-levels, and robustness and flexibility. *Id*. at 3-5.

142. BEINHOCKER, *supra* note 22 , at 185.

143. Odell, *supra* note 67.

144. Citizens solve local problems, combining resources and expertise in the form of new technologies. Steven Johnson describes how early cities evolved around new farming mechanisms, with urban emergence intensifying as fossil fuel technologies were developed. "And with that new flow of energy, new kinds of cities emerged: the factory towns of Manchester and Leeds, and the great metropolitan superorganisms of London, Paris, and New York." JOHNSON, *supra* note 139, at 113.

145. W. BRIAN ARTHUR, THE END OF CERTAINTY IN ECONOMICS (1994), http://www.santafe.edu/~wbarthur/Papers/Pdf_files/End_of_Certainty_Web.pdf.

146. *Id*. at 6.

hand, emergent structures can fall apart when their basic conditions are altered in such a way that they work against the health of the system as a whole. If the ants stop leaving pheromone trails, they can no longer cooperatively feed the colony. If corrupt stockbrokers are allowed to unethically manipulate the system, the complex structure of price signals falls apart. If cities are saddled with stagnant industries, their growth falters, and their economies can crumble. As our current economic woes illustrate, the line between emergence-fostering actions and emergence-stifling actions sometimes can be difficult to discern.

Agents' actions in turn affect the other agents, setting off both positive and negative feedback loops. Beinhocker uses the helpful metaphor of adjusting shower temperatures.[147] The delay between adjusting the knob and the change in temperature means that one is likely to over-shoot, oscillating back and forth until finally settling on the right temperature.[148] But this is a simple case with a single agent. In a recent study, an economist and a physicist sought to understand what happens in youth hostels where many showers share the scarce "market" for hot water.[149] They found that:

> Tuning one's shower in some hotels may turn into a challenging coordination game with imperfect information. The temperature sensitivity increases with the number of agents, making the problem possibly unlearnable. Because there is in practice a finite number of possible tap positions, identical agents are unlikely to reach even approximately their favorite water temperature.[150]

Fortunately we have developed some understanding of what types of conditions lead away from such negative feedback loops, and towards more productive emergence. Generally speaking, a greater ability of agents to connect and explore new modes of production will facilitate the chance connections that a top-down designer might not foresee. Better global information sharing and feedback between agents facilitates better local decisions. The system as a whole can take a leap forward when new innovations come out of this process and are replicated throughout the network. Inductive tinkering by a single agent can lead to breakthroughs with widespread payoff.[151]

---

147. *See* BEINHOCKER, *supra* note 22, at 101, 394.

148. *Id.*

149. *See* CHRISTINA MATZKE & DAMIEN CHALLET, TAKING A SHOWER IN YOUTH HOSTELS (2008), http://arxiv.org/pdf/0801.1573v1; *Tweaking Taps for a Constantly Warm Shower*, NEW SCIENTIST, Feb. 16, 2008, at 18.

150. MATZKE & CHALLET, *supra* note 149, at 1.

151. In this important sense, the ant colony analogy falls short. Ants are not known to innovate their basic rules for foraging or their colony structure: they do not build new tools for finding food, nor do they have diverse motivations and modes of compensation for their work.

In place of Old School Economics' conventional wisdom of the market's "invisible hand," Beinhocker emphasizes a notion of "fitness functions."[152] Various emergent structures may be more or less fit for the environment and the task at hand. The best chance of finding good fitness functions lies in leaving the emergent system open to subsequent emergence.

As we shall see, emergence can take several different forms, including ideas, innovation, economic growth, and spillovers. Emergent phenomena include economic patterns such as oscillations, punctuated equilibrium, and power laws. Economic growth comes primarily from new ideas; people, ideas (instructions) and things (materials). An evolutionary approach to economics admits that we do not know now, nor will we ever know for sure in the future, the ideal set of market rules. Instead, we should be content to develop supporting institutions that preserve a bounded space for evolutionary activity, and at most look to shape the inputs to the fitness function of the marketplace.

II. NETWORKED ECONOMY: THE INTERNET AS THE ULTIMATE EMERGENT PLATFORM

'Tis true, there's magic in the web of it.
William Shakespeare[153]

Just as economic theory has been turned upside-down thanks to innovative new analytical and empirical work on many fronts, so have the staid assumptions of telecommunications technology been cast aside by the rise and success of the Internet. In many ways, the Internet is the ultimate emergent phenomenon: a platform for untold forms of economic, social, and personal connectivity and interaction. As we have seen, every network of agents operates under a certain set of rules, developed over the course of time in contingent ways. To understand better how the Internet is a novel creation of history—one which can and should play a significant role in shaping our public policy framework—we need to understand what makes the Internet so unique and successful.

For starters, it is important to understand that the "network of networks" we call the Internet is not some neutral, value-free assemblage of routers and servers and fiber optics. Generally, technology may be viewed from a certain perspective as "neutral," but how we design and use it reflects a distinctive social and psychological bias. As an artifact of human ingenuity, technology expresses deep-seated desires, wants, needs, and fears. While component parts may be used for a variety of

152. *See* BEINHOCKER, *supra* note 22, at 195.
153. WILLIAM SHAKESPEARE, OTHELLO, THE MOOR OF VENICE act 3, sc. 4.

purposes—think for example, of the assemblage of mechanical systems into either exploratory rocket ships or atomic weapons—the design and assembly and uses of those components inevitably reflects very human impulses.

In the case of the present-day Internet, that built-in bias is reflected in the key elements of its architecture and infrastructure. As Lawrence Lessig already has shown us, "Code is Law,"[154] or rather, computing technologies are products of human design that affect our behavior.[155] The structure of the Internet reflects the ethos that produced it.[156] Those who struggled to bring forth the Internet did so in the full knowledge that they were imbuing it with specific characteristics that reflected their personal and professional value systems. Those values include interconnectivity, openness, flexibility, and the lack of a pervasive centralized authority.[157] The Net is also oriented towards user activities at the so-called "edge" of the network, as opposed to network activities at the network's "core." At the same time, the Internet has no fixed, inherent nature, except for what we build into its architecture. The Net is what we make it.[158]

A. *The Net's Origins*

1. Overlooked Components: The Social, Economic, and Legal Backdrop

In describing the essential architectural and modular ingredients that make up the Internet, many tend to neglect some of the most critical elements: namely, the social, economic, and legal environment within which the Internet operates. Some have referred to a technology's "context of use," which describes the society and the web of other artifacts within which technologies are always embedded.[159] A technology is not severable from the culture in which it is embedded. "Material artefacts encode, embody, convey, or transmit whole systems of

---

154. LAWRENCE LESSIG, CODE VERSION 2.0 5 (2006).

155. *See id*.

156. BALL, *supra* note 12, at 374. *See also* MANUEL CASTELLI, THE INTERNET GALAXY 36 (2001) ("The culture of the producers of the Internet shaped the medium.").

157. *See* JOHN NAUGHTON, A BRIEF HISTORY OF THE FUTURE 275-77 (2000).

158. Depending on your viewpoint, the Internet at any one moment is a technical architecture (physical assets, logical protocols, and software), or a complex of providers (who owns, operates, and manages the technical components), or a complex of users and their applications and content, or a substrate for economic and non-economic activity, or a process of human interactions. No single conceptual metaphor can hope to capture all of these elements at once.

159. Nelly Oudshoorn & Trevor J. Pinch, *Introduction to* HOW USERS MATTER 1-2 (Nelly Oudshoorn & Trevor J. Pinch eds. 2005).

immaterial ideas and behavioural patterns."[160]

Technology evolves with us, our human capacities, our culture, and our environment. Susan Crawford helpfully has called it the "code/law background medium,"[161] but it actually involves a richer and more complex mix of elements. Indeed, one cannot divorce the Net from its social, economic, and legal context. The ecosystem of the Internet is but a part of the larger ecosystem of human life.

Starting at least in the 17th Century, and extending to today, the "constitutive choices" about the modern media—the press, postal and telecommunications networks, cinema, and broadcasting—have taken place in the context of larger political and economic transformations.[162] In particular, U.S. government policymakers undertook supremely political objectives with important economic consequences. In short, politics created our media world, from the emergence of the first newspapers and postal systems to the rise of the mass press, telecommunications, motion pictures, and broadcasting in the 20th Century. Critical choices about freedom of expression, ownership of media, the architecture of networks, secrecy, privacy, and intellectual property have made the modern media as much a political as a technological invention.[163]

The Internet is no different. Now that the post-industrial, information society has come, what kind of society it proves to be ultimately will be a political choice. The Net is subject to the very same social, economic, and political forces that affect any other part of the world, real or virtual. Because of this rich backdrop, government officials and policymakers potentially have an enormous role in shaping the architecture and uses of the Internet. As we shall see, the U.S. Government in particular can, and inevitably will, to some extent "regulate the Internet."

2. An Unlikely Birth

It has become a truism that the commercial Internet, and particularly the World Wide Web, is a phenomenon built largely by end users operating at the periphery of the network. Nonetheless, surprisingly few bother to stop to ponder exactly what that truism may mean, or what specific implications can be drawn for the future.

Certainly the Internet did not start out that way. After all, despite

160. John Ziman, *Evolutionary Models for Technological Change*, *in* TECHNOLOGICAL INNOVATION AS AN EVOLUTIONARY PROCESS 1, 8 (John Ziman ed. 2000).

161. Susan P. Crawford, *The Biology of the Broadcast Flag*, 25 HASTINGS COMM. & ENT. L.J. 603, 606 (2003).

162. *See generally* PAUL STARR, THE CREATION OF THE MEDIA (2004).

163. *Id.* at 1-19, 385-402.

some of the more extreme rants of self-proclaimed "cyberlibertarians," the Internet is a creature spawned not in the rich soil of the valleys around San Jose, but in windowless conference rooms at the Pentagon, with the aid of government-sponsored academia. In particular, government, military, and academia provided the structure and financial support for the nascent network. As a result, "the Internet was born at the unlikely intersection of big science, military research, and libertarian culture."[164]

Indeed, "the real history of the Internet reaches back to that terribly traditional, often-reviled institution of our collective aspirations: government."[165] The "gift culture of the ARPANET"—the secret scientific research project funded by the U.S. military—became a prolific incubator of many innovations.[166] As one noted historian has concluded, "public investment in science and technology—channeled through institutions that continued, however, to be decentralized and competitive—proved instrumental in the emergence of computer sciences, advanced telecommunications, and other developments that led directly to the contemporary phase of the information revolution."[167] Of course, it is highly ironic that centralized decision-making led to a decentralized Internet, that military desire to create a resilient and efficient system led to a highly interconnected, distributed network, and that the top-down mandate to use a particular root protocol allowed the Net to become a platform for bottom-up user choice and freedom.[168]

The Internet required three decades of subsidies to reach commercial market introduction.[169] It has been estimated that the U.S. Government spent some $125 million building the Internet's predecessor networks.[170] Only government, it seems, can afford to be that patient.[171]

Networking pioneer and entrepreneur Charles Ferguson has observed that new technologies like the Internet typically come from neither the venture capital industry nor from the free market.[172] Instead,

---

164. MANUEL CASTELLS, THE INTERNET GALAXY 17 (2001).

165. DAVID BOLLIER, SILENT THEFT 101 (2003).

166. *Id*. at 103.

167. STARR, *supra* note 162, at 18.

168. BALL, *supra* note 12, at 377-79.

169. For more on this often-misunderstood history, see JANET ABBATE, INVENTING THE INTERNET (1999); KATIE HAFNER & MATTHEW LYON, WHERE WIZARDS STAY UP LATE: THE ORIGINS OF THE INTERNET (2000); NAT'L RESEARCH COUNCIL, THE INTERNET'S COMING OF AGE (2001).

170. Larry Press, *Seeding Networks: The Federal Role*, 39 COMM. OF THE ACM 10, 15 (1996).

171. Lee W. McKnight, *Internet Business Models: Creative Destruction As Usual*, *in* CREATIVE DESTRUCTION 39, 59 (Lee W. McKnight, Paul M. Vaaler, & Raul L. Katz eds. 2002).

172. *See* CHARLES H. FERGUSON, HIGH STAKES, NO PRISONERS: A WINNER'S TALE OF GREED AND GLORY IN THE INTERNET WARS (1999).

he explains that "virtually all the critical technologies in the Internet and Web revolution were developed between 1967 and 1993 by government research agencies and/or in universities."[173] During that same time period, a $10 billion commercial online services industry arose in the free market. "The comparison between the two," he argues,

> is extremely clear and extremely unflattering to private markets. The commercial industry's technology and structure were inferior to that of the nonprofit Internet in every conceivable way, which is the primary reason that they were so rapidly destroyed by the commercial Internet revolution. Internet technology was around and available for more than twenty years, continuously evolving under the noses of companies like AT&T, IBM, CompuServe, AOL, and even Microsoft. But somehow these companies managed not to notice. Neither, by the way, did most VCs.[174]

It certainly is unclear whether the free market alone could or would have created such a thing as the Internet, but the available evidence is not promising. As Ferguson points out, in the 1980s and early 1990s, wholly incompatible, proprietary computer networks arose—bulletin boards, online service providers, private networks, email services. Without the existence of a ready alternative like the Internet, such "closed" networks may well have become the prevailing marketplace norm. Kevin Werbach has noted that "the victory of the interconnected outcomes over the centralized ones was always contingent on historical, regulatory, economic, and cultural factors."[175] The Internet may be viewed as an example of a path-dependent creation, a "telecommunications anomaly,"[176] and even a historic accident.[177] Some may see the Internet as the "Black Swan" of the communications world, a wholly unexpected event that came out of nowhere to bring a profound and widespread impact to the economy.[178]

173. *Id*. at 13. The emergence of the home computer out of the "Homebrew Computer Club," an eclectic San Francisco-based hobbyist group, rather than IBM, HP, or Xerox, provides an interesting parallel to the Net's rise over proprietary alternatives. As Richard Ogle puts it, "[t]he failure of the mainstream computer industry to anticipate the arrival of the personal computer—an enormous failure of insight and imagination—exemplifies once again the fact that being in thrall to the wrong idea-space can blind you to what seems obvious to others." OGLE, *supra* note 67, at 78.

174. FERGUSON, *supra* note 172, at 13. Some have noted that the Internet did not so much drive out its competitors as subsume them. *See* Mark A. Lemley & David McGowan, *Legal Implications of Network Economic Effects*, 86 CAL. L. REV. 479, 552 (1998).

175. Werbach, *supra* note 88 at 18.

176. *See* Paul A. David, *Economic Policy Analysis and the Internet: Coming to Terms with a Telecommunications Anomaly*, *in* OXFORD HANDBOOK ON INFORMATION AND COMMUNICATION TECHNOLOGIES (Robin Mansell et al. eds., 2007).

177. DIANE COYLE, THE SOULFUL SCIENCE 57 (2007).

178. Taleb calls the Net "unplanned, unpredicted, and unappreciated upon its discovery,

The U.S. Government's role certainly was not limited to funding, research, prodding, and eventual privatization. On the regulatory front, policymakers made key decisions that dictated whether and how the Internet would develop into a mass-market phenomenon.[179] Beginning in the late 1960s with the original *Computer Inquiries*, the Federal Communications Commission (FCC) explored ways to protect the nascent online environment from regulation, and give it access to vital communications links. The FCC's *Computer Inquiry* safeguards governed consumer access to last-mile ramps—ordinary phone lines—owned and controlled by the incumbent local exchange carriers (ILECs), and used to access online services. This regulatory framework essentially buttressed the Internet's own open and end-to-end design principles.[180]

The *Computer Inquiry* rules did several important things. First, the world was divided into basic communications services (regulated as common carriage),[181] and enhanced information services (left unregulated). Enhanced services were defined as computer-based software applications and services that utilized the public switched telephone network (PSTN).[182] Second, providers of enhanced services (known as ESPs) gained the right to access basic services, on a nondiscriminatory basis, using the ILECs' commercial rates and terms. This end user right eventually became known as ISP open access. Third, ESPs and others had the concomitant right to attach lawful devices, such as computer modems, to the ILECs' phone networks. The end result was a modular regulatory framework, with targeted common carriage regulation of the lower infrastructure layers of the network, and an "unregulation" regime applicable to the upper applications, devices, and content layers.[183]

---

and . . . well after." TALEB, *supra* note 67, at 135. One question is whether this supposed Black Swan event is still endogenous to (arising from within) the market, or whether the U.S. Government's extensive involvement in the Net's birth and success makes it an exogenous happenstance. The answer to that question well could dictate how one approaches the Internet as an economic phenomenon.

179. For a more fulsome discussion of this regulatory history, see Richard S. Whitt, *A Horizontal Leap Forward*, 56 FED. COMM. L.J. 587, 597-600 (2004).

180. *See* Robert Cannon, *The Legacy of the Federal Communications Commission's Computer Inquiries*, 55 FED COMM. L.J. 167, 204-05 (2003).

181. Common carriage conveys a raft of legacy regulations, including market entry and exit requirements, tariffing of service offerings, cost-based pricing, consumer complaint processes, and general oversight by federal and state regulators. *See generally* 47 U.S.C. §§ 201-231 (2000).

182. The FCC's rules define enhanced services as those services "offered over common carrier transmission facilities used in interstate communications, which employ computer processing applications that act on the format, content, 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." 47 C.F.R. § 64.702(a) (2008).

183. Kevin Werbach observes that the FCC's decision "meant that data services, which

The FCC also took other important steps, such as classifying ESPs as end users, thus protecting them from the excessive per-minute telephony access charges normally applicable to carriers for long-distance telephone traffic that originates and terminates to the ILEC networks.[184] In addition, following the breakup of AT&T in 1984, U.S. District Court Judge Harold Greene presided over a consent decree that barred the Bell Operating Companies from providing interLATA information services until 1991.[185] Finally, in the Telecommunications Act of 1996, Congress retained the FCC's basic/enhanced split in the form of new definitions of "telecommunications services" and "information services,"[186] and added a statutory provision decreeing that the Internet should remain unfettered by regulation.[187] Thus, in the *Computer Inquiry* era spanning roughly from 1980 to 2005, the United States had an open and unregulated communications platform by design (the Internet) that was married by regulation to open end points (the local telephone network).

B. *The Net's Architecture*

The Internet today is a network of networks, an organic hodgepodge of disparate infrastructure melded together through common software protocols. Understanding the what, where, why, and how of this architecture goes a long ways towards understanding how the Net fits into the rough formula of emergence we discussed above, and in turn the implications for communications policy going forward.

1. The Law of Code: Modularity

The modular nature of the Internet describes the "what," or its overall structural architecture. The use of layering means that functional tasks are divided up and assigned to different software-based protocol layers. For example, the "physical" layers of the network govern how electrical signals are carried over physical wiring; independently, the "transport" layers deal with how data packets are routed to their correct destinations, while the application layers control how those packets are used by an email program, web browser, or other user application or service. This simple and flexible system creates a network of modular

---

could ride transparently on top of the voice telephone network, were effectively outside of that network's sphere of influence." Kevin Werbach, *Only Connect*, 22 BERKELEY TECH. L.J. 1233, 1259 (2007).

184. *See* 47 C.F.R § 69.01 (2008).

185. United States v. Am. Tel. & Tel. Co. 552 F. Supp. 131, 197 (D.D.C. 1982), *aff'd sub nom.* Maryland v. United States, 460 U.S. 1001 (1983).

186. 47 U.S.C. §§ 153(20), (46) (2000).

187. 47 U.S.C. § 230 (2000).

"building blocks," where applications or protocols at higher layers can be developed or modified with no impact on lower layers, while lower layers can adopt new transmission and switching technologies without requiring changes to upper layers. Reliance on a modular system of layers greatly facilitates the unimpeded delivery of packets from one point to another. [188]

Put simply, the Internet is comprised of Code, stacked in Layers. One can view Code, the software and hardware components of the network, as the bricks and mortar.[189] Writ large, these components constitute "[a] set of procedures, actions, and practices, designed in particular ways to achieve particular ends in particular contexts."[190] By contrast, layers constitute the architectural features of the Internet, in this case its modular structure. The layers are what we build using the raw materials of Code as the building blocks:

> [E]ngineers use multiple protocols that partition a communication problem into disparate sub-problems and organize the software into modules that handle the sub-problems. Functions are allocated to different protocol layers or levels, with standardized interfaces between layers. The flexibility offered through the layering approach allows products and services to evolve by accommodating changes made at the appropriate layer, rather than having to rework the entire set of protocols. In other words, layering allows changes to implementation of one layer without affecting others, as long as the interfaces between the layers remain constant.[191]

Layers create a degree of "modularity," which allows for ease of maintenance within the network. This modularity, or independence, of each layer creates a useful level of abstraction as one moves through the layered stack. In particular, the user's ability to alter functionality at a certain layer without affecting the rest of the network can yield tremendous efficiencies when one seeks to upgrade an existing application (higher layer) that makes extensive use of underlying physical infrastructure (lower layer).[192]

2. Smart Edges: End-to-End

The end-to-end ("e2e") design principle describes the "where," or the place for network functions to reside in the layered protocol stack. The general proposition is that the core of the Internet (the network

188. *See generally* Whitt, *supra* note 179, at 601-09.
189. ALBERT-LÁSZLÓ BARABÁSI, LINKED 174 (2003).
190. ALEXANDER R. GALLOWAY, PROTOCOL xii (2004).
191. Whitt, *supra* note 179, at 602-03.
192. *Id.* at 604.

itself) tends to support the edge of the Internet (the end user applications, content, and other activities).[193] Some have rendered this broadly as dumb networks supporting smart applications.[194] A more precise technical translation is that a class of functions generally can be more completely and correctly implemented by the applications at each end of a network communication.

The e2e principle suggests that "[s]pecific application-level functions usually cannot, and preferably should not, be built into the lower levels of the system—the core of the network." [195] Instead, such functionality ideally operates on the edges, at the level of client applications that individuals set up and manipulate.[196] E2e architecture "[i]s designed to be fairly simple, open and stable at the network level while allowing users the freedom to develop innovative applications to run on top of it."[197] Thus, users remain the driving force in such a system.[198] Rather than relying upon the creativity of a small group of innovators who might work for the companies that control the network, the e2e design enables anyone with a network connection to design and implement a better way to use that network.[199] As Lee McKnight has observed, "most Internet businesses operate on the edge of the Internet, which is where the intelligence and processing power resides by design."[200] The resulting explosion of innovative applications on the Internet likely never would have happened but for the incorporation of the end-to-end design into the network.[201] Thus, innovation and creativity become decentralized. This differs from traditional telephony and cable networks, where control over security, protocols, or permitted applications and content are handled in the core (in headends and central offices), away from the users at the edge. As a result, the power and

193. *Id.* at 604-05.

194. *See* David S. Isenberg, *The Dawn of the Stupid Network,* ACM NETWORKER, Feb.-Mar. 1998, at 24-31, *available at* http://www.isen.com/papers/Dawnstupid.html.

195. David D. Clark & Marjory S. Blumenthal, *Rethinking the Design of the Internet: The End-to-End Arguments vs. the Brave New World*, 1 ACM TRANSACTIONS ON INTERNET TECH. 70, 71 (2001).

196. Susan P. Crawford, *Someone to Watch Over Me: Social Policies for the Internet* 21 (Cardozo Law School Legal Studies Research Paper No. 129, 2006).

197. BOLLIER*, supra* note 165, at 102.

198. *Id*.

199. Ashish Shah, Douglas C. Sicker, & Dale N. Hatfield, Thinking About Openness in the Telecommunications Policy Context 6 (2003), http://web.si.umich.edu/tprc/papers/2003/244/openness2.pdf (paper presented at the Telecommunications Policy Research Conference).

200. McKnight*, supra* note 171, at 47.

201. *See, e.g.*, Mark A. Lemley & Lawrence Lessig, *The End of End-to-End: Preserving the Architecture of the Internet in the Broadband Era* 14 (Stanford Law School, Working Paper No. 207, 2000) (explaining role of "e2e" design in producing the "extraordinary innovation" of the Internet), *available at* http://cyberlaw.stanford.edu/e2e/papers/Lemley_Lessig_e2epaper.pdf.

functionality of the Internet is left in the hands of the end users.[202]

With regard to the Internet, the end-to-end argument now has been transformed into a broader principle "[t]o make the basic Internet protocols simple, general, and open, leaving the power and functionality in the hands of the application."[203] In the words of one commentator, e2e has become "a policy preference of potentially profound meaning."[204] Of course, the e2e principle can be prone to exaggeration. One cannot have a modern data network without a core, and in particular the transport functionality to connect together the myriad constituents of the edge, as well as the widespread distribution of the applications and content and services provided by the edge. Elements of the core network, while erecting certain barriers (such as firewalls and traffic shaping) that limit pure e2e functionality, may still allow relatively unfettered user-to-user connectivity at the applications and content layers. To have a fully functioning network, the edge and the core need each other. And they need to be connected together.

3. A Network of Networks: Interconnection

Werbach has recently pointed out an often under-appreciated aspect of the Internet's architecture: connectivity.[205] This aspect of the Net goes to its "why," which is the overarching rationale of moving traffic from Point A to Point B. Werbach believes that "the actual development of the Internet focused not on the edges, but on the links."[206] The early Internet was designed with an emphasis on internetworking and interconnectivity, and moving packets of data transparently across a network of networks:

> The defining characteristic of the Internet is not the absence of discrimination, but a relentless commitment to interconnection. . . . The engineers and entrepreneurs who laid the foundations for today's commercial Internet developed a set of technical protocols, business norms, and contractual arrangements to link together diverse

202. By precluding discrimination, e2e also "[s]ets the conditions necessary for a fair fight, so that what survives is truly the fittest and not merely the favored." Tim Wu, *The Broadband Debate*, *A User's Guide*, 3 J. ON TELECOMM. & HIGH TECH. L., 69, 83 (2004). "E2e can help erase through competition the invariable mistakes that a centralized network planner will make." *Id*. at 83-84.

203. Timothy Wu, *Application-Centered Internet Analysis*, 85 VA. L. REV. 1163, 1164-65 (1999).

204. Jonathan Sallet, *Just How Open Must an Open Network be for an Open Network to be Labeled Open*, 8 FIRST MONDAY 1, 6 (2003), http://www.firstmonday.org/issues/issue8_3/sallet/.

205. *See* Werbach, *supra* note 88, at 3.

206. Werbach, *supra* note 183, at 507.

networks.[207]

Interconnecting then is the baseline goal embedded in the Internet's architecture, creating incentives and opportunities for isolated systems to come together, and for edges to become embedded in tightly interconnected networks.[208] Werbach has shown that interconnectivity creates both decentralizing and centralizing trends in the Internet economy, and both centripetal force (pulling networks and systems into the Internet commons) and centrifugal force (towards the creation of isolated gated communities). He expresses concern that the Net increasingly is being pushed towards disaggregated, proprietary islands of connectivity.[209]

4. Agnostic Protocols: IP

The design of the Internet Protocol ("IP"), or the "how," allows for the separation of the networks from the services that ride on top of them. IP was designed to be an open standard, so that anyone could use it to create new applications and new networks. By nature, IP is completely indifferent to both the underlying physical networks, and to the countless applications and devices using those networks. In particular, IP does not care what underlying transport is used (such as fiber, copper, cable, or radio waves), what application it is carrying (such as browsers, e-mail, Instant Messaging, or MP3 packets), or what content it is carrying (text, speech, music, pictures, or video).[210] Thus, IP enables any and all user applications and content. IP also was designed to follow the e2e principle.[211] Thus, using IP, individuals are free to create new and innovative applications that they know will work on the network in predictable ways.

In 1974, Vint Cerf and Robert Kahn issued their seminal paper on the TCP/IP protocol suite, in which the authors "present a protocol design and philosophy that supports the sharing of resources that exist in different packet switching networks."[212] Based in large part on how Cerf and Kahn designed that protocol suite (plus more than a little help from

207. Werbach, *supra* note 88 at 3. Some of the flavor of this scheme can be found in Postel's Law, named after John Postel: "be conservative in what you do, be liberal in what you accept from others." Werbach, *supra* note 183, at 518.

208. Werbach, *supra* note 183, at 529.

209. Werbach, *supra* note 88, at 8.

210. Whitt, *supra* note 179, at 604-07.

211. Timothy Wu, *Network Neutrality, Broadband Discrimination*, 2 J. ON TELECOMM. & HIGH TECH. L. 141, 146 (2003).

212. Vinton G. Cerf & Robert E. Kahn, *A Protocol for Packet Network Intercommunication*, 22 IEEE TRANSACTIONS. ON COMM. No. 5 637 (1974), *available at* http://cs.mills.edu/180/reading/CK74.pdf.

the U.S. Government to ensure its universal use on the networks), the Internet Protocol has become the ubiquitous "bearer" protocol at the heart of the Internet.[213]

C. The End Result: A "Virtuous Feedback Network"

From these various architectural components of the Internet, the end result is that IP helps fashion a "virtuous hourglass" from disparate activities at the different network layers. In other words, the Net drives convergence at the IP (middle) layer, while at the same time facilitating divergence at the physical networks (lower) and applications/content (upper) layers. The interconnected nature of the network allows innovations to build upon each other in self-feeding loops. This network topology and universal connectivity gives meaning to what some have labeled the Net's three golden rules: nobody owns it, everybody uses it, and anyone can add to it.[214] One might refer to this as a "virtuous feedback network."

From the above discussion of the Internet's different yet related design components, one can see the resulting whole: that, generally speaking, no central gatekeeper exerts unilateral control over activities on the Internet.[215] This governing principle allows for vibrant user activity and creativity to occur at the network edges. Moreover, the values imbued into the Net's architecture were there from the beginning.[216] In such an environment, entrepreneurs need not worry about getting permission for their inventions to reach end users. In essence, the Internet has become a distributed, yet connected, platform for emergence.[217] Indeed, technology platforms such as the Internet are both open (accessible) and communal (adaptable). One could think of it like the electric grid, where the ready availability of an open, standardized, and stable source of electricity allows anyone to build and use myriad of

213. *See* Whitt, *supra* note 179, at 629.

214. DON TAPSCOTT & ANTHONY D. WILLIAMS, WIKINOMICS 273 (2006).

215. Of course, as alluded to earlier, talk about an "edge" versus "core" dichotomy should not obscure the messy reality that the Net includes crucial top-down core elements. ICANN runs the domain name system (DNS) so that there is common addressing scheme for users, while the IETF makes critical engineering and standards decisions that then are implemented throughout the system. One irony is that a certain centrality of common standards and addressing schemes may well be necessary in order for the Net's decentralized nature to fully emerge.

216. NAUGHTON, *supra* note 157, at 268-77.

217. These aspects may help explain why the Internet is different from previous forms of communications and transportation infrastructure. *See, e.g.*, VARIAN, ET AL., *supra* note 5, at 4 (while the Internet represents one instance of "combinatorial innovation," its uniqueness may stem from its foundational software protocols, which are quite different from the physical devices that drove previous technologies).

different electric devices.[218]

The Internet is more than that: it is a complex adaptive system, whose architecture is much richer than the sum of its parts.[219] As such, Net-based human activities produce emergent and self-organizing phenomena. Metaphors seem to fall short when describing the Internet; it is by various accounts an object and a process, a place and an idea.

As the networks and users that comprise it continue to change and evolve, the Net's core principles of modularity, e2e, interconnectivity, and agnosticism are constantly being pushed and prodded by technology, market, and legal developments. That is not to say these developments are inherently unhealthy. Clearly there are salient exceptions to every rule, if not new rules altogether, and the Internet needs to adjust to the realities of security concerns like denial-of-service (DoS) attacks, and the needs of latency-sensitive applications like streaming video. The question is not whether the Net will evolve, but how. Will the inevitable changes come organically, or will they be imposed unilaterally? And by whom?

III. GROWTH ECONOMY: THE EMERGENCE OF IDEAS, INNOVATION, AND "NET EFFECTS"

The emergent phenomena of new ideas and innovation, channeled through generative networks of agents such as the Internet, provide powerful fuel for economic growth and other important effects. Growth long has been a concern for economists as they seek to understand what drives nations to build and maintain wealth. In a highly networked economy, the benefits of innovation in physical and social technologies go beyond traditional economic growth, and generate a diversity of what we call "Net effects."

An initial point is to understand that the Internet as a platform for new ideas and innovations has been slighted in Old School Economics as a mere "exogenous" influence. In fact, general platform technologies like the Internet are endogenous elements, which in turn fuel growth within the system. Beinhocker puts it succinctly:

> [A] change in technology, such as the invention of the Internet, can be seen as an exogenous shock to the economic system . . . . The problem with this approach is that it gives economists an escape hatch and allows them to put the most difficult and often most interesting questions outside the bounds of economics. For example,

218. Richard Lanham finds that the Net reflects "the comedy of the commons," as it is developing into an ever-richer community resource that "combines the power of a free market, where individual gain leads to collective benefit, with the cooperative ownership of the cultural conversation." RICHARD A. LANHAM, THE ECONOMICS OF ATTENTION 13 (2006).

219. BARABÁSI, *supra* note 189, at 174.

> if technological change is treated as a random, outside force (like the weather), then one doesn't need a fundamental theory of the interaction between technological change and changes in the economy.[220]

Yochai Benkler also notes that "our theories of growth and innovation assume that industrial models of innovation are dominant."[221] Economics for too long has focused only on production, labor, and capital as the key elements of the market. To these, Romer, who helped found the New Growth school of economics, now has added knowledge and technology.

The "Net effects" we discuss below are a variety of pecuniary and non-pecuniary benefits that emerge when networked agents interact. Economists often treat these effects as "externalities"—meaning that the forces cannot be accounted for purely in terms of traditional market transactions. This includes "spillovers" (non-affiliated entities benefit from others' innovations), peer-production (networks allow diversely motivated agents to collaborate), and all social, political, and cultural benefits outside the purview of standard market analysis. It is tempting to think of these Net effects as consisting of "primary" benefits (economic growth) versus "secondary" benefits (miscellaneous "economic" and "non-economic" advances). Terms like "spillovers" suggest as much, connoting an unintended minor consequence of a major economic activity. Yet we should be hesitant to impose such a dichotomy on this complex mesh of human activities. Not only do "Net effects" help fuel core growth, they can have profound positive impacts on human life.

Countless things emerge from a networked, layered, end-to-end platform like the Internet. For purposes of this paper, the next two sections will delve into those emergent phenomena that have a direct bearing on the public policy landscape. In brief, ideas and innovation emerge from the Net, which in turn brings economic growth and various "Net effects."

A. *The Nature of Ideas and Innovation*

So where do ideas, and then innovation, come from, and why? Ideas have a diverse and unpredictable variety of sources and uses. As we will see, ideas can be wedded to things, and to other ideas, in ways that drive innovation, and in turn create a host of positive economic and non-economic benefits. As Douglass North puts it, ideas and their creation are "the fundamental driving force of the human condition."[222]

---

220. BEINHOCKER, *supra* note 22, at 55.
221. BENKLER, *supra* note 67, at 460.
222. DOUGLASS C. NORTH, UNDERSTANDING THE PROCESS OF ECONOMIC

Ideas are the raw material for innovation. Crawford observes that "ideas are not like goods; they are potentially far more valuable."[223] In the ordinary transformational cycle, ideas become concepts, which become inventions, which are utilized for commercial or other purposes. They are the recipes for combining atoms into useful things; while the atoms are limited, the ideas themselves essentially are unlimited. Innovation, by contrast, is the application of ideas—invention plus implementation. Ideas and innovation form an essential feedback cycle, where input becomes output, becomes input again.[224]

Hayek claimed that "there is no simple under-standing of what makes it necessary for people under certain conditions to believe certain things. The evolution of ideas has its own laws and depends very largely on developments that we cannot predict."[225] But we can still try.

1. Ideas

One reason that economic growth defies simple explanation is that ideas beget future ideas, amplifying total output. It is difficult to know at any given time how much a particular idea will produce, as its cascading effects have yet to be realized. Any one innovation is likely to build on another. This "standing on the shoulders of giants" concept is the familiar motivation for much of our intellectual property law, which seeks to balance incentives for one innovator to produce with the benefits to innovators down the road. For years, economists had given short shrift to this reality in their models of economic growth because it was deemed simply too complex. Technological progress, and the ideas that led to it, were considered "exogenous"—outside of the system.

However, Kenneth Arrow noted in 1962 that "[i]nformation is not only the product of inventive activity, it is also an input."[226] This simple observation articulates in the language of economics something that seems almost intuitive today. Still, seeing ideas as inputs is critically important and fundamentally different from the results in simple linear economic models. Arrow described innovation as an inherently uncertain process, and discussed ways in which a society might spur innovation despite this risky environment. He ultimately concluded that there was

---

CHANGE 18 (2005).

223. Susan P. Crawford, *The Internet and the Project of Communications Law,* 55 UCLA L. REV. 359, 391 (2007).

224. Others draw lines in the technological change process between invention, innovation, and diffusion. LIPSEY ET AL., *supra* note 125, at 12.

225. Thomas W. Hazlett, *The Road from Serfdom: Forseeing the Fall*, REASON, July 1992, http://www.reason.com/news/show/33304.html.

226. Kenneth J. Arrow, *Economic Welfare and the Allocation of Resources for Invention* 13 (Rand Corp. Paper No. P-1856-RC, 1959), *available at* https://www.rand.org/pubs/papers/2006/P1856.pdf.

no clear, single, optimal path, but that the best approach is to foster a diversity of modes of production.[227] In any event, an economy that provides fertile ground for idea creation, reuse, and adaptation tends to spur growth and future innovation more readily than one that does not.

In Beinhocker's telling, ideas that make up Physical Technologies or Social Technologies are particularly valuable in the process of evolution toward more productive systems. Because they feed back into the economy in the form of recipes for production, and because they can be adapted in new and unexpected ways, ideas generate increasing returns. Further, Romer found that knowledge builds on itself, "which means that as we learn more, we get better and better at discovering new things. It also means there's no limit to the amount of things we can discover."[228]

A second overlooked aspect of ideas is that they can be re-used infinitely. In 1990, economist Paul Romer published a landmark paper entitled "Endogenous Technological Change."[229] Building upon Arrow's description of the self-feeding nature of information, Romer further examined the nature of ideas. His key observation was that ideas are non-rival, meaning that any number of persons can simultaneously make use of them.[230] Whereas two people cannot both eat the same apple, for example, ideas can be copied and shared without depriving anyone of their use. However, ideas also are partially excludable, meaning that through law and other constructs we can sometimes prevent this sharing from occurring. Nevertheless, as the cost of transmitting ideas approaches zero, the marginal cost approaches zero as well.[231] From the perspective of social welfare, these ideas would be shared for free. Since information is a non-rival good, it takes only one person to invent an idea, which an entire group then can adapt.[232]

On the Internet, ideas and the resulting innovation could not

---

227. *Id.*

228. Kurtzman, *supra* note 33, at 2.

229. *See* Paul M. Romer, *Endogenous Technological Change*, 98 J. OF POL. ECON. S71 (1990).

230. Of course Romer was not the first to have this insight. Thomas Jefferson noted that

> If nature has made any one thing less susceptible than all others of exclusive property, it is the action of the thinking power called an idea, which an individual may exclusively possess as long as he keeps it to himself; but the moment it is divulged, it forces itself into the possession of every one, and the receiver cannot dispossess himself of it.

Letter from Thomas Jefferson to Isaac McPherson (Aug. 13, 1813), *in* THE WRITINGS OF THOMAS JEFFERSON 333 (Albert E. Bergh ed. 1905) (1907). Albert Einstein similarly remarked, "[i]f I give you a pfennig [penny], you will be one pfennig richer and I'll be one pfennig poorer. But if I give you an idea, you will have a new idea, but I shall still have it, too." The MacTutor History of Mathematics archive, A meeting with Einstein, Mar. 1996, http://www-groups.dcs.st-and.ac.uk/~history/Extras/Einstein.html.

231. DAVID WARSH, KNOWLEDGE AND THE WEALTH OF NATIONS 366 (2006).

232. ROBERT WRIGHT, NONZERO 48 (2001).

behave less like physical goods. By contrast, something like crops of corn require physical goods to grow (water, fertilizer, soil). Crops also must be grown in a physical place, transported physically for sale, and once purchased cannot be shared without depriving the original owner of the good. Online, ideas exhibit very different characteristics: they are built on top of other ideas, they sometimes exist only ethereally on hard drives, and they are transmitted instantly and cheaply. These intangible ideas increasingly drive the growth of the economy as a whole, and Romer's explanation of this phenomenon (to be explored further shortly) helps explain how some economies are able to grow much more rapidly than their linear counterparts.

A third attribute of ideas is that they are relatively easy and inexpensive to share, at least outside the confines of intellectual property law. Economist Ronald Coase first explained in the late 1930s how modern firms are formed to reduce transaction costs that would otherwise make many types of production prohibitively inefficient.[233] When using traditional price mechanisms, certain costs are introduced in each transaction, such as matching buyer and seller, negotiating the contract, and other overhead related to exchanging the good. Within a firm, these costs can be dramatically reduced or eliminated entirely because activities can be coordinated without negotiating prices, paying sales taxes, or incurring other costs.

In a highly networked environment, sharing certain types of goods becomes much easier and cheaper than ever before. Specifically, nonrival ideas can spread effortlessly to the extent that they are non-excludable. Sharing of innovations has historically involved low transaction costs; with the advent of the Internet, these costs approach zero. Once an idea has been created, eliminating barriers to sharing it can lead to the most efficient use and further innovation. Coase's bright line between individual and firm begins to blur in an innovation economy that takes full advantage of this structure.[234] In some cases, loosely related individuals operating outside of market dynamics develop critical components of the technological infrastructure that multiply production throughout a variety of sectors of the economy.

A final important aspect of ideas is that they flourish in open systems. Whereas industrial economies based on physical capital require large firms, networked economies thrive when small businesses and entrepreneurs innovate in a maximally open environment. The network ideally should enable access to markets and connect people to ideas, regardless of size. These diversely configured actors then introduce

233. Coase, *supra* note 92, at 395-98.

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

growth-fueling ideas for the next generation of producers. Throughout history, goods have been manufactured primarily for a purpose that was known by the producer. Nevertheless, they have often been modified to the great benefit of the overall market. Ideas are especially flexible in this way, and the platforms we use to generate them are most efficient when they facilitate that flexibility.

2. Innovation

If there is any one business lesson of the last decade that has acquired near-universal empirical support and expert agreement, it is this: innovation is a good thing. The creation of new and different objects, processes, and services are at the heart of any rational conception of economic growth and the fulfillment of human potential. No matter what you call it—creativity, entrepreneurism, novelty, ingenuity—the global economy feeds on the constant infusion of the products of innovation.

The 20th Century will be credited by many as the century of innovation.[235] Indeed, one historian has demonstrated that the "accelerating growth [of useful knowledge] . . . has affected the world more [profoundly] than all other social and political changes taken together."[236] Innovation is a much-admired concept, yet in many ways still rather mysterious and elusive. It has been defined in some quarters as invention plus implementation.[237] Where ideas are the raw makings of a recipe, innovation is the fashioned process or product. More specifically, innovation involves the process of taking a raw idea and developing it into a concept, which "yields some type of invention, and which is finally implemented and commercialized."[238] However one chooses to define it, "[i]nnovation is the source of economic variation,"[239] and "the key factor enabling growth and change in capitalist economies."[240]

Research shows conclusively that innovation tends to flow from the users, not the consumers or providers; from the many, not the few; from the connected, not the isolated; from individuals and small groups, not larger organizations; from the upstarts, not the established; from the decentralized, not the concentrated; from the flat, not the hierarchical; from the autonomous, not the controlled. [241] Innovation is produced from

235. GERARD H. (GUS) GAYNOR, INNOVATION BY DESIGN 1 (2002).
236. JOEL MOKYR, THE GIFTS OF ATHENA 297 (2002).
237. GAYNOR, *supra* note 235, at 3.
238. *Id.* at 3, 7.
239. COYLE, *supra* note 177, at 189.
240. McKnight, *supra* note 171, at 39-41.
241. *See, e.g.*, CARLISS Y. BALDWIN & KIM B. CLARK, DESIGN RULES (1999); JOHN SEELY BROWN & PAUL DUGUID, THE SOCIAL LIFE OF INFORMATION (2000); BHASKAR CHAKRAVORTI, THE SLOW PACE OF FAST CHANGE (2003); CLAYTON M.

those users motivated by many incentives, including profit, pride, and personal fulfillment. The arrival of innovation is not usually predictable or orderly;[242] indeed, "invention often is the mother of the unforeseen."[243] While individual innovations tend to be minor and incremental, collectively they create technical progress. There is also a separate "demand side" perspective to innovation, based on extensive research showing that "venturesome" consumers adopting and using technology are crucial to maintaining economic prosperity.[244]

Clayton Christensen placed the concept of innovation squarely before the general public in his acclaimed trilogy.[245] His writings focus on what he calls sustaining innovations—those allowing firms to provide better and more profitable products to their customers—as opposed to disruptive innovations—those offering initially poorer performance along the dimension that existing customers care about the most. Christensen found that modularity can have a profound impact on industry structure, because it enables independent, nonintegrated organizations to sell, buy, and assemble components and subsystems. "[I]n a modular world, [firms] can prosper by outsourcing, or by supplying just one element."[246] Such firms can become, not just a mere link in a value "chain," but an integral component of a complex and evolving value "net."

CHRISTENSEN, SCOTT D. ANTHONY, & ERIK A. ROTH, SEEING WHAT'S NEXT? USING THE THEORIES OF INNOVATION TO PREDICT INDUSTRY CHANGE (2004); BORU DOUTHWAITE, ENABLING INNOVATION (2002); GAYNOR, *supra* note 235; ANDREW HARGADON, HOW BREAKTHROUGHS HAPPEN (2003); FRANS JOHANNSON, THE MEDICI EFFECT (2004); LARRY LESSIG, THE FUTURE OF IDEAS (2002); CONSTANTINOS MARKIDES & PAUL GEROSKI, FAST SECOND (2004); MCKNIGHT, VAALER & KATZ, EDS., *supra* note 171; JOHN MCMILLAN, REINVENTING THE BAZAAR (2002); MOKYR, *supra* note 236; DAVID NYE, TECHNOLOGY MATTERS (2006); HOWARD RHEINGOLD, SMART MOBS (2002); OGLE, *supra* note 67; NELLY OUDSHOORN & TREVOR PINCH, EDS., HOW USERS MATTER (2003); SCOTT PAGE, THE DIFFERENCE (2007); JOHN THACKARA, IN THE BUBBLE (2005); JAMES M. UTTERBACK, MASTERING THE DYNAMICS OF INNOVATION (1994); VON HIPPEL, *supra* note 67.

242. *See* SCOTT BERKUN, THE MYTHS OF INNOVATION 30 (2007).

243. NYE, *supra* note 241, at 159. Importantly, "since innovation means doing something never done before, there is an element of genuine uncertainty in all innovative activity." LIPSEY ET AL., *supra* note 125, at 30.

244. AMAR BHIDE, THE VENTURESOME ECONOMY (2008). Bhide argues that "the willingness and ability of users to undertake a venturesome part plays a critical role in determining the ultimate value of innovations." *Id*. at 323. He cites a number of supportive elements of the U.S. economic system, including a high level of inclusiveness and participation, a wide variety of organizational forms, venturesome beliefs that embrace new technologies and goods, and a premium on growth. *Id*. at 409.

245. CLAYTON M. CHRISTENSEN, THE INNOVATOR'S DILEMMA: WHEN NEW TECHNOLOGIES CAUSE GREAT FIRMS TO FAIL (1997); CLAYTON M. CHRISTENSEN & MICHAEL E. RAYNOR, THE INNOVATOR'S SOLUTION (2003); CHRISTENSEN, ANTHONY, AND ROTH, *supra* note 241.

246. Clayton M. Christensen & Scott D. Anthony, *Disruption, Disintegration, and the Impact of New Telecommunications Technologies*, *in* THE BROADBAND EXPLOSION 91, 99 (Robert D. Austin & Stephen P. Bradley eds., 2005).

Christensen believes that the Internet's decoupling of services and transport creates innovative new business models across customers and markets. In his words, "IP is the ultimate modular interface."[247] Users operating at the so-called edge of the Internet are responsible for many of the key innovations that we enjoy today. The Internet in itself can be seen as a rare breakthrough innovation.[248] Lee McKnight posits that the Internet facilitates rapid development and diffusion of innovations by network users. IP acts as a "bearer service"—the general purpose platform technology linking technologies, software, services, customers, firms, and markets—so that the Internet becomes "an innovation engine that enables creation of a remarkable range of new products and services."[249] Thus, "the Internet works its magic through rapid development, diffusion, and validation of innovations."[250] Benkler describes how the Internet helps disrupt the traditional producer/consumer model by empowering the rise of end users who can play both roles as part of a continuing conversation and exchange of information. The "Great Shopping Mall" can be transformed into the "Great Agora," featuring "unmediated conversation of the many with the many."[251] "The Internet may be considered a disruptive innovation, but in essence it's a new way of doing business—a new tool to accomplish the same result."[252] As Crawford puts it, the central presumption of Internet innovation is that "everything not prohibited is permitted."[253]

Obviously, these observations amount to a generalization, one not true for all times, places, and people. Certainly, there are innovative large, entrenched organizations—think Apple—and countless uncreative small ones. Nor can a market system survive only with innovation-churning entrepreneurs; after all, "[b]ig firms remain essential to refine and mass-produce the radical innovations that entrepreneurs have a greater propensity to develop or introduce."[254] With regard to the Internet, innovations also are not limited to the content and applications layers, or to consumer-facing retail offerings; they happen deep in the

247. *Id.* at 104.

248. GAYNOR, *supra* note 235, at 34.

249. McKnight, *supra* note 171, at 40 (citation omitted).

250. *Id.* at 41 (citation omitted).

251. Yochai Benkler, *From Consumers to Users: Shifting the Deeper Structures of Regulation Toward Sustainable Commons and User Access*, 52 FED. COMM. L.J. 561, 565 (2000).

252. GAYNOR, *supra* note 235, at 31. "Software platforms have been found to accelerate further the process of creative destruction, mainly because code is digital and malleable," and can be distributed over the Net to potentially billions of computing devices around the world. DAVID S. EVANS ET AL., INVISIBLE ENGINES 338 (2006).

253. Susan P. Crawford, *Shortness of Vision: Regulatory Ambition in the Digital Age,* 74 FORDHAM L. REV. 695, 724 (2005).

254. WILLIAM J. BAUMOL ET AL., GOOD CAPITALISM, BAD CAPITALISM, AND THE ECONOMICS OF GROWTH AND PROSPERITY 92 (2007).

logical and physical infrastructure of the network. Indeed, layering with IP at the center allows for significant network innovation below, as well as above, the IP layer. If nothing else, however, the concept of "innovation from the edge" provides a useful corrective to present-day presumptions about how markets actually work in a capitalist society, and highlights the importance of the edge of the Internet to the rest of us.

B. *Economic Growth*

So what is growth? To most economists it means a rising standard of living for a country's citizens, measured according to the increase in gross domestic product (GDP) per capita. More generally, growth is measured according to how much value a nation is perceived to produce for each citizen. Growth in economic output per person is "the most important measure and determinant of economic performance . . . ."[255] To Emergence Economics, growth arises from the discovery of new recipes, and the transformation of things from low-value to high-value configurations. In shorthand, it is turning ordinary sand into semiconductors. Romer explains it this way:

> Economic growth occurs whenever people take resources and rearrange them in ways that are more valuable. A useful metaphor for production in an economy comes from the kitchen. To create valuable final products, we mix inexpensive ingredients together according to a recipe. The cooking one can do is limited by the supply of ingredients, and most cooking in the economy produces undesirable side effects. If economic growth could be achieved only by doing more and more of the same kind of cooking, we would eventually run out of raw materials and suffer from unacceptable levels of pollution and nuisance. Human history teaches us, however, that economic growth springs from better recipes, not just from more cooking. New recipes generally produce fewer unpleasant side effects and generate more economic value per unit of raw material.[256]

Some see undesirable effects from growth, like the disruption of traditional culture, congestion, and damage to the environment.[257] While these very real social costs should not be downplayed, "conventional thinking about economic growth fails to reflect the breadth of what growth, or its absence, means for a society."[258] Most think only of a

255. ATKINSON, *supra* note 44, at 141. GDP measures both hours worked and productivity; Atkinson explains that only the latter should be considered relevant for purposes of expanding economic growth. *Id*. at 239-40.

256. Paul Romer, *Economic Growth*, *in* THE CONCISE ENCYCLOPEDIA OF ECONOMICS 128 (2007).

257. For a brief recitation, see VERMEIJ, *supra* note 102, at 295-302.

258. BENJAMIN M. FRIEDMAN, THE MORAL CONSEQUENCES OF ECONOMIC

higher material standard of living,[259] but there are also significant social, political, and moral benefits not priced by the market. Moreover, "changes in per capita GDP radically understate the impact of economic growth on the average person."[260] More often than not, economic growth "fosters greater opportunities, tolerance of diversity, social mobility, commitment to fairness, and dedication to democracy."[261] However, economic progress needs to be broadly based if it is to foster real social and political progress; conversely, stagnation in living standards can lead to rising intolerance and incivility.[262] Growth's byproduct of increased leisure also has had "the most liberating and enriching impact on the citizens of the West."[263] Further, the "enhancement of human freedom is both the main object and the primary means of [economic] development."[264] Romer sums it up that "better growth policy could have implications for the quality of life in all dimensions that are so large that they are hard to comprehend."[265]

So economic growth is a key component to a country's well being. As Romer observes, "[b]y far the most important characteristic of capitalist economies, which distinguishes them from all other previously and currently existing societies, is their slow but steady underlying rate of real economic growth."[266] Still, economists long have sought to understand the mystery of how economic growth happens, and why some nations seem to make sudden jumps whereas others grow slowly.[267] To be sure, the economy is a complex network of interactions; individual agents, acting according to diverse incentives, create growth as an emergent phenomenon.[268] In the late 1980s, however, a new generation

---

GROWTH 4 (2005).

259. "It is estimated that the purchasing power of the average American a century ago was one-tenth what it is today." BAUMOL ET AL., *supra* note 254, at 1-2; *see also* LIPSEY ET AL., *supra* note 125, at 5.

260. LIPSEY ET AL., *supra* note 125, at 6.

261. Friedman, *supra* note 258, at 4. Moreover, "[w]hile economic growth makes a society more open, tolerant, and democratic, such societies are in turn better able to encourage enterprise and creativity, and hence to achieve ever greater economic prosperity." *Id.* at 15.

262. *Id.* at 9.

263. PAUL ORMEROD, THE DEATH OF ECONOMICS 23 (1994).

264. AMARTYA SEN, DEVELOPMENT AS FREEDOM 53 (1999). "Individual capabilities crucially depend on, among other things, economic, social, and political arrangements . . . . [H]uman beings are not merely the means of production, but also the end of the exercise." *Id.* at 53, 296.

265. Paul Romer, *Should the Government Subsidize Supply or Demand in the Market For Scientists and Engineers* 47 (Nat'l Bureau of Econ. Research, BER Working Paper No. 7723, 2000).

266. ORMEROD, *supra* note 6, at 48.

267. Atkinson observes, for example, that neoclassical economics does not have the tools or orientation needed to understand why the United States entered a 20 years period of slowing growth beginning in 1973. ATKINSON, *supra* note 44, at 143-47.

268. Lipsey explains that economic growth can be understood in terms of five partially

of economists began to appreciate the concept of growth in the context of technological progress.

New Growth Theory reminds us that growth flows from within the system itself, and is directly and profoundly affected by conscious decisions made by economic actors. As Crawford puts it in the context of networked economies, "[t]he economic growth-based . . . [story] is straightforward: the greatest possible diversity of new ideas that will support our country in the future will come from the online world, because of its special affordances of interactivity, interconnectivity, and unpredictable evolution."[269] If we wish to secure the Internet as an engine for growth going forward, however, we first must understand how to preserve this generative quality in the midst of complex network effects.

1. The Road to Romer

Adam Smith's foundational 1776 work, *The Wealth of Nations*, theorized that as a firm developed specialized roles for workers, their skills would benefit the productivity of the firm and thus the market overall. The cost of goods they produced would be disciplined by the "invisible hand" of competitive pricing, and the market would converge on an optimally efficient equilibrium.[270] In the mid-20th century, Joseph Schumpeter modified this competitive hypothesis, pointing out that firms often form temporary monopolies and subsequently are unseated by other firms in an act he called "creative destruction." The critical advantage of these winning new entrants is their improved technology. Through this process, innovation occurs in a stair-step fashion rather than a continuous line.[271]

Much of economic growth theory has focused on how best to encourage development of these technologies. Nobel Prize winning economist Robert Solow recently observed that Schumpeter:

> worked out his conception of the entrepreneur, the maker of "new combinations," as the driving force and characteristic figure of the

---

independent, partially interacting subsystems: economic, technological, scientific, political, and cultural. LIPSEY ET AL., *supra* note 125, at 374-77. The coordination of these dynamic evolving systems—semi-autonomous and semi-interdependent—occurs as an emergent property, resulting from the actions of countless individuals and groups. *Id.*

269. Crawford, *supra* note 223, at 6-7.

270. ADAM SMITH, THE WEALTH OF NATIONS 447 (N. Kelly, vol. 1, 1801) (1776), *available at* Google Books. Robert Reich calls Smith's invisible hand "the most famous, or infamous, bodily metaphor in all of social sciences." TAYLOR, *supra* note 6, at 85. Mark Taylor claims that the image originated not with Smith, but John Calvin, who used it to describe God's providence in the world. Smith then appropriated Calvin's doctrine of providence to explain the machinations of the market. *Id.* at 4, 85. Interestingly, the "invisible hand" also can be reinterpreted for modern ears as unguided, emergent behavior of the market system.

271. *See generally* SCHUMPETER, *supra* note 35.

> fits-and-starts evolution of the capitalist economy. He was explicit that, while technological innovation was in the long run the most important function of the entrepreneur, organizational innovation in governance, finance, and management was comparable in significance.. . . I think that this is Schumpeter's main legacy to economics: the role of technological and organizational innovation in driving and shaping the growth trajectory of capitalist economies.[272]

This distinction between technological and organizational innovation is a mirror of the Physical Technologies and Social Technologies that Richard Nelson has identified at the heart of complex economic growth. Even though Solow refers only to the first type of innovation explicitly as a "technology," he is saying the same thing—new ways of working with things and new ways of organizing people are the most important contributions to economic growth.

Solow's own work on growth theory in the 1950's was highly influential, but ultimately failed to fully explain the stair-step pattern of technological progress that Shumpeter described. In Solow's growth model, technology fed into the system at a steady rate. When it came to explaining what generated this innovation, however, the Solow model was at a loss, because it treated this technological advance as something that happened *exogenously,* coming from outside the economy itself. To be sure, technology had assumed a place of importance, but the core question of how to encourage technology and the resulting growth remained unanswered.

2. Enter New Growth Theory

In fact, Schumpeter's core claims about how technological change happens would lay somewhat dormant until the 1980s. By the end of the decade, younger generations of economists were hard at work on the "increasing returns" problem. In short, they asked "why do some economies appear to grow very rapidly, despite the assumption that all the traditional inputs are increasing at a steady rate?"

Exogenous factors are background conditions and givens that lay outside an economic model. In traditional economic theory, the factors of production are land, labor, and capital. Knowledge and human nature were simply "givens," a fixed part of the background.[273] In 1990, then-unheralded economist Paul Romer released a paper where he concluded that the new factors of production should be classified as people, ideas, and things. More importantly, he found that technological change and

---

272. Robert M. Solow, *Heavy Thinker*, NEW REPUBLIC, May 21, 2007, at 2, 3, *available at* http://www.powells.com/review/2007_07_12.

273. COYLE, *supra* note 177, at 39.

the growth of knowledge should be viewed as endogenous to the system. Romer cited a clear distinction between rival goods (corporeal goods of absolute possession and limited sharing) of objects (atoms), versus nonrival goods (non-corporeal goods that can be copied or shared and used by many people at the same time) of ideas (bits).[274]

So there are objects, and there are ideas. And to Romer, ideas are what truly matter in generating economic growth. He accepted Kenneth Arrow's observation that information (and therefore technological progress) is not only a product of the economy but also an input back into it, creating a positive feedback effect.[275] At the same time, he noted Schumpeter's point that firms can be spurred to innovate to gain or retain their market power.[276] But he also altered these basic ideas in critical ways. Arrow's feedback loop of technological knowledge became not simply learning-by-doing within firms, but rather a global multiplier of productivity when this *non-rival* information resource was shared. Ideas, Romer explained, cannot be over-used. Schumpeter's "creative destruction" could happen, according to Romer, in situations where monopoly was neither complete nor highly difficult to overcome.[277]

A world of objects does not lead to sustained growth, let alone exponential growth. Instead, growth happens when information is fed back into the economy for ongoing re-use, never diminishing the usefulness of the information itself. Mechanisms generating new ideas are as important as access to abundant resources for economic growth.[278] *Optimal* growth happens when the non-rivalry of information is balanced by the appropriate degree and type of excludability, giving innovators incentive to undertake research and development in the first place. In short, "technological change . . . lies at the heart of economic growth."[279]

Romer's work ignited a wave of research on *endogenous* growth—the explanation of how growth is fed by economic factors.[280] It was called "new growth theory" to distinguish it from the "neoclassical" approach derived from Solow. In new growth theory, technological progress became a critical fourth component of economic growth models—both on the input and the output sides of the equation. Professor Charles Jones asserts that Romer's papers "lay out with startling clarity the link

274. Romer, *supra* note 229, at S71-D102.
275. *See id*. at S77-S76.
276. *See id.* at S76-S78.
277. *See id.*
278. VERMEIJ, *supra* note 102, at 310.
279. Romer, *supra* note 229, at S72.
280. *See generally* ELHANAN HELPMAN, THE MYSTERY OF ECONOMIC GROWTH (2004); *see also* Philippe Aghion & Peter Howitt, *Market Structure and the Growth Process*, *in* 1 REV. OF ECON. DYNAMICS 276 (1998).

between economic growth and ideas."[281]

The economics of ideas is different from the economics of objects. For example, because ideas are non-rival, my use of an idea does not inherently reduce the "amount" of the idea available for you to use.[282] There also is an increasing return to ideas, and to ideas and objects together. This notion of the increasing return of ideas is one of the central elements of Romer's theory. Professor Jones points out one ground-breaking consequence: the accumulation of ideas is able to sustain growth in a way that the accumulation of capital cannot.[283] The strongest growth, and the most "virtuous" complex network, feeds ideas back into the system.

The mystery of why some nations grow at faster rates than others can now be explained, at least in part, in terms of how effectively a particular nation's economy optimizes the creation of these new innovative recipes for growth.[284] As Romer says, "it is ideas, not objects, that poor countries lack."[285] Pointing to a different component of the "rough equation" for emergence, Benkler has observed that "what most poor and middle-income countries lack is not human creativity, but access to the basic tools of innovation."[286]

3. Implications For Technological Change

So "long-term growth is driven mainly by technological change," and in turn "new technologies radically alter more or less everything in the socio-economic order."[287] Still, Romer's work, while monumental, leaves important details to be worked out. How does trade between nations affect this dynamic? What is the appropriate balance between the incentives of exclusion and the increasing returns of openness? Are there modes of production in which the innovators share freely from the start? What types of technologies afford the greatest growth potential when they are not restricted by exclusion?

One potential answer to the final question has been articulated in

---

281. C.I. Jones, *Growth and Ideas*, *in* HANDBOOK OF ECONOMIC GROWTH 1070 (2005). By contrast, as Business Week's chief economist puts it, many economists "grudgingly acknowledge the importance of technological change, but they do not understand it or trust it." MICHAEL MANDEL, RATIONAL EXUBERANCE xii (2004).

282. Jones, *supra* note 281, at 1066.

283. *Id.* at 1075.

284. Economist Douglass North argues convincingly that the development of legal institutions and norms that support market performance also are necessary to undergird a nation's successful economic growth. NORTH, *supra* note 222, at 155-56.

285. Romer, *supra* note 256, at 129.

286. BENKLER, *supra* note 67, at 468. *See also* EASTERLY, *supra* note 52, at 176 ("[D]ifferences in productivity growth explain over 90 percent of the differences across countries in per capita growth between 1960 and 1992.").

287. LIPSEY ET AL., *supra* note 125, at xv.

the ongoing research on "General Purpose Technologies" ("GPTs"). A GPT is a special type of technology that has broad-ranging enabling effects across many sectors of the economy. Some define a GPT as a generic technology that eventually comes to be widely used, to have many uses, and to have many spillover effects.[288] The foundational work on GPTs was first published by Timothy Bresnahan and Manuel Trajtenberg in 1992.[289] They describe how this particular type of technology is most likely to generate increasing returns in line with Arrow and Romer, with growth coming from specific applications that depend on ideas in the "general" layer of technology. Specifically, GPTs play a role of "enabling technologies" by opening up new opportunities rather than offering complete, final solutions.[290] The result is "innovational complementarities," meaning that "the productivity of R&D in a downstream sectors increases as a consequence of innovation in the GPT technology. These complementarities magnify the effects of innovation in the GPT, and help propagate them throughout the economy."[291]

Whereas Romer focused generally on the overall economy, the GPT literature makes clear that some technologies are especially important when it comes to non-rival reuse and follow-on innovation. Over the past decade, economists have expounded upon how electricity, motors, personal computers, and software platforms all exhibit this characteristic.[292] Citing Trajtenberg's work, Joel Mokyr makes a persuasive case that the semiconductor is the greatest "macroinvention" since the emergence of electricity.[293] The Internet in particular is a GPT, with "the potential to contribute disproportionately to economic growth" because it generates value "as inputs into a wide variety of productive activities engaged in by users."[294] One lesson for policymakers is that

---

288. *Id*. at 98.

289. Timothy Bresnahan & Manuel Trajtenberg, *General Purpose Technologies 'Engines of Growth'?* (1992), *reprinted in* 65 J. OF ECON. 1, 83 (1995).

290. *Id*. at 84.

291. *Id*.

292. *See, e.g.*, EVANS ET AL., *supra* note 252 (software platforms are a type of GPT). Lipsey and his colleagues put GPTs into six major categories: materials technologies (domesticated animals), power (the dynamo), information and communication technologies (the Internet), tools (the wheel), transportation (railway), and organization (the factory system). LIPSEY ET AL., *supra* note 125, at 133.

293. MOKYR, *supra* note 236, at 112 (explaining that the semiconductor's unusual properties as an innovation merit its status as a GPT, including its ability to recombine with other techniques, its complementarities with downstream innovations, and its consequent pervasiveness in many applications).

294. Brett M. Frischmann & Barbara van Schewick, *Network Neutrality and the Economics of an Information Superhighway: A Reply to Professor Yoo*, 47 JURIMETRICS 383, 424 (2007). Lipsey observes that the current ICT revolution is not unique; there have been other GPT-driven "new economies" in the past. Further, while "GPTs have not been common in human

when considering the appropriate balance between various market incentives, one must also consider the extent to which a particular type of technology is a GPT. Looking back at the development of the IT industry more than ten years after his key GPT paper, Bresnahan commented:

> But let us be clear that the lesson here for Schumpeterian Economics is far more general than the narrow and specific point about "open architecture," which seems like a technical concept from computing. Instead, the point is about the role of a permissive, forward-looking system of innovation in which inventions can come from multiple sources. . . . The most economically important use of a general purpose technology need not be determined by the inventors of the GPT, but rather by the inventors of complements, applications.[295]

The additional lesson drawn from Schumpeterian economics is that not all market dynamics are strictly Schumpeterian—at least in the way that term often is employed in contemporary policy rhetoric. To be sure, dynamism, waves of destruction, and temporary incumbency all are part of healthy markets. Nonetheless, the "Schumpeterian" perspective too often becomes twisted into bald assertions that policymakers have no useful role whatsoever because market power inevitably is transitory, and leads invariably to innovation. This is an unhelpful dumbing-down of Schumpeter's insights, one which also overlooks additional significant progress in economic thinking since his time.

Keeping GPTs "general" is not always in the clear interest of a firm that might seek to control them. That firm might envision greater profits or efficiency through making a tremendously useful resource more scarce, by charging much higher than marginal cost, or by customizing solely for a particular application.[296] While these perceptions might be true in the short term, or for that one firm's profits, they can have devastating effects for growth of the economy overall. The more general purpose the technology, the greater are the growth-dampening effects of allowing it to become locked-down in the interest of a particular economic agent.

Relatedly, Jonathan Zittrain of the Oxford Internet Institute has

experience," the rate of innovation in GPTs has accelerated greatly over the last 10,000 years. LIPSEY ET AL., *supra* note 125, at 131-32.

295. Timothy Bresnahan, *Creative Destruction in the PC Industry*, *in* PERSPECTIVES ON INNOVATION 105, 114, 118 (Franco Malerba & Stefano Brusoni eds., 2007).

296. This is not to say that firms cannot create specialized implementations of GPTs. On the contrary, much of the value of GPTs comes from specific instantiations. Nobody would think of toting a desktop computer on a plane in order to work en route, but most laptop computers are not fundamentally different with respect to their general-purpose nature than desktops. If, however, a firm obtained and exercised control over the fundamental PC, operating system, or network platforms, welfare-enhancing specializations would be foreclosed.

discussed how "the generative Internet" has great "capacity to produce unanticipated change through unfiltered contributions from broad and varied audiences."[297] The important feature of generative platforms, such as the Internet, is that users easily can do numerous things with them, many of which may not have been envisioned by the designers. If, for example, the Internet had been built solely as a platform for sending email, and required retooling to do anything else, most applications and business models never would have developed.[298]

Finally, technological change is a historical process in which there is a clear arrow of time. Because "agents' behavior and choice sets are path-dependent, technological change is replete with the possibility of multiple equilibria, lock-ins, and possible 'butterfly effects.'"[299] When it comes to new generative platforms like the Internet, history matters.

C. The External is in: "Net effects"

Of course, the Internet is not just about reducing costs and increasing the supply of goods and services. To many people, the Net is about the less-tangible values. The market is not everything.

Old School Economics defines the "market" in a fairly narrow way, as the result of trading capital and goods and services, and recognizes no values that are not expressed in actual market choice behavior. However, the market can be seen as far more than the sum of economic transactions. Viewed from a broader perspective, the market is more akin to The Great Agora of ancient Greece, the marketplace that held many important human interactions, whether for pecuniary gain or other intrinsic benefits.[300] Under this view, people are more than a bundle of economic wants and needs. Not just consumer or users, we have meaning beyond our economic activity. We value things that we will never purchase.[301] To the extent Old School Economics fails to account for these values and activities, it paints an incomplete picture of human behavior.

The value of new ideas and innovation goes far beyond the sum of explicit capital exchange. As outlined below, new ideas result in innovation spillovers, social production, and even the rise of a new "social layer" of the Net. Benjamin Friedman makes a similar point when he reminds us that economic growth creates a host of social, political, and

---

297. JONATHAN L. ZITTRAIN, THE FUTURE OF THE INTERNET AND HOW TO STOP IT 70 (2008).

298. *See id.*

299. LIPSEY ET AL., *supra* note 125, at 14.

300. BENKLER, *supra* note 251, at 565.

301. STANOVICH, *supra* note 60, at 261.

political goods that private markets do not trade or price.[302] These additional benefits are not just along for the ride; in addition to being important in their own right, they form a synergistic relationship with economic gain and growth. Recognizing these very real "externalities" goes against the pervasive bias in Old School Economics that only that which can be modeled or quantified as directly benefiting pecuniary interests is meaningful.

1. Innovation Spillovers

When new technologies are developed, they can benefit the private innovator, the public, or both. If the innovator has the ability to exclude others from using the technology, he can charge for its use and thus capture its value. However, it is uncommon for innovations to be completely excludable, and even rarer that the technology controller can find a way to "internalize" all of the value. Invariably, some of the value "spills over." Thus, spillovers are a type of externality from those transactions. The private benefits or costs do not completely capture the social benefits or costs, so one market participant's actions affect others without compensation being paid or received. Pollution is a classic example of a negative externality; scientific research is one form of positive externality.

Professors Frischmann and Lemley, among others, wish to bring back into economic thinking the overall benefits of preserving spillovers, here defined as "uncompensated benefits that one person's activity provides to another."[303] Spillovers generally fall into two categories: unanticipated consumer surplus, and third party benefits.[304] Regardless of the classification scheme employed, it is a simple fact that "no innovator has captured all or even most of the social benefits of his or her invention."[305] Frischmann and Lemley's primary thesis is that the social value of innovations often far exceeds the private value, and these spillovers in turn actually encourage greater innovation.[306]

Economists sometimes try to distinguish between "real" spillovers that truly benefit parties outside the market, and "pecuniary" spillovers that are ultimately resolved and accounted for within a market and between private parties.[307] However, real spillovers often end up feeding

302. Friedman, *supra* note 258, at 400.

303. Brett M. Frischmann & Mark Lemley, *Spillovers*, 100 COLUM. L. REV. 101, 102 (2006).

304. *Id.* at 105. Others have used the language of knowledge "leaks" which create a gap between social returns and individual (or firm) returns. EASTERLY, *supra* note 52, at 152-53.

305. Frischmann & Lemley, *supra* note 303, at 103.

306. Moreover, if all spillovers were completely captured by the original innovator, the economic conditions of everyone else would not improve. In essence, spillovers lift all boats. *Id.*

307. Those spillovers resulting in a wedge between private and social returns—and thus

back into the economy in the form of overall social welfare, while pecuniary spillovers have unforeseen long-term effects through nonrival reuse and follow-on innovation. Economic literature on various types of spillovers reflects this ambiguous nature of externalities versus internalities. For example, some economists speak of "network externalities" while others insist on "network effects."[308] The dynamic and social value of ideas makes them difficult to account for purely in terms of internalized private transactions.

Traditional economic thinking dictates that the economy functions best when firms maximally internalize "real" spillovers, using property rights and price signaling to allocate resources efficiently. The economics of innovation paint a somewhat different picture, particularly when it comes to general purpose technologies. GPTs help "rejuvenate the growth process by creating spillovers that go far beyond the concept of measurable externalities" and far beyond those agents that initiated the change.[309] This has important implications when trying to tally the sum total of beneficial value and activity generated by the Internet.

2. Peer Production

Yochai Benkler's *The Wealth of Networks* lays out the case for social production. The wealth of networks is in their potential for widespread participation in making, sharing, and experiencing information.[310] Benkler discusses the current and potential benefits from social production, the "networked information economy," and the possibility of reversing the control focus of the industrial information economy.[311] He argues that social production serves important values, including autonomy, democratic participation in the political sphere, construction of culture, justice and human development, and community.[312] He also points out that "advanced economies rely on non-market organizations for information production much more than they do in other sectors."[313]

Benkler's major work addresses some nine different "ideal-type

affecting net welfare—are deemed to be "real" or "technological;" examples include the social benefits of education, and the social costs of pollution. Those spillovers involving only a transfer of surplus wealth between producers and consumers are deemed "pecuniary;" the classic example is a consumer purchasing a product for less than he or she was willing to pay. *Id.* at 106-07.

308. *E.g.* Stan Liebowitz & Stephen Margolis, *Network Externalities (Effects)*, *in* THE NEW PALGRAVE'S DICTIONARY OF ECONOMICS AND LAW (1998).

309. LIPSEY ET AL., *supra* note 125, at 98,100.

310. *See also* COOPER, *supra* note 86, at 127 (new forms of collaborative production "bind people together in productive, social, and economic relations to produce and self-supply an increasing array of micro-products to met their needs.").

311. BENKLER, *supra* note 67, at 29-32.

312. *Id.* at chs.5-10.

313. *Id.* at 47.

information production strategies."[314] Here we will only distinguish between two broader modes of human behavior intrinsic to his work. At root, Benkler is interested in *peer production*, a form of activity outside the traditional producer/consumer relationship. Information agents in highly connected networks with low transaction costs find new ways of working together and generating productivity. The Internet, of course, provides plentiful examples of this type of production

In some of these modes of production, traditional economic incentives play a part, and agents seek traditional economic gains.[315] The resulting innovation is important because users get precisely what they want, which in turn increases social welfare.[316] Social welfare is very probably increased by the presence of innovations freely revealed by users.[317] This does not count innovations made available in commerce. "Social welfare is likely to be higher in a world in which both users and manufacturers innovate than in a world in which only manufacturers innovate."[318] Social welfare equals the total income of a society. Similarly, the principles of "Wikinomics" (being open, peering, sharing, and acting globally) lead to what its proponents call mass collaboration. The claim is that, "with peer production we will harness human skill, ingenuity, and intelligence more efficiently and effectively than anything we have witnessed previously."[319] The "Long Tail" of economic abundance (where the mass of niche markets creates a long-tail graphical distribution) helps power this peer production.[320]

Another group of these modes of production constitutes *social production*, in which agents operate under non-traditional incentives. Here, Benkler points out that people produce things for different reasons, including a variety of social-psychological rewards.[321] So incentives clearly matter to people, but not all are purely economic. Philip Weiser estimates that 70 percent of all Web pages "are built by individuals from their desire to share ideas, rather than to make money."[322] Research suggests that computer programmers working for money are likely to be less creative than those programming as a hobby in their own time. There can be an inverse relationship between creativity

314. *Id*. at 43.
315. *Id*. at 462-63.
316. VON HIPPEL, *supra* note 67, at 2-3.
317. *Id*. at 11-12.
318. *Id*. at 107.
319. TAPSCOTT & WILLIAMS, *supra* note 214, at 18.
320. CHRIS ANDERSON, THE LONG TAIL 73, 219 (2006).
321. BENKLER, *supra* note 67, at 96.
322. Philip Weiser, *Law and Information Platforms*, 1 J. ON TELECOMM. & HIGH TECH L. 1, 33 n.147 (2002) (*quoting* Kevin Kelly, *The Web Runs on Love, Not Greed*, WALL ST. J., Jan. 4, 2002, at A8).

and external reward.[323] Volunteer contributors of code to widely used software products are often strongly motivated to innovate by the joy and learning they find in their work.[324]

Benkler argues that "the basic technologies of information processing, storage, and communication have made nonproprietary models more attractive and effective than was ever before possible." Among other things, this allows new "patterns of social reciprocity, redistribution, and sharing . . . ."[325] In a sufficiently open and ubiquitous network, the benefits of the traditional firm can apply to all individuals.

3. The "Social Layer"

Crawford takes perhaps the most expansive view of the non-pecuniary benefits of the Internet as an ideas and innovation platform that enables human interactivity. In "The Project of Communications Law," she claims that a focus only on future applications-layer innovation from the Internet (as promulgated by Web companies), or the Net as a "content-delivery supply chain" (as seen by the broadband companies and the FCC), provides an "impoverished (or at least incomplete) perspective on communications."[326] Interestingly, she includes Yochai Benkler and his social production vision in the "application-layer" centric camp.

Crawford views human online communities as a form of complex adaptive system, which generate not only economic growth, but also new forms of persistent social interaction and dynamic human relationships, which evolve in complex and unpredictable ways. She urges "a changed perspective on the internet that takes as central the evolution of human connections and relationships online."[327] As part of that mission, she touts "cognitive diversity," which ensures that "people with diverse experiences training, perspectives, predictive models, interpretations, and tools are online."[328] She sees the Net as allowing "innovation in social relationships at a system level," which goes beyond seeing the "content" layer of the Internet as the "social layer."[329] The existence of such a social layer promotes diversity, the democratization of information (in creation, distribution, and access), and the decentralization of democracy. Crawford sums it up nicely:

---

323. DOUTHWAITE, *supra* note 241, at 125.
324. VON HIPPEL, *supra* note 67, at 8.
325. BENKLER, *supra* note 67, at 462.
326. Crawford, *supra* note 223, at 3.
327. *Id.* at 23.
328. *Id.* at 27, n.109. Scott Page has substantiated a similar point about the benefits of diversity: individuals with vastly different backgrounds and life experiences can yield superior outcomes versus like-minded experts. *See* PAGE, *supra* note 241.
329. Crawford, *supra* note 223, at 388 n.134.

> Treating the internet like just another proprietary, competing network that is no different from the telephone network will cause as-yet-unborn technologies, applications, collaborations, human creativity, devices, growth, economic development, and innumerable other intangible and tangible valuable and interesting things never to come into existence.[330]

D. *A new vision of wealth*

The American quest to understand prosperity was founded in Adam Smith's discussion of the wealth of nations. In his analysis, wealth comes from specialized production moderated by perfect competition. The "invisible hand" operating in the marketplace generates optimal outcomes.

From the perspective of Emergence Economics, the picture is more complex, but also more true to our contemporary reality. Physical constraints no longer need limit most forms of production, and shareable ideas motivate core growth. These ideas come from an evolutionary process and are fed back into the economic system. "Wealth is knowledge, and its origin is evolution."[331] Prescribing the exact nature of competition takes a back seat to understanding whether competitors are motivated to innovate. The interconnected nature of our real and ethereal networks multiplies the potency of these technologies, and offers new ways to work and cooperate.

The market's successful role in generating wealth does not necessarily imply that this wealth is distributed in an optimally equitable manner. Policymakers certainly can and should address concerns about wealth distribution, but should strive to do so outside the context of the market's evolutionary processes. Any programs ideally would remain consistent with the premises of Emergence Economics by harnessing, rather than impeding, market forces. Further, as we discuss later, the Internet's ability to democratize the "Long Tail" sources and distribution of innovations suggests that widespread access to the Net through more, bigger, and open broadband on-ramps may help alleviate some of these equity concerns.

Promoting wealth involves safeguarding the generative potential of these technologies, whether they are at risk from government hubris, undue market power, or other forces. Along the way, many ideas will fail, but unforeseen breakthroughs will eclipse the losses of these failed experiments. Spillovers are central. Social production is a potent model

---

330. Crawford, *supra* note 196, at 37; *see also* Susan P. Crawford, *The Ambulance, the Squad Car, and the Internet*, 21 BERKELEY TECH. L.J. 873, 931 (2006).

331. BEINHOCKER, *supra* note 22, at 450.

for innovation. The "social layer" enriches our existence in incalculable ways. Most of all, the Internet experience reminds us that wealth emerges—as if from an "invisible hand"—when diverse agents can connect and evolve.

IV. POLITICAL ECONOMY: EMERGING IMPLICATIONS FOR U.S. COMMUNICATIONS POLICY

Politics is the art of looking for trouble, finding it everywhere, diagnosing it incorrectly, and applying the wrong remedies.
Groucho Marx[332]

So after several dozen pages of intermediate economics and a smattering of Internet history and technology, where are we? Hopefully at this point the patient reader has come to a new appreciation for the complexities of agents and networks evolving together in adaptive markets, the Internet as an optimal platform for massive emergence, and the importance of innovations, economic growth, and other emergent market and non-market phenomena.

This paper will not have accomplished its primary objective, however, with a mere overview of the teachings of what we have come to call here Emergence Economics. We have something more constructive in mind. To repeat what now should be obvious, the marketplace of goods, services, technologies, and ideas does not exist in a pristine state, carved out of equations and metaphors on a university lecture hall blackboard. The market is the living, breathing incarnation of all of us. And a considerable part of that "us" is "The State."

This final section will address briefly how Emergence Economics can help us gain a new outlook on the appropriate roles of government and market in our daily affairs. We will see that these two human constructs are not polar opposites, but rather two distinct and different ways of approaching matters of considerable importance to each of us. Indeed, some argue that public policy is its own complex adaptive system, co-evolving with the economic sector.[333] By enlisting the assistance of our new learning, we will recommend some useful mechanisms for markets and states not just to coexist uneasily, but to reinforce each other's strengths, hopefully in a manner that maximizes tangible and intangible benefits for all concerned. Our aim here is not to

332. AND I QUOTE 265 (Ashton Applewhite et al., eds., rev. ed. 2003).

333. *See* Barbara A. Cherry, *The Telecommunications Economy and Regulation as Coevolving Complex Adaptive Systems: Implications for Federalism*, 59 FED. COMM. L.J. 369 (2007); *see also* Barbara Cherry & Johannes Bauer, *Adaptive Regulation: Contours of a Policy Model for the Internet Economy* (Quello Center for Telecomm. Mgmt. & Law Working Paper, 2004), http://www.quello.msu.edu/images/uploads/wp-04-05.pdf.

argue for one exclusive approach to communications policy, or grapple with specific thorny policy issues, but rather to suggest ways to shift the terms and ground of the debate so that they more faithfully reflect economic realities.[334]

Others have sought to bring to bear on the law the most recent insights of complexity science, network science, and neuroscience.[335] Here we will use the overarching topic of communications law and policy to suggest a course of possible action for lawmakers and regulators to encourage—cautiously and deliberately—the discovery and proliferation of More Better Ideas. We will suggest that this goal can be accomplished largely through market-driven policies that favor more, bigger, and open broadband network facilities leading to a generative Internet platform.

A. The Overall Role of Government

1. The Co-Evolution of Markets and Governments

Some have called economics and politics the 8th and 9th layers of the Internet.[336] While largely accurate, one misleading element of that metaphor is that the two are not separate and distinct spheres of influence. Politics and economics form the background context for each other, and for the Net itself. Each is a particular and extraordinary manifestation of evolutionarily-constrained human intelligence. Like economies, political systems—laws, edicts, regulations, principles, bully pulpits, norms, and the various agents who employ them—are a social construction, a form of human-made culture.[337] "Law itself is a self-organized emergent property of thousands of informal mores and

---

334. A more comprehensive approach to "adaptive policymaking," including devising a policy design space, will be presented in a forthcoming paper by one of the authors. *See* Richard S. Whitt, *Adaptive Policymaking: Evolving and Applying Emergent Solutions for U.S. Communications Policy,* 61 FED. COMM. L.J. No. 3 (forthcoming 2009). Here we are deliberately confining ourselves to economics—writ large via the newer schools of economic thought—as a foundational basis for public policy. We do not mean to suggest that other, more normative factors should not, and will not, also play a supporting role in the policymaking process.

335. *See, e.g.*, Michael Katz & Howard Shelanski, *Mergers and Innovation*, 74 ANTITRUST L.J. 1 (2006); Lemley & McGowan, *supra* note 174, at 479; Strandburg et al., *supra* note 82, at 1296 (Strandburg urges legal scholars to "jump on the network bandwagon in greater numbers because of the important conceptual advances and analytical tools that network science provides."). Some have even given book-length treatment to the combination of law and science. *See, e.g.*, STEVEN WINTER, A CLEARING IN THE FOREST (2001) (proposing to unite cognitive science and the law).

336. *See* Rohit Khare, *Building the Perfect Beast, Dreams of a Grand Unified Protocol*, 3 IEEE INTERNET COMPUTING 89 (1999), *available at* http://www.ics.uci.edu/~rohit/IEEE-L7-applcore.html.

337. *See* HENRY PLOTKIN, THE IMAGINED WORLD MADE REAL (2003).

restrictions that were codified over time . . . ."[338] To Crawford, "[a]ttention should be paid to the evolutionary ecosystem of the law as the background medium in which innovation occurs, business models evolve, and social factions grow and prosper."[339]

Old School Economics tends to view government as a corrupting exogenous force, an unwelcome outside influence that usually does far more harm than good. Milton Friedman famously remarked that "the government solution to a problem is usually as bad as the problem."[340] Or as Philip Ball puts it, "free-market fundamentalists argue that total noninterventionism is the best way to let the economy reach equilibrium."[341]

Emergence Economics comes at the question from a slightly different direction. Obviously, we now know that the idea of an equilibrium market, one of perfect efficiency and optimal outcomes, lacks serious foundation. More critically, in Ball's words, "it is time to recognize such claims for what they are: expressions of faith, unhindered by facts and based largely on predetermined views about the role of governments, taxation, and legislation."[342] As the history of the Internet amply demonstrates, government can and does play a constant, active, supporting role in the market, shaping the parameters of what companies and individuals can do, even if from the sidelines.[343]

From one perspective, government can be seen as a separate, yet interconnected agent, operating within the market itself. Putting aside the appealing but misguided notion of the pure "free market," the economy simply could not survive without basic laws to prop it up. Statutes and regulations concerning contracts, property, torts, securities, criminal activity, worker and consumer protection, intellectual property—these and more provide the grounding for modern day economic activity. Beinhocker has observed that "the economic evolutionary system is constructed out of a vast array of Social Technologies, many of which rely on government."[344] The state, with its

338. SHERMER, *supra* note 2, at 6.

339. Crawford, *supra* note 161, at 605.

340. MILTON FRIEDMAN, THERE'S NO SUCH THING AS A FREE LUNCH 6 (1975).

341. BALL, *supra* note 12, at 222.

342. *Id*. at 224. Evidence abounds that recent failings by U.S. financial markets stem from poor or non-existent institutions, based in part on market agents relying on half-truths from Old School Economics. *See, e.g.*, CHARLES R. MORRIS, THE TRILLION DOLLAR MELTDOWN 167 (2008) (to a great extent the financial meltdown can be attributed to the U.S. Government coddling the financial industry). One can trace many of these issues as far back as the late 19th Century, when the rise of the stock market helped engender a "speculation-based" economy, with its attendant social benefits and costs. *See* LAWRENCE E. MITCHELL, THE SPECULATION ECONOMY (2007).

343. BEINHOCKER, *supra* note 22, at 172, 472.

344. *Id*. at 425.

uniquely coercive authority, is critically important in setting the rules of the game.[345] Of course, that role has far greater legitimacy in democratic societies.[346]

This legal superstructure also serves the vital purpose of instilling trust and cooperation in strangers, so that they are willing to engage in market transactions. Without such trust, economies cannot hold. In the case of U.S. financial entities in 2008 and 2009, the result has been a frozen credit market, with banks unwilling to lend money. Modern capitalism cannot hope to flourish where legal institutions do not function properly. The law of the jungle decidedly is not the law of the free market.

So the workings of the economy rely on the government. But in turn, economics should inform the law. And of course, the question comes down to how much government is enough. A key distinction between a capitalist and socialist economy is whether the ultimate arbiter of economic fitness is a market or a hierarchy.[347] The easy assumption is that only the state can be a hierarchy, and thus improperly attempt to impose ill-fitting top-down solutions on the market. Beinhocker reminds us, however, that firms too are hierarchies, with similar cognitive constraints that can lead to flawed judgments. Regulation can be public, or private, and the impact on other agents in the market can be much the same: restraints on freedom of choice and action.

Moreover, the economy is a social process, one that does not exist apart from the rest of our lives as social beings. As such, citing behavioral psychology, Beinhocker argues that the true market perspective on human behavior is neither Left (humans are inherently altruistic) nor Right (humans are inherently self-regarding). In truth, we are actually a mix of both, or what behavioral economists call "strong reciprocity."[348] This means we are predisposed to cooperate in social situations, but also to punish group members who behave selfishly.[349] Further, the Right is

345. ORMEROD, *supra* note 19, at 96. No less a conservative authority than Milton Friedman observes, "[t]he existence of a free market does not of course eliminate the need for government. On the contrary, government is essential both as a forum for determining the 'rule of the game' and as an umpire to interpret and enforce the rules decided on." MILTON FREEDMAN, CAPITALISM AND FREEDOM 15 (Chicago University Press 1980) (1960).

346. In addition to serving as "market rule-setter and referee," the state also can provide goods and services that markets otherwise would undersupply." MCMILLAN, *supra* note 241, at 149.

347. BEINHOCKER, *supra* note 22 , at 422.

348. *Id*. at 419.

349. As one set of researchers puts it: "The behavioral sciences have traditionally offered two contrasting explanations of cooperation. One, favored by sociologists and anthropologists, considers the willingness to subordinate self-interest to the needs of the social group to be part of human nature. Another, favored by economists and biologists, treats cooperation as the result of the interaction of selfish agents maximizing their long-term individual material interests . . . . [We show that] a significant fraction of people fit neither of these stereotypes.

correct that the economy is too complex for central planning to work effectively. Hayek in particular pointed out that policymakers have knowledge coordination problems, possess no perfect rationality, and utilize no good market feedback mechanism.[350] On the other hand, the Left also is correct that markets, while useful and necessary, are not optimally efficient. The question is not states versus markets, Beinhocker explains, but "how to combine states *and* markets to create an effective evolutionary system."[351]

Surprisingly (at least to some), Hayek endorses an active role for government—for some purposes:

> It is important not to confuse opposition against . . . [central] planning with a dogmatic laissez faire attitude. The liberal argument is in favor of making the best possible use of the forces of competition as a means of coordinating human efforts, not an argument for leaving things just as they are . . . . To create conditions in which competition will be as effective as possible, to supplement it where it cannot be made effective, to provide the services which, in the words of Adam Smith, "though they may be in the highest degree advantageous to a great society, are, however, of such a nature, that the profit could never repay the expense to any individual or small number of individuals"—these tasks provide, indeed, a wide and unquestioned field for state activity. In no system that could be rationally defended would the state just do nothing.[352]

Put differently, the question is not whether government necessarily is part of the market, but what that role should be. Instead, we should take "a pragmatic approach to the market, against the quasi-religious views that it is always right or fundamentally evil . . . . Markets are not magic, nor are they immoral."[353] If we no longer assume (as Old School Economics does) that markets invariably converge on optimal efficiency, there is reason to believe that government intervention may in some instances be beneficial. Policymakers can have a role in facilitating positive outcomes from this ecosystem, but this role should be carved out carefully, guided in part by the various schools subsumed within Emergence Economics.

Rather, they are conditional cooperators and altruistic punishers . . . which we call strong reciprocators." HERBERT GINTIS ET AL., MORAL SENTIMENTS AND MATERIAL INTERESTS xi (Herbert Gintis et al. eds., 2005). It is thought that gene-culture coevolution resulted in strong reciprocity in human beings. *Id*. at 27-28.

350. *See* HAYEK, *supra* note 101.

351. BEINHOCKER, *supra* note 22, at 427 (emphasis in original).

352. F. A. HAYEK, THE ROAD TO SERFDOM 85, 88 (Bruce J. Caldwell ed., Univ. of Chicago Press 2007) (1944).

353. MCMILLAN, *supra* note 241, at 226.

2. The Policymaker As Adaptive Economic Agent

Easterly correctly observes that "[g]overnment is not a single, all-knowing actor. Government instead is a coalition of politicians representing different factions."[354] The policymaker, whether an individual or an agency, a member of Congress or the Federal Communications Commission (FCC), acts as a representative of government. They are authoritative agents in the economic system, as well as their own complex adaptive systems.[355] As such, the policymaker needs to overcome the typical analytical flaws of economic agents, and try to see the market with fresh eyes.

In fact, as just another agent operating within the economic system, the policymaker possesses all of the cognitive constraints—and adaptive flexibility—of any other agent. True, policymakers in modern democracies are answerable to political hierarchies (if elected, their constituents; if selected, both the elected and their constituents). Further, policymakers have a unique power: the unilateral coercive authority that comes from the state. Whether a policymaker acts or refrains from acting has repercussions in the world. So the policymaker invariably has an impact on the surrounding ecosystem of the market.

The human desire to predict and control runs deep. And planning inherently is a significant and unavoidable part of a policymaker's job. Yet short-term prediction and control of the economy is inherently impossible, due largely to our data gathering and processing shortcomings as agents, and the inherent complexity and unpredictable movements of markets.[356] The pitfalls of central planning of markets are fundamentally problems of information.[357] Much economic policy in the West has been and remains conducted on the basic of short-term forecasts of the economy. "Politicians have sought to change the world. But the point is to interpret it correctly."[358]

The vigor of markets comes from their decentralized nature; they empower people to find creative solutions to problems. Government laws and policies inevitably help shape part of the fitness environment within which companies compete and other agents make their choices.[359] The crucial question, then, is how those government actions affect that fitness

354. EASTERLY, *supra* note 52, at 258.

355. Geoffrey Hodgson has written extensively about the evolution of political institutions as the stuff of social life. *See, e.g.*, GEOFFREY HODGSON, THE EUROPEAN ASS'N FOR EVOLUTIONARY POLITICAL ECON., EVOLUTION OF ECONOMIC INSTITUTIONS (2007).

356. ORMEROD, *supra* note 19, at 75-90; *see also* TALEB, *supra* note 67, at 180 (corporations and governments overestimate their ability to understand the subtle changes that constitute the world).

357. MCMILLAN, *supra* note 241, at 149.

358. ORMEROD, *supra* note 19, at 182.

359. *See* BEINHOCKER, *supra* note 22 , at 425.

environment. Given policymakers' all-too-apparent constraints, the complexity of the market itself, and the effective workings of the market in providing growth and other emergent benefits, considerable caution is warranted.

Ideally, then, U.S. policymakers will not attempt to intervene in the day-to-day processes of the American marketplace. This means, in brief, that government generally should leave to market mechanisms the workings of the evolutionary algorithm: agents differentiating, selecting, and amplifying specific Physical Technologies, Social Technologies, and Business Plans. But this does not mean that policymakers need remain on the sidelines. Where policymakers have identified important public policy goals and objectives, the key is to employ market forces, as much as feasible, to achieve those desired ends. In other words, policymakers should "tinker" with the fitness landscape in ways that can bolster, and not hinder, evolutionary processes. In general, then, government should do less, not more—but less still can become more, if done better.[360]

B. A Communications Policy Approach For Innovation And Growth

Our new economic and technology foundations necessarily implicate significant changes to our public policy thinking. We need a new approach to our nation's communications policy, one rooted in Emergence Economics and its useful lessons.

We next will sketch out one approach to communications policy that should support greater levels of innovation and growth. This treatment necessarily will be brief at this point, and only provides some suggestions on ways to use Emergence Economics as a guiding instrument. It is our contention that policymakers should have as their ultimate aim to foster an ecosystem in the communications sector that imparts greater economic and non-economic benefits for all agents—producers and consumers, policymakers and citizens. Market forces—defined broadly as the sum total of human productive activities serving a range of pecuniary, social, and personal purposes—remain the most effective mechanism for those benefits to fully emerge.

1. Why Communications Policy?

Why focus on communications policy? Because first and foremost, human communications matter. Our species can only survive and flourish when our power of communication is fostered. Modern technologies have enabled us to build powerful shared platforms where all variants of person-to-person interaction are possible. As Mokyr has demonstrated,

360. ORMEROD, *supra* note 19, at 96, 184.

widespread access to such platforms greatly aids in the dispersion of useful knowledge.[361] Here are several different yet related ways that the concept has been described:

> Information and communications are core elements of autonomy and of public political discourse and decision making. Communication is the basic unit of social existence. Culture and knowledge, broadly conceived, form the basic frame of reference through which we come to understand ourselves and others in the world . . . . The basic components of human development also depend on how we produce information and innovation, and how we disseminate its implementations.[362]

> The complexity of human interactions has been fostered throughout the ages by communications technology, which facilitates the exchange of information on all levels, from individuals to governments. The more information is exchanged, the more feedback processes occur and thus, in general, the more complexity. Computer networks are now transforming the nature and speed of such communication, and the sheer volume of accessible information.[363]

> The parameters of the current Information Age become clear when we understand the information revolution not only as a major sociocultural change but also as something like an orbital movement in which information revolves in such a way that it begins to act on itself. The information revolution occurs when information turns on itself and becomes self-reflexive. This turn has been made possible by new electronic and telematic technologies, through which information acts on information to form feedback loops that generate increasing complexity. This is why the information revolution issues in the moment of complexity.[364]

As we have seen, the Internet so far has been an optimal platform for generating new ideas and innovation, economic growth, and other Net effects. Government policies inevitably affecting the Net specifically, and the communications sector more generally, have a profound impact, for good or ill, on the national economy. In short, communications policy should be seen potentially as a major lever, whether upward or downward, for economic development and growth.

---

361. MOKYR, *supra* note 236, at 290-91.
362. BENKLER, *supra* note 67, at 464.
363. PETER COVENEY & ROGER HIGHFIELD, FRONTIERS OF COMPLEXITY 338 (1995).
364. TAYLOR, *supra* note 115, at 106.

a. *The Compelling Need To Rethink Our Priorities and Approaches*

In conjunction with the critical role of communications in economic and non-economic human endeavors, there is a real need to correct decades of flawed thinking that underpins what passes for communications policy in this country. The forces of Old School Economics have found fertile ground in the communications field.

For the most part, incumbent actors and industries have embraced Old School Economics as a basis to argue for less economic regulation. They claim that the "free market" should have primacy, and "perfect competition" produces the optimal results in the public interest.[365] They also argue that a relatively modest version of the nation's antitrust laws offer the only way to deal with competition/market power concerns.[366] Yet policy opponents of the incumbents tend to argue from the other side of the same coin: that government regulation inevitably is the best response to deal with economic or social concerns. For some of these players, market failure is endemic, and governments are best equipped to rectify the market's many perceived failings.[367] In short, both camps see the market as a coldly efficient machine, one side with approval, the other with approbation. As Julie Nelson puts it, the mechanical metaphor of the market can lead to "naïve and irresponsible neoliberal probusiness policies" versus "naïve and impractical antimarket alternatives."[368] Unfortunately neither those on the self-proclaimed "Right" or "Left" appear to realize that in many cases they are operating

365. For example, the Cato Institute cites its support for traditional America "principles of limited government, free markets, individual liberty, and peace." Cato's Mission, www.cato.org/about.php. As we have seen these concepts are not entirely self-evident, or at least well defined. Of course it is not the ultimate aims, but how they are to be achieved, that deserves the closest scrutiny.

366. As just one example, a recent white paper argues that there are almost no forms of "bundling" and tie-in sales that raise anticompetitive concerns in technology markets, because they invariably create efficiency and do not foreclose competition. Stan Liebowitz & and Stephen Margolis, *Bundles of Joy: The Ubiquity and Efficiency of Bundles in New Technology Markets*, PERSPECTIVES FROM FSF SCHOLARS, Jan. 24, 2008, http://www.freestatefoundation.org/images/Bundles_of_Joy.pdf, at 2. While we acknowledge that bundling often can be pro-consumer and pro-competition, we think it goes too far to claim that to be the case in nearly all instances, or that "a product achieves a degree of market power… wherever an innovation succeeds." *Id*. at 46. As we explain above, successful innovations do tend to create market power, but certainly do not constitute the only source.

367. For example, Free Press states that the "broken" media system "isn't natural," and that media should be compelled to "serve the public interest" by being "vibrant, diverse and independent." Free Press and the Free Press Action Fund, http://www.freepress.net/node/121. Again, whether one agrees or not with those goals, it is not obvious that reliance on regulatory fiat alone is the optimal means for fulfilling them.

368. JULIE A. NELSON, ECONOMICS FOR HUMANS 53, 57 (2006). Nelson suggests as an alternative metaphor the economy as a beating human heart, connoting a living, vital organ. *Id*. at 59.

from false premises. Without a deeper and richer appreciation for economic realities, though, it is impossible to discern whether and how any of these viewpoints should be given credence.

The Federal Communications Commission apparently can do little in this environment but follow the prevailing economic notions. In one order after another, the agency tends to parrot the stated views of the dominant players, on both sides of an issue, and couches its policies in the vernacular of Old School Economics. One recent prominent example is the FCC's 2005 decision deregulating broadband services provided by the incumbent local exchange carriers (ILECs).[369] Here the FCC defined the ILECs' combined Internet access/broadband transmission services as a unitary information service, and thus outside traditional common carriage regulation such as the *Computer Inquiry* nondiscriminatory access safeguards. That *Wireline Broadband Order* exemplifies many of the flaws of relying on traditional economic thinking, at least in this case as articulated by the incumbent LECs and their allies.

The sole question the Commission saw fit to ask and answer in that order is whether the costs of the *Computer Inquiry* regulations outweigh their benefits to the broadband providers. The analysis focuses tightly on a traditional analysis of costs and benefits, and only of the broadband providers themselves—not the Internet, or its users.[370] Tellingly, nowhere in the 86-page order does the Commission discuss broadband as a platform to the Internet, or any potential impact on the generative Internet itself. Indeed, aside from briefly discussing and dismissing concerns raised by independent Internet service providers, the order rarely utilizes the word "Internet."

The FCC also adopts easy assumptions about the state of the broadband market, without recourse to record evidence, save citations to filings by broadband providers themselves.[371] For example, the FCC claims that the then-current broadband market is competitive, and growing far more competitive with expected imminent entry by providers of fixed and mobile wireless services, satellite services, and broadband over powerline (BPL) services.[372] To those who point to persistent market concentration between the cable companies and telephone

369. Appropriate Framework for Broadband Access to the Internet Over Wireline Facilities, *Report & Order & Notice of Proposed Rulemaking*, 20 FCC Rcd. 14,853 (2005) [*hereinafter Wireline Broadband Order*].

370. *Id.* ¶¶ 43, 65-69.

371. Notably the order is littered with phrases including words like "expect" and "anticipate" and "predict." The agency at one point even admits that much of its analysis is based on "what our predictive judgment tells us about how [the broadband] market is likely to develop." *Id.* ¶ 43. Unfortunately, the conclusions are rendered with far more certainty, and finality, than this couched language otherwise would warrant.

372. *Id.* ¶¶ 33, 50.

companies, the FCC claims that such arguments are premised only on "snapshot data" that are "both limited and static," as compared to "larger trends" in "the dynamic nature of the marketplace forces."[373] Of course, while those phrases are couched in the language of Emergence Economics, such duopoly "snapshots" endure to this day.[374]

Amazingly, the FCC casually dismisses one of its more singular achievements of the late 20th Century. "The *Computer Inquiry* rules themselves reflect a fairly static picture of network development, and an assumption that a line could be drawn between the network functions and computer processing without impeding technological innovation."[375] The counter-assertion that such line drawing resulted, not in an arguable reduction of broadband network innovation, but an explosion of online innovation, leading to the Internet itself, seems never to have been seriously contemplated, although it was plainly presented.[376]

The Commission further "expect[s] that facilities-based wireline carriers will have business reasons to continue making broadband Internet access transmission services available to ISPs without regard to the *Computer Inquiry* requirements."[377] The implicit assumption is that the broadband "market" and its incentives system would function properly, allowing for mutually-satisfactory agreements between broadband providers and independent ISPs. This expectation is reached

---

373. *Id*. ¶ 50.

374. Robert D. Atkinson, *Framing a National Broadband Policy*, 16 COMMLAW CONSPECTUS 145, 175-76 (2007) [*hereinafter* Atkinson, *Framing*]. *See generally* Robert D. Atkinson, *The Role of Competition in a National Broadband Policy*, 7 J. ON TELECOMM. & HIGH TECH. L. 1 (2009) [*hereinafter* Atkinson, *Competition*]. The Commission also cites the incumbent LECs themselves for the self-serving proposition that "the additional costs of an access mandate diminish a carrier's incentive and ability to invest in and deploy broadband infrastructure investment." *Wireline Broadband Order*, *supra* note 369, ¶ 44. Absent the ISP access rules, the FCC posits, broadband providers could "produce new or improved services in response to consumer demand." *Id*. ¶ 71. Even if these arguments have merit—and current evidence is spotty at best—strong counter-claims about stifling independent ISP-based innovations were not afforded similar weight.

375. *Id*. ¶ 70.

376. As one example, the BroadNet Alliance, a coalition of national, regional, and local independent ISPs, submitted pleadings in the FCC's docket explaining how the FCC's ISP-related policies have played a pivotal role in enabling the Internet. Reply Comments of the BroadNet Alliance in Appropriate Framework for Broadband Access to the Internet Over Wireline Facilities, CC Dkt. No. 02-33 (July 1, 2002). In particular, the coalition filed a detailed white paper showing that the *Computer Inquiry* rules "in large part enabled the rise and amazing success of the online world," by creating conditions that allowed consumers to reach the online providers of their choice. *Id.* at 2. As a later ex parte explained, "should this critical access to those facilities no longer be made available to ISPs under the Computer Rules, the only remaining choice for broadband Internet access will be the incumbent's ISP. . . ." Ex Parte Filing of the BroadNet Alliance in Appropriate Framework for Broadband Access to the Internet Over Wireline Facilities, CC Dkt. No. 02-33 (July 24, 2002). That prediction proved prescient.

377. *Wireline Broadband Order supra* note 369, ¶ 64.

despite the fact that "we cannot state unequivocally that incumbent LECs would not otherwise provide wholesale access, absent this compulsion."[378] Nonetheless, despite this lack of confidence, and little record evidence of current or expected market competition to generate the necessary economic incentives, the FCC concludes that "the public interest is best served if we permit competitive marketplace conditions to guide the evolution of broadband Internet access service."[379] The Commissioners felt obliged to add, "this does not mean that we sacrifice competitive ISP choice for greater deployment of broadband facilities."[380] But that is precisely the high-stakes calculus the agency utilized here. The sad reality is that the independent ISP industry all but disappeared in the wake of the FCC's decision, to the point where policymakers and industry players today refer to the integrated broadband/Internet access providers as "ISPs"—effectively acknowledging the reality that there are no others.[381] Causality is always difficult to assign, but the FCC's decision must have played at least some pernicious role.

So the FCC—an independent regulatory body charged with holding industry expertise and operating in the public interest—renders decisions that appear to display little appreciation for the ways that actual markets function, and sometimes fail to function. In the case of the *Wireline Broadband Order*, the agency in particular: (1) relied on arguments and evidence largely from interested party agents; (2) kept a tight focus on the costs and benefits to a single set of agents, the incumbent broadband providers; (3) conversely failed to factor in the potential impact on the Internet as a generative platform; (4) recognized the broadband carriers' supposed market incentives to invest in broadband networks, without appreciating their more obvious incentives not to strike commercially-viable deals with competing ISPs; (5) rejected without a more searching analysis its own *Computer Inquiry* precedent drawing lines between Internet access and broadband networks;[382] (6) acted through revamped statutory definitions, rather than more flexible deregulatory tools such as forbearance;[383] and (7) generally showed a

378. *Id.* ¶ 63.

379. *Id.* ¶ 85.

380. *Id.* ¶ 79.

381. As one commentator sums it, "connectivity has been vertically integrated." MARTIN FRANSMAN, THE NEW ICT ECOSYSTEM 31 (2007).

382. For example, if in fact certain computer processing is so tightly wedded to underlying communications network functionality, as the Commission claims, why would facilities-based Internet access not be defined as a telecommunications service? The agency still could deal separately with the regulatory implications through more tailored statutory tools, such as the forbearance provision of the Communications Act. *See* 47 U.S.C. § 10 (2000).

383. Among other advantages, the forbearance approach would have offered the agency a more empirically-based, provisional, and reversible statutory tool. Instead, one of the many ironies of the FCC's ostensibly deregulatory decision is that, by defining its way out of

decided lack of skepticism about its own predictive judgment. Moreover, the Commission added to its failings by not allowing for any post-decision accountability, which would allow the agency to revisit the decision's factual support after a certain period of time. Such a revisit could include, for example, asking whether broadband deployment and competition were developing as anticipated, and whether independent ISPs were able to strike adequate wholesale deals to provide competing ISP services to consumers.

As we transition from existing communications-related industries to a new world ruled by IP-based networks and applications, we should look towards a different approach. The FCC, for one, needs a framework and some conceptual tools that do a better job of respecting the economic world as actually lived.

b. Caution Ahead

Generally speaking, we have more to lose than to gain from direct government involvement in Internet-based markets, especially if such intervention is based on misunderstanding or even ignorance of market dynamics. As Ormerod observes, imperfect markets plus imperfect regulators equal a strong dose of caution.[384]

In particular, government clearly can impede innovation. Technological creativity has proven to be politically vulnerable; "the history of technological progress is the history of an endangered and much-resisted species."[385] Mokyr finds that in centralized bureaucracies, whether governmental or corporate, "there is a built-in tendency for conservatism" and resisting innovation.[386] But there is more to it than that:

> The political economy of technological change thus predicts that it will be resisted by well-organized lobbies, whereas its defenders will usually be a motley group of consumers and inventors and perhaps a few groups with a direct interest in economic growth. The struggle between the two parties will always take the form of a non-market process, because relying on market forces alone would, by definition, lead to the triumph of the new technology.[387]

---

common carriage treatment of broadband networks under Title II, the agency instead has opened wide the door to equally damaging—and far less bounded—regulation of Internet-based services, applications, and content under Title I. This disturbing trend currently is playing itself out in the VoIP arena.

384. ORMEROD, *supra* note 6, at 138-39.

385. MOKYR, *supra* note 236, at 223-34.

386. *Id.* at 238.

387. *Id.* at 253-54.

Putting aside this "purposeful self-interested resistance to new technology,"[388] by the same token we do not intend to imply that the Internet inevitably is the only source of innovation or growth, or that public policy should be skewed relentlessly in that direction. Instead, we are offering a needed corrective to the static thinking of the recent past, largely informed by outdated if not fatally flawed economic theory. Similarly, government policymakers need not approach the Internet as some precious museum piece, which should be forever fixed in the same configuration. The Net in many ways is a living thing, a constantly changing process reflecting countless human choices. It would be as much a mistake for a government official to tamper with the evolutionary algorithm so as to attempt to preserve the Net as it is, as it would have been to prevent its original creation and launch in the name of preserving the original commercial online companies.

At the same time, because the Net has brought such amazing benefits, and promises so much more, there is real value in trying to retain certain core elements against counter forces. Werbach has nicely summed up the challenge: to engage in a balancing act between the centripetal and centrifugal forces that shape the Net. In our view, this balancing act must begin with a presumption (but only that) that interconnected, end-to-end, layered networks like the Internet provide real economic and "spillovers" value. Beyond that, policymakers must be vigilant, open to new ideas, and flexible in devising and implementing policy measures.[389]

An important distinction to keep in mind is the means versus the ends. As we shall see, the ultimate end goal of more good ideas, and the follow-on objectives of many more open and big broadband pipes, may follow reasonably, but the means of achieving those objectives are not so obvious. In particular, the means are influenced by a healthy skepticism that legislators and regulators will get the formula right. That skepticism should be tempered by an understanding of agent limitations—users and firms alike—and an often-opaque marketplace.

2. Defining Our Mission: Goals and Objectives

In sketching out a game plan for revamping U.S. communications policy, first we will need to distinguish between different elements of an

388. *Id.* at 220.

389. The Internet community has developed many successful mechanisms for evolving itself through iterative self-governance. One remarkable example is the Internet Engineering Task Force (IETF), a voluntary group that collaboratively has defined the ever-evolving core protocols for decades. The open, shared standards that come out of this group reflect the diverse constituencies involved, but to date have preserved agents' ability in the network to connect in a transparent fashion.

overarching policy design space. One way to understand this framework is to break it down into its component parts, which include goals, objectives, projects, and tools.[390] The goals are the largest, longest term elements to be accomplished (for example, landing on Mars). The objectives are the intermediate term aims (building and testing a rocket ship to send to Mars). The projects are the specific, short-term aims (devising elements of the engine that will power the rocket), while the tools are the practical mechanisms utilized for achieving all of the above (computer programs that model different components of the rocket ship). The organizational and institutional elements of the design space (the policy players and the policy rules, respectively) are important as well. Consistent with our discussion in previous sections, the chief aim is to be bold about the vision of goals and objectives, while more modest yet flexible about the particular programs and tools used to accomplish them.

a. One Goal: More Good Ideas

At this point, we trust that the turn in the discussion will appear almost self-evident. As we have seen, ideas are the fodder, the raw material, for economic growth and other beneficial Net effects. New technologies—products, processes, and forms of organization—are the most important determinant of long-term economic growth.[391] The free flow of information between and among people can lead directly to a raft of BPs, PTs, and STs competing vigorously and effectively in the marketplace, along with every form of information, entertainment, political discourse, and commercial and non-commercial speech. One overarching goal for policymakers, especially in the communications field, should be to see the market generate a greater number of useful ideas so as to drive the evolutionary process to optimal heights. Romer calls for a "combinatorial explosion" of ideas.[392] By furthering an increased quantity of beneficial new ideas, more potential innovation is enabled.

Crawford for one appears to agree with this goal of More Good Ideas. She recently argued that a key organizing principle for communications law should be to support the emergence of diverse new ideas online.[393] Crawford interprets this as "allowing the end-to-end,

390. In a subsequent paper, one of the authors will further explore and expand upon this "policy design space" framework in the context of federal communications policies. Whitt, *supra* note 334.

391. LIPSEY ET AL., *supra* note 125, at 11.

392. Kevin Kelly, Paul Romer: The Economics of Ideas, http://www.versaggi.net/ecommerce/articles/romer-econideas.htm.

393. Crawford, *supra* note 223, at 35.

content-neutral, layer-independent functions of the internet to flourish and allowing groups and human attention to pick and choose from among the bad ideas presented online, enabling good ideas to persist and replicate."[394]

Of course, what is "good" or "bad" should not be for any to decide unilaterally for anyone else. The market's role, through each and all of us, is to churn through the various options—BPs, PTs, STs, and all other instantiations of useful knowledge—and select what is most fit. The "Long Tail" of the Internet suggests that fitness landscapes can be enabled for a much wider array of options than otherwise has been available previously to market agents. Even a "failed" idea for most agents in the market increasingly can manage to succeed with at least some of us in the deeper niches.

b. One Objective: Harnessing Broadband As An Optimal Internet Platform

With the ultimate end goal in mind—More Good Ideas—we next need a more near-term set of policy objectives that will help us achieve that goal. While obviously there are a number of salient possibilities in the communications sector, we select by way of example the objective of harnessing broadband networks to serve as an optimal platform for the public Internet. Here we will touch on three interrelated components of such a policy objective: Open Platforms, More Platforms, and Bigger Platforms. In each case, the focus is on broadband providing enhanced access to the generative Internet (as opposed to other uses, such as private networks broadcasting proprietary content). Our discussion here necessarily will be brief, and is intended for illustrative purposes.[395]

i. Open Broadband Platforms

First, we should want to promote "open" platforms leading to the Internet, capable of adaptive power by the myriad of end users interacting, innovating, and competing with each other. As we have seen, the Net is not just an e-commerce platform, but also a means of distributing and validating ideas, and other aspects of human communications. The Long Tail, among other things, also helps extend economic growth beyond the "winner take all" mentality, to numerous niches served by smaller players (who also have a chance to become big

394. *Id.* at 35-36.

395. Author Whitt has produced a companion paper that focuses exclusively on U.S. broadband policy; this section necessarily provides only a modified and truncated portion of that work. Richard S. Whitt, *Evolving Broadband Policy: Taking Adaptive Stances to Foster Optimal Internet Platforms*, 17 COMMLAW CONSPECTUS (forthcoming 2009).

players). An enormous gradation of ideas, whether fit to all, or some, or a few, can exist on the Net.

So how do we define openness? There are a variety of ways to analyze the question. As just one example, Jonathan Sallet has written about the various ways of thinking about open networks.[396] He observes that openness can vary based on different perspectives on the network—content and applications accessed by end users, for example, versus the network and ISP connectivity utilized by competing network providers.[397] As Sallet notes, the Bell System traditionally was completely open at the content layers, because any end user could communicate with any other end user, but almost completely closed at the network layers, because there was no right to attach terminal equipment or interconnect competing networks.[398] Kevin Werbach reminds us to think about openness at, and between, the physical interfaces, where the network meets content (such as technical standards for modems), and logical interfaces, where the content moves through the network (such as unique identifiers and routing databases).[399] For now it is useful to note that "openness" can occur at different interfaces within the broadband network, between and among elements of the physical layer, the logical layer, and the applications and content layers.

The Internet itself provides important clues about the degrees of openness for the on-ramps that serve our larger goal of more good ideas.[400] The combination of layering, network connectivity, IP as an agnostic

396. Jonathan Sallet, *supra* note 204, at 3.

397. *Id.* at 6-7.

398. Notably the FCC's "Internet Policy Statement" principles focus only on one end of the broadband connection—the consumer—and all but ignore the other end—the providers of applications, content, and devices. The Statement indicates that "consumers are entitled to access the lawful Internet content of their choice, . . . run applications and use services of their choice, . . . [and] connect their choice of legal devices that do not harm the network . . . ." Appropriate Framework for Broadband Access to the Internet Over Wireline Facilities, *Policy Statement*, 20 FCC Rcd. 14,986 (2005), *available at* http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-05-151A1.pdf. This approach overlooks the fact that, if broadband providers take certain actions that end up reducing the supply of such applications and content, the consumer will suffer, even if his or her own "right to access" remains untouched. Given the critical importance of innovation and competition from the edge of the Internet, we should want both ends of the broadband pipe to be open.

399. Kevin Werbach, *Breaking the Ice: Rethinking Telecommunications Law for the Digital Age*, 4 J. ON TELECOMM. & HIGH TECH. L. 59, 80-95 (2005).

400. In a sense, it is true that the Internet today is not an absolutely "neutral" place, in that the various servers, routers, and content delivery networks that comprise it in part can and do distinguish between various forms of traffic. The whole point of the Internet is that it is a robust freewheeling marketplace of ideas and commerce and interaction. We have less concern about the "neutral" state of the Net's architecture than local broadband networks because, among other differences, commercial arrangements between Net players (a) do not otherwise deliberately block or degrade other parties' access to the Net, and (b) are struck today in a comparatively competitive environment, with no single decision-maker able to impose its will on others.

"bearer" protocol, and of course the e2e principle, have allowed end users to utilize the Net as a ubiquitous platform for their activities. The mantra of "innovation without permission" also helps clarify what we mean here. At bottom, we should want users to operate freely, and not be required to secure before-the-fact approval from network operators for their lawful activities. Given the enormous positive values springing from the Internet, then, the government should see as one of its objectives adopting broadband policies that do not interfere with—if not enhance—the salient features of the Net. In particular, agnostic protocols, interconnectivity, and e2e functionality are positive network characteristics that should be preserved and promoted for the Internet's on-ramps.[401]

What we are touting here is an overall environment of openness on broadband platforms, and not necessarily a prescription. How one achieves an open network can be very different from compelling it as a legal or regulatory mandate. A related point comes from the debates over "network neutrality," which inaptly is referred to by too many proponents and opponents alike as a regulation-first approach to preserving openness, rather than (in the authors' view) as the desired outcome.[402] To argue for open platforms to the Net as an objective is not to suggest any particular approach to making it happen. Further, it is not all good out there on the Net. As a faithful reflection of humanity, the Net has its share of unseemly conduct and criminality. To be open to the best does not imply to be open to all.[403] Still, going forward we should remain "open" to other ways of thinking about openness.

ii. More and Bigger Broadband Platforms

Another major national policy objective to support More Good Ideas should be more and bigger broadband networks. The broadband-enabled Internet is rapidly changing the world in countless beneficial

401. Elsewhere author Whitt has discussed optimal Internet access as including the twin "openness" dimensions of sufficient capacity for, and overall integrity of, Internet access on broadband networks. Whitt, *supra* note 395.

402. The current debate over network neutrality is focused only on last-mile broadband connectivity, where the relative lack of competition renders the threat of unilateral gatekeeping more significant and tangible. Thus, it is a misnomer to refer to "net" neutrality, as if the Internet itself is supposed to be completely neutral to all traffic. Of course, it serves the rhetorical objectives of some to criticize various broadband openness proposals as amounting to "regulation of the Internet." We are really talking about the end points serving the consumer, meaning something more like "broadband neutrality" or "broadband openness"—phrases which at this point we doubt will catch on in the current political "marketplace of ideas."

403. Jonathan Zittrain has discussed how open systems like the Internet can be prone to abuse, and argues convincingly for a strategy that "blunts the worst aspects of today's popular generative Internet and PC without killing those platforms' openness to innovation." Zittrain, *supra* note 297, at 150.

ways. From a purely economic perspective, broadband connectivity is becoming a catalyst for innovation, productivity growth, job creation, and global competitiveness. In addition to enabling far richer uses of the Internet, broadband is an innovation platform of its own, and a source of network-based innovations. Broadband also serves as a platform to educational opportunity and civic discourse.

At the same time, there are major challenges both in terms of the reach and the depth of today's broadband offerings. Put simply, we need bigger broadband pipes, and more of them. In the United States, we have relied on a policy that leaves broadband infrastructure largely in the hands of private entities. In conjunction with such private investment, however, many still see a salient role for public policy. Robert Atkinson for one offers an approach to fashioning "a national broadband policy," which he separates into (1) broadband everywhere (providing separate incentives for rural deployment), (2) broadband for everyone (developing digital literacy and broadband applications), (3) greater speeds, and (4) more competition.[404]

Broadband's externalities present a unique challenge to any objective of fostering More and Bigger Platforms. We have every reason to want to give broadband providers the proper economic incentives to further invest in their networks, and to incent new entrants where the economics make sense. By the same token, as we have seen, communications infrastructure typically generates large social benefits not captured by the infrastructure provider. For broadband networks, "because effects" (money made because of something) are greater than "with effects" (money made from selling that something).[405] Atkinson sees four kinds of positive externalities (what we have called innovation spillovers) attributable to broadband networks: (1) network externalities (network effects), both direct and indirect; (2) prosumer investments (consumers become both users and producers); (3) competitiveness externalities (international leadership in technology); and (4) regional externalities (particularly impacts on rural communities).[406] He points out that broadband is unique in that "the social returns of broadband investment exceed the private returns to companies and consumers . . . . [T]here is considerable reason to believe that there are significant externalities from high-speed broadband, and that if left to themselves, market forces alone will lead to less investment in broadband than is societally optimal."[407]

Broadband providers seek to capture (internalize) those externalities by serving as an Internet platform—meaning they want to gain at least

404. Atkinson, *Framing*, *supra* note 374, at 164.
405. Frischmann & Lemley, *supra* note 303, at 102-05.
406. Atkinson, *Framing*, *supra* note 374, at 153-64.
407. *Id.* at 145, 154.

some of the "because effects" revenues and profits. For many this incentives system is the root cause of the "network neutrality" policy debates, which we will not address in any depth here.[408] However, the larger point is that these positive externalities do not appear to register in the incentives structure of the broadband providers. Thus, "[t]he profit/loss statements of individual firms fail to take into account the positive externalities from a widely deployed broadband network, including economic growth, lower-cost health care, and higher quality education."[409] Where these positive externalities exist, so that private investment generates total social value in excess of individual firm value, it may be appropriate for the government to get involved.

As we will argue in the next section, government involvement in contestable markets generally should be limited to a form of "tinkering" with the fitness landscape, through a mix of additional inputs, connectivity, incentives, and feedback. With regard to the need for additional competition in the broadband market, for example, it is not clear that the government has a major prescriptive role, save to remove any regulatory hurdles and get out of the way.[410] While it is one thing to take away impediments that prevent new broadband platforms from emerging in the fitness environment, it is quite another to seek to compel private parties to uptake a specific technology where the economics normally would not support such a result. Atkinson warns that the role of government should not be proactively to compel or subsidize the deployment of additional broadband networks, largely because it is not clear that otherwise it makes sense economically for a "third pipe" competitor to enter the consumer broadband market.[411] Given the complexities and uncertainties of the market, policymakers would be wise to heed that warning.

408. We will only note that Joseph Farrell and Philip Weiser describe how broadband providers might tend to internalize "complementary" externalities ("ICE"), and argue that such private internalization mitigates competitive and consumer harm. Joseph Farrell & Philip J. Weiser, *Modularity, Vertical Integration, and Open Access Policies: Towards a Convergence of Antitrust & Regulation in the Internet Age*, 17 HARVARD J. L. & TECH. 85 (2003). At the same time, they identify a variety of exceptions to this tendency. *Id*. Even where these exceptions do not apply, however, the nature of innovation spillovers would argue that overall social welfare in this domain rarely is served by private firm internalization alone. Frischmann & Lemley, *supra* note 303, at 135-39. Furthermore, as Barbara van Schewick has explained, broadband providers retain certain economic incentives to disadvantage non-affiliated application innovators despite the ICE principles. Barbara van Schewick,. *Towards an Economic Framework for Network Neutrality Regulation*, 5 J. ON TELECOMM. & HIGH TECH. L. 329 (2007).

409. JOHN WINDHAUSEN, A BLUEPRINT FOR BIG BROADBAND 5 (2008), http://net.educause.edu/ir/library/pdf/EPO0801.pdf.

410. Atkinson, *Competition*, *supra* note 374, at 15. In Atkinson's words, "[e]nable, but don't promote." *Id.*

411. *Id.*

C. *Applying the Adaptive Policy Framework: Enabling without dictating*

So what specific lessons can Emergence Economics impart to policymakers, particularly those involved with the information and communications technology sector? One suggestion is to adopt a conceptual framework that separates out those market activities that should not be susceptible to the employment of policy tools, from those that should. We believe in brief that policymakers generally should endeavor—at most—to foster the market's processes, rather than interfere with or even attempt to replace those processes. As we will explain in the sections to follow, this "adaptive policymaking" dichotomy would still allow certain "tinkering" with the fitness environment—providing useful inputs, connectivity, incentives, and feedback. However, the basic workings of the evolutionary algorithm—agents differentiating, selecting, and amplifying various technologies and business plans—should be left to the effectiveness, merits, and complexity of the open market.

Admittedly the line drawing here between harmful dictating and beneficial enabling is not yet a rigorously grounded exercise. But we submit that adopting and utilizing a conceptual framework such as this is preferable to current policy rationalizations based on Old School Economics. We also recognize that this proposed framework remains contingent to circumstances. In an arguably contestable market, where adequate supply and choices for end users allows the evolutionary processes to function, and institutions foster accountability and social trust among market agents, the "tinkering without tampering" formula should prove most effective. Where, however, one market sector may require only minimal tinkering to maximize innovation, economic growth, and emergent public benefits, another sector might benefit from more extensive intervention designed to prop up insufficient market forces, or repair or replace damaged institutions. On balance, we believe that the provisional nature of the enterprise speaks well to our overall emphasis on flexible, tailored, context-specific, and reversible steps by policymakers.

1. Not Dictating the Evolutionary Process

Trusting the evolutionary process of the market amounts to trusting ordinary people to make the right decisions about their lives. Such a stance is democratic with a small "d." Moreover, while the results may not be optimal or efficient for all, the market comes closest to the meritocracy we should want. The self-organizing market process promotes effectiveness over efficiency, and coordinates economic decisions better than any known options—not optimally, just better.[412] To paraphrase Winston Churchill, "markets are the worst form of economic system—except for all the others."[413]

Effective policymakers need to possess at least two qualities: the ability to make reasonably accurate forecasts, and the ability to understand the effects of changes in policy on the system in question.[414] As we have seen, there are deep underlying reasons for the inability to plan and control outcomes successfully. Ball notes that "[t]here are few easier targets than governmental, regulatory, and planning decisions that have had the opposite of their intended effects. In many such cases these unwanted outcomes can be put down to a failure to appreciate the interconnected and interactive nature of the system concerned."[415] Government officials' predictive ability, and the likelihood of unintended consequences in a CAS like the market, should loom large as potential drawbacks to market interventions.

There will be losers as well as winners from economic growth. The evolutionary process between market players involves weeding out the good from the not-as-good, or preferred over not-as-preferred, or adapted to not-as-adapted. As the path of growth proceeds, old industries die and new ones are created. Agents in the old industries typically will plead for protection against the new technology. The government should resist mightily such entreaties,[416] and refrain from interfering in the weeding-out process by leveling the proverbial playing field to benefit one company, or sector, or industry. Favoring any particular outcome interferes with the effectiveness and meritocratic nature of the contest itself. Under Ormerod's "Iron Law of Failure," evolution necessarily includes extinction. Policymakers should respect that process and not try to disrupt what is beneficial to the system as a whole. Much as forest rangers sometimes allow fires to burn out the ecosystem for its own sake, policymakers should want the fitness threshold—the minimum productivity level necessary for survival—to be sufficiently high for a healthy overall system.[417]

At the same time, this does not suggest that the government must refrain from intervening to ameliorate the *effects* of adaptive change

412. LIPSEY ET AL., *supra* note 125, at 45.

413. In a 1947 speech to the House of Commons, Churchill said that "it has been said that democracy is the worst form of government except all those other forms that have been tried from time to time." RALPH KEYES, THE QUOTE VERIFIER: WHO SAID WHAT, WHERE, AND WHEN 43 (Macmillan 2006).

414. ORMEROD, *supra* note 6, at 55-57.

415. BALL, *supra* note 12, at 454.

416. EASTERLY, *supra* note 52, at 181-82.

417. ORMEROD, *supra* note 6, at 231. As an aside, in the public policy arena often it is those who shout most loudly and vociferously about the disciplinary virtues of the free market, who are in most need of them.

through targeted efforts, such as worker education and retraining programs. However, where markets are functioning adequately to provide agents with sufficient choice and opportunities to act, the forces of change themselves should not be impeded, else innovation and growth are threatened.

a. Don't Differentiate

First, adaptive policymakers should not be in the habit of creating, proposing, or emphasizing particular market alternatives. Businesses, working according to a myriad of strategies, are far better at generating new ideas, and having them tested in the marketplace, than government entities. Beinhocker discusses how the balance between bureaucratic and entrepreneurial tendencies in the market can sustain the evolutionary process of consistent incremental process and occasional big jumps.[418] When the government steps directly into this process, it risks over-emphasizing narrowly-conceived technologies or business plans. For example, the Federal Communications Commission has recognized the innovation-hampering nature of its traditional "command and control" approach to mandating how spectrum licensees use their frequencies.[419]

This is not to say that there should be no governmental role in encouraging the differentiation process. For example, where markets are not functioning properly, it may be helpful to facilitate convening the relevant market actors to fashion cooperative solutions. Whole sectors often face critical dilemmas or limitations, while individual businesses erect intentional or unintentional walls between themselves. If the government can provide a venue for collaborative differentiations, it can productively support the process without controlling it. Similarly, there may be a limited role for the policymaker when businesses rely on scarce government-controlled resources, such as rights-of-way or radio spectrum.

b. Don't Select

Second, the adaptive policymaker should not have any direct role in business plan selection. Market actors must be free to select the Physical

418. BEINHOCKER, *supra* note 22, at 152-56.

419. SPECTRUM POLICY TASK FORCE, FCC, REPORT, ET Dkt. No. 02-135, at 65, 67 n.400 (2002), *available at* http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-228542A1.pdf. Another example of this dictating approach is the now-defunct U.S. Civil Aeronautics Board (CAB), which prior to the Airline Deregulation Act of 1978 was found to have limited market entry by new carriers, mandated point-to-point routing systems, subsidized various routes, and even set formulas governing ticket prices and profitability. *See* U.S. Centennial of Flight Comm'n, Deregulation and its Consequences, http://www.centennialofflight.gov/essay/Commercial_Aviation/Dereg/Tran8.htm.

and Social Technologies that make up their Business Plans, and to innovate toward what they think will be "most fit" for the economic landscape. Selection is the heart of evolution, and the heart of markets. Here, Beinhocker discusses how the "Big Man" process of selection can slow down or even halt the process of evolution. If a single agent controls the system, often according to self-interest that does not align with overall growth, the process is distorted. He explains that:

> In a big man system, the fitness function maximized is the wealth and power of the Big Man (and his cronies), rather than the overall economic wealth of the society. Thus, the creative, entrepreneurial, and deductive-tinkering energies of the population are directed toward pleasing the Big Man.[420]

Policymakers need to be attuned to Big Man thinking that arises from within government structures, as well as similar thinking that comes from outside. The instinct to pick outcomes can manifest itself in well-intentioned bureaucratic design, or in regulatory capture by market actors. Even tempting calls for solutions in the name of enabling the market—like creating new or stricter property rights, or allocating resources to the highest bidders—can become a form of selection favoring certain preferred agents over others.

c. Don't Amplify

Third, it should not be the adaptive policymaker's role to amplify the "most fit" business plans. New technologies undoubtedly are exciting, but interfering with the evolutionary process at this stage is very risky. Artificially bolstering a successful but nascent approach at the same time threatens to push aside competing innovations, or successful current plans. Amplifying either legacy or new approaches threatens the ability of the market to sort itself out according to the wishes and actions of market players.[421]

---

420. BEINHOCKER, *supra* note 22, at 288.

421. Some argue, for example, that the State of California's foray into energy deregulation promoted market conduct that tended to create higher prices, but discouraged other conduct that could have led to price reductions. Darren Bush and & Carrie Mayne, *In (Reluctant) Defense of Enron: Why Bad Regulation is to Blame for California's Power Woes (or Why Antitrust Law Fails to Protect Against Market Power When the Market Rules Encourage its Use)*, 83 OR. L. REV. 207, 211 (2004). The authors there believe that the government's "deregulatory" structure, which among other things relied on non-existent excess capacity to discipline wholesale prices, and consumer price caps to freeze retail prices, actually legitimized anticompetitive conduct by Enron and others, "and made that conduct the norm." *Id.* at 212. Whether one accepts or not this particular interpretation of the complex set of events in California during 2000-2001, it is clear that government actions amplifying specific business plans can lead to negative market outcomes.

Instead, amplification should happen only at the level of individual agents as they navigate the fitness landscape. Business Plans that fail will be de-emphasized; technological combinations that work will be repeated and spread. As this amplification occurs, agents will explore variations on the successful strategy by tweaking it and iterating through the differentiation and selection processes once again.

Amplification of good ideas across multiple agents or sectors happens when agents observe or actively exchange knowledge. This presumes a certain degree of connectivity or cooperation, and here the government may again have a limited and indirect role. When agents close off access to knowledge—through obfuscation, strict propertization, or resistance to shared standards—their actions are counter-productive to the market-wide process of amplification.

2. Enabling the "Econosphere"

A far more appropriate role for government, where important policy goals and objectives are at stake, is to experiment with different changeable elements of the fitness environment within which the evolutionary algorithm operates. As Beinhocker puts it:

> Policies that get the government involved in differentiating, selecting, and amplifying Business Plans would be seen as interfering in economic evolution . . . . In contrast, policies that *shape the fitness environment,* while leaving Business Plan selection and amplification to market mechanisms, are a different matter . . . . As long as markets provide the mechanism for selecting and amplifying Business Plans, then the economic evolutionary process will innovate and adapt in response to those regulations.[422]

One can characterize this role with different metaphors: enabling the "econosphere," filling in market "gaps," tinkering with inputs, or revising rules of the "contest."[423] In one sense, the market constitutes a giant search engine, with economic agents competing algorithmically to determine the optimal results. The fundamental point is to improve the market's ability to formulate and present different options (the quantity function), while leaving the selection processes themselves undisturbed (the quality function). To the extent that growth comes not only from capital markets or government subsidy, but also, if not primarily, from technological progress, the government's role should be to generate conditions in which such growth can occur, without picking or

422. BEINHOCKER, *supra* note 22, at 426.

423. Francois Jacob first popularized the notion that "evolution is a tinkerer." François Jacob, *Evolution and Tinkering*, 196 SCIENCE 1161 (1977).

subsidizing the winners, or hindering the losers.[424]

For purposes of this discussion of communications policy, we believe environmental "tinkering" by adaptive policymakers can be accomplished in at least four different ways: (1) feeding the evolutionary algorithm through diversifying inputs, such as Business Plans and their accompanying Physical Technologies and Social Technologies; (2) fostering connectivity between agents, so that communications links are optimized; (3) shaping the fitness landscape to create economic incentives and increased market trust for certain activities; and (4) enhancing market feedback mechanisms, to facilitate better decisions through generating greater flows of timely and accurate information. Again, to suggest these potential steps of supplying inputs, connectivity, incentives, and feedback is not to endorse their use in any or all situations. Only where an overarching policy decision requires some form of market implementation should one or more of these steps even be considered, and perhaps implemented. But if done correctly, these relatively modest steps can provide major emergent benefits.

The notion of enabling from within the given construct of the market in part has its roots in the insight that some constraints, such as lack of foresight and uncertainty about outcomes, are simply inevitable. Policymakers would be wise to heed this insight, and act within the inherent limitations of human endeavors. In the words of Mark Taylor, a noted complexity theory expert:

> One of the perennial promises of visionaries is that in the future, all things will be possible. Whatever constraints we suffer in this world will disappear and we will be able to enjoy a freedom now barely imaginable. Such promises, however, are always cruel because they cannot be fulfilled. Possibilities are inevitably limited by constraints that can never be overcome. The only viable freedom is not freedom *from* constraints but the freedom to operate effectively *within* them . . . . Constraints provide the parameters within which thinking and acting must occur.[425]

A related issue is that policymakers of all persuasions appear

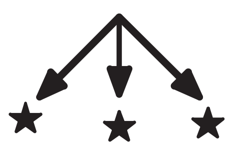
424. A salient example of "tinkering without tampering" outside the ICT space would be various proposals to curb greenhouse gas emissions, including carbon dioxide, by using "carbon offsets." This so-called "cap and trade" system is intended to harness market forces by establishing a total cap on carbon emissions, and then allowing entities to mitigate their own emissions by purchasing "credits" generated from more efficient, alternative fuel sources. While some proposals raise concerns about verification and enforceability—and claims that the government is only licensing pollution—the approach itself is consistent with the notion of productive "tinkering" by tapping into market incentives to achieve larger public policy objectives.

425. TAYLOR, *supra* note 115, at 224.

chronically unable to admit that any single aspect of their policy has failed.[426] They must come to accept the reality that both "market failure" and "policy failure" are inevitable, and learn from the mistakes made, or fitness not achieved. More to the point, failure creates fodder for future growth. "Paradoxically, failure at the detailed, individual level, whether plant or animal, company or government policy, is absolutely necessary for the health and vitality of the system as a whole. We need change and evolution to make progress."[427] Policymakers should prize their own unique position and ability to tinker, and thereby encourage "perpetual novelty, adaptation as a function of entire populations, the role of variety and experimentation, and the potential of decentralized and overlapping authority."[428]

In all respects, then, policy decisions in these contexts should be seen not as enduring mandates, but as a series of experiments that compete to evolve over time. Adaptive strategy suggests that policymakers should levy many small bets, in a trial-and-error (or better, trial-and-success) fashion. One should be willing to execute for today, and adapt for tomorrow.

a. Feed the Algorithm

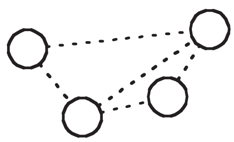


First, the adaptive policymaker can "feed the algorithm" of evolution by adding additional inputs to the process. These inputs include Business Plans, Physical Technologies, and Social Technologies. In some ways, this puts the government in the role of a lab technician, providing different plans and technologies for agents to experiment with in the market through selection.

By allowing, and even nudging, additional inputs to feed the algorithm, optimal amounts of novelty, knowledge, and growth are generated. A diversity of inputs serves as the raw material for differentiation. Ideas are the key input, because they can become innovation (when combined with implementation), physical technologies (when combined with things), and social technologies (when combined with processes). By the same token, supplementing market forces from within via inputs to the emergence algorithm can strengthen the evolutionary process, and yield a richer outcome. The key is to influence the *quantity* of inputs, without disturbing the *quality* of decisions derived ultimately from the algorithm itself. Even light-touch moves can have big downstream effects, both good and bad.

426. ORMEROD, *supra* note 6, at vii.
427. *Id.* at viii.
428. AXELROD & COHEN, *supra* note 19, at 29.

An inescapable conclusion of Romer's work is the need to find ways to increase economic growth. By all accounts, the market by itself is not sufficient to provide every useful input to the rough equation for emergence. One way to feed the evolutionary algorithm is to use the government's spending authority to channel resources.[429] Many experts have discussed the urgent need for technology policy to support research and development.[430] As the Net's own origins plainly show, government-sponsored R&D can help create generative platforms, big and small, for economic growth.



b. Foster Agent Connectivity

The adaptive policymaker also can foster connectivity and networking between various agents in the market. This can be done, for example, by strengthening or adding links (lines of communication) between nodes (agents).

New growth will not happen if the right infrastructure, or institutions—of science and the markets, of conventions and rules—are not in place.[431] What some call the "New Alexandrians," like their ancient counterparts, understand that "creating a shared foundation of knowledge on which large and diverse communities of collaborators can build is a great way to enhance innovation and corporate success."[432] We cannot always rely on competition and short-term self interest alone to promote an optimal infrastructure for ideas. "Vibrant markets rest on robust common foundations: a shared infrastructure of rules, institutions, knowledge, standards, and technologies provided by a mix of public and private sector initiatives."[433] Of course, the Internet is the single best example of such a shared infrastructure, emerging from a mix of first public, and then private actions. So at minimum policymakers should facilitate ways for agents to communicate and interact via the Net.

Joel Mokyr has produced a masterly historical and analytical account of how the costs of accessing useful knowledge (roughly equivalent to our Physical Technologies and Social Technologies) determine how likely it

---

429. Romer, *supra* note 265.

430. As just one example, the National Academy of Sciences issued a joint paper calling for enhancing "the human, financial, and knowledge capital necessary for US prosperity," in part by increasing federal support for various R&D-related tax credits, and providing additional funding for scholarships and fellowships in science, math, and engineering. National Academy of Sciences, *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future* 2 (2007), *available at* http://www.hq.nasa.gov/office/oer/nac/documents/Gathering_Storm.pdf.

431. Bailey, *supra* note 130.

432. TAPSCOTT & WILLIAMS, *supra* note 214, at 178.

433. *Id.*

is that such knowledge will expand.[434] He found that the lower the costs of access, and the greater the supply, the more knowledge will be cumulative.[435] "[T]he knowledge revolution in the eighteenth century was not just the emergence of new knowledge; it was also better access to knowledge that made the difference."[436] Policymakers can look to this salient example as a model for fostering greater connectivity between agents in the market.

c. Shape the Landscape

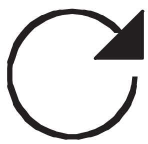
Encouraging greater increases in income over a shorter period of time arguably is the central economic policy task of any nation.[437] And in that quest, incentives for growth obviously matter.[438] With regard to four decades of repeated attempts to turn poverty into prosperity, Easterly concludes, "[n]either aid nor investment nor education nor population control nor adjustment lending nor debt forgiveness proved to be the panacea for growth. [These formulas did not work] . . . because [they ignored] . . . the basic principle of economics: people respond to incentives."[439] Countries where activities that promote growth are rewarded will grow faster than countries where this is not the case.[440] The contribution of the entrepreneur in the growth process is substantial; it has been argued that economies that want to advance faster should embrace a mix of entrepreneurial and big-firm capitalism.[441]

Thus, the policymaker can serve as a "fitness function shaper," which amounts to acting so that "the evolutionary processes of the market can be better shaped to serve society's needs."[442] Because incentives provide useful signals to all agents in the market, the best way to use the fitness landscape to achieve policy objectives is to employ market-based incentives. This can be accomplished by, for example, setting broad policy goals, and then allow agents operating under unfettered economic and non-economic conditions to meet those goals. By shaping the metaphoric fitness landscape within which agents operate—providing incentives to scale particular mountains, or supporting the discovery and sharing of path shortcuts—policymakers encourage policy objectives without interfering with the core activity of

434. *See* MOKYR, *supra* note 236.
435. *Id.* at 8.
436. *Id.* at 74-75.
437. AMARTYA SEN, DEVELOPMENT AS FREEDOM (1999).
438. EASTERLY, *supra* note 52, at 177.
439. *Id.* at 143.
440. BAUMOL ET AL., *supra* note 254, at 13.
441. *Id.*
442. BEINHOCKER, *supra* note 22, at 427.

market evolution.

One of the best examples of a public policy built on a correct understanding of "shaping the landscape" is the FCC's *Computer Inquiry* precedent. As discussed previously, the Commission created its basic/enhanced regulatory dichotomy largely as a way to fence off the online world from unwarranted carrier-style regulation. The FCC's rules established the basis for market forces eventually to evolve new and beneficial Social and Physical Technologies. There would have been no Internet (at least as we now understand it) without that prescient policy decision taken years before the successful rise of commercial online services for consumers.



d. Enhance Feedback Mechanisms

A final form of potentially beneficial tinkering involves creating or enhancing market feedback mechanisms, essentially filling in various information or transparency gaps in the market. This means providing agents with more and better information, and perhaps enhanced decision-making tools as well, so they can make informed decisions. Agents as consumers or users typically lack information, and foresight, and can be easy victims in a marketplace tilted against them. Bounded rationality, asymmetric information flows, cognitive biases, linear thinking—these findings and more suggest that users often stand little chance when negotiating with more powerful agents. The policymaker can help even the odds, at least to some degree. Because consumers and users are adaptable and able to learn and grow, policymakers should give them what they need to take that leap: more information, and a voice.

To be clear, the government should not attempt directly to alter the market outcome. However, policymakers could have a role in maximizing the voices in the marketplace, and trying to ensure they are clearly heard. "As a general rule, democratic interests tend to favor greater transparency, openness, intelligibility, and cheap access to information."[443] One way to do this is to arm the users with tools to better discern for themselves truth from falsehood.

Transparency not only alerts and educates those who make themselves educated agents, consumers, citizens, and producers. It also acts as a form of self-discipline on the affected firms and other entities. Those entities would be less likely to pursue anti-competitive or anti-consumer practices if they must advertise them to the world. Moreover, education can and does go both ways, with students/users imparting

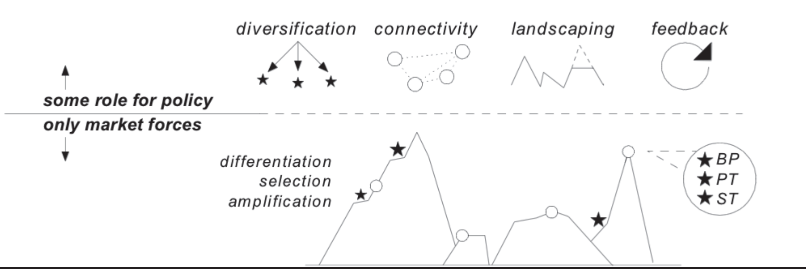
443. STARR, *supra* note 162, at 17.

knowledge by challenging various assumptions, and asking questions not previously considered.[444] By allowing users greater transparency into market processes, and more information about market decision points, governments can initiate a virtuous cycle of interaction between policymaker and citizen. A mutually beneficial "cognitive diversity" can result.[445]

Some lean on user transparency as a key remedy to the network neutrality conundrum. Among other suggestions, Phil Weiser and Rob Atkinson call on the FCC to adopt a "notice and monitoring regime" that would require broadband providers to announce details about their provision of service to consumers, and then adhere to such policies.[446] Atkinson separately has discussed another tool for policymakers, creating user-generated mapping interfaces to track broadband deployment.[447] More information also can help promote self-help; after all, if even a small fraction of end users are more aware of the policies and limitations on service, they can use software or hardware tools to engage in their own efforts to monitor their broadband connections—and, if possible, act accordingly.

Thus, in the context of the fitness landscape metaphor introduced in Section I, this dichotomy between acceptable "tinkering" and unacceptable "tampering" in the workings of the market might be usefully conceptualized as such:

Fig. 2: Tinkering Versus Tampering in a Fitness Landscape



444. JARED M. DIAMOND, COLLAPSE: HOW SOCIETIES CHOOSE TO FAIL OR SUCCEED 419 (2005). Reflecting echoes of behavioral economics, Diamond goes on to dissect various paths to failure by group decision-making, including an inability to anticipate, perceive, attempt to solve, and actually solve major problems. *Id.* at 419-40.

445. PAGE, *supra* note 241. Nonetheless, more information is not always better. Consumer choice can be taken to an extreme; too many options can mean confusion, and even paralysis, while increasing costs unnecessarily to providers. TALEB, *supra* note 67, at 142-45. As with all things, policymakers need to seek an appropriate balance.

446. ROBERT D. ATKINSON & PHILIP J. WEISER, A "THIRD WAY" ON NETWORK NEUTRALITY 14 (2006), http://www.itif.org/files/netneutrality.pdf.

447. Atkinson, *Framing*, *supra* note 374, at 168 n.107.

CONCLUSION

What we have here labeled and critiqued as Old School Economics—that form of economics that has become received wisdom by too many in the U.S. public policy community—still holds many important truths about our human condition. At the same time, some of the key assumptions and verities of that influential form of economic thinking have been proven overstated, or even wrong. The market is a far more rich, dynamic, and complex place than has been assumed. While the larger field of academic economics has been incorporating the newer ways of thinking, for the most part news of these developments has not reached the chambers of the U.S. Congress, or the West Wing of the White House, or the eighth floor of the Federal Communications Commission. For our country's larger economic, social, and political interests to be better served, that situation should change.

What we have distilled and call here Emergence Economics offers us the promise of a new conceptual framework, a way of approaching and understanding the growth-oriented network economy that is being brought about by the Internet. That framework seeks neither to deterministically engineer this dynamic economy, nor to blindly assume that it is evolving toward perfect efficiency. But with new frameworks come new ways of seeing. Romer and others have amply demonstrated that knowledge and technology are not just outputs of the economy, but also essential inputs that drive economic growth and countless other social benefits. Numerous researchers also have shown how game-changing, disruptive innovations tend to emerge from the edges of the Net. These innovations in turn create far-reaching benefits to unaffiliated entities, in the form of innovation "spillovers," and further inputs, throughout the network. This sort of edge-driven, broadly beneficial, mutually reinforcing activity thrives in an environment of open "generativity," where no market player—whether government or firm—unilaterally can pick winners and losers.

The government's unique role in all this, at best, should be to experiment with the optimal background conditions for a dynamic, unpredictable, and evolving environment. In particular, adaptive policymakers should determine whether and how to tinker with the market's inputs, connectivity, incentives, and feedback—and then stand back to let the process itself unfold. With empowered agents working through connected networks via evolutionary processes, we are more likely to unlock the full-blown emergence of new ideas and innovation, of economic growth and other "Net effects." Only when private markets and public policies learn to work constructively with each other, and not in needless conflict, can those emergent benefits be more fully realized in our everyday lives.

COPYRIGHT AND THE RULE OF REASON

CHRISTOPHER SPRIGMAN*

INTRODUCTION

In this essay I will describe several ways to reduce the growing disconnect between the utilitarian rationale that undergirds copyright law and the copyright law itself. Before I come to the arguments, however, a few words regarding the premise: What do I mean by the "disconnect" between copyright law and its justification?

Copyright is sometimes justified as the appropriate reward for an author's creative labor.[1] And copyright is also sometimes justified as a way that we acknowledge an author's strong interest in a creation that reflects and embodies his or her personality.[2] But the dominant justification for copyright, at least in the United States, is explicitly utilitarian. Congress's power to create patent and copyright laws is provided for explicitly in our Constitution, and—uniquely among the

* Associate Professor, University of Virginia School of Law. Thanks to participants at a conference on antitrust and intellectual property organized by Tom Cotter at the University of Minnesota Law School. Thanks also to Albert Choi, Scott Hemphill, Bill McGeveran and Dan Ortiz for helpful comments on an earlier draft, and to Brady Cox for expert research assistance. All errors are, of course, mine.

1. For a description and critique of the labor justification for copyright, see Seana Valentine Shiffrin, *Lockean Arguments for Private Intellectual Property*, *in* NEW ESSAYS IN THE LEGAL AND POLITICAL THEORY OF PROPERTY 138 (Steven R. Munzer, ed., 2001).

2. *E.g.*, Justin Hughes, *The Philosophy of Intellectual Property*, 77 GEO. L.J. 287, 330-65 (1988) (describing, critiquing, and providing arguments to bolster Hegelian personality justification for copyright).

provisions describing Congress's powers—the grant contains a purpose clause that sets out an explicitly utilitarian rationale. Congress is given the power to pass patent and copyright laws "[t]o promote the Progress of Science and useful Arts."[3]

Congress has operationalized that command in laws that provide copyright owners with the right to prevent others from copying (either in whole or in any not-insubstantial part), performing or displaying, or distributing their works without authorization.[4] Our copyright laws are based on our expectation that, by creating these property rights, we will encourage the creation of new artistic and literary works.[5] And at the level of theory, this makes sense: without some form of inducement, we would have cause to worry about the amount of new works produced.[6] Creative works are often quite expensive to produce, and yet cheap to copy. Without some control over copying, we would expect that competition from copyists would force the price of creative works down toward the marginal cost of making a copy. By providing property rights that rightsholders can assert against copyists, we enable rightsholders—at least those in possession of commercially-valuable works[7]—to charge a

3. U.S. CONST. art. I, § 8, cl. 8.

4. *See* 17 U.S.C. § 106 (2006).

5. *See, e.g.,* U.S. CONST. art. I, § 8, cl. 8.

6. The incentives theory may be sound as a general matter, but that does not mean that it applies equally to every form of creativity. Recent scholarship has demonstrated that in areas as diverse as academic research, creative cuisine, and production of open source software, maintenance of incentives to create either does not depend on or appears frequently to ignore the formal intellectual property law. *See, e.g.,* Yochai Benkler, *Coase's Penguin, or, Linux and The Nature of the Firm,* 112 YALE L.J. 369 (2002) (detailing collaborative production in open source software); Christopher J. Buccafusco, *On the Legal Consequences of Sauces: Should Thomas Keller's Recipes be* Per Se *Copyrightable?*, 24 CARDOZO ARTS & ENT. L.J. 1121 (2007) (examining anti-appropriation norms among elite chefs); Dotan Oliar & Christopher Sprigman, *There's No Free Laugh (Anymore): The Emergence of Intellectual Property Norms and the Transformation of Stand-up Comedy*, 94 VA. L. REV. 1787 (2008) (detailing social norms governing creativity among stand-up comics); Emmanulle Fauchart & Eric Von Hippel, *Norms-Based Intellectual Property Systems: The Case of French Chefs* (Mass. Inst. of Tech. Sloan Sch. of Mgmt., MIT Sloan Working Paper 4576-06, 2006), *available at* http://ssrn.com/abstract=881781 (describing the informal norms system discouraging appropriation without attribution among French haute cuisine chefs); Katherine J. Strandburg, *Sharing Research Tools and Materials: Homo Scientificus and User Innovator Community Norms* (May 23, 2008) (unpublished manuscript, *available at* http://ssrn.com/abstract=1136606) (examining social norms governing the sharing of academic science research). This phenomenon of non-IP incentives is not confined to the economic margins. For an example of a major industry in which widespread copying and derivative reworking does not appear to suppress innovation incentives, see Kal Raustiala & Christopher Sprigman, *The Piracy Paradox: Innovation and Intellectual Property in Fashion Design,* 92 VA. L. REV. 1687 (2006) (examining the impact of piracy on the fashion industry).

7. Note the qualification; many artistic and literary works have no commercial value at all, and for these works, copyright creates neither demand nor scarcity and therefore it cannot produce economic benefits for rightsholders. It is only in the case of commercially valuable works where copyright's incentives rationale has any bite. I have previously argued for

higher than competitive price. We hope that this lure will stimulate investment in creative works.

Of course, stimulating the production of more creative works is not the only thing required to "promote progress." We must also worry about *access* to the new works we have incentivized authors to produce—artistic and literary works are valuable because they create our culture, but they do so only if they are widely accessible. Technological developments, including, most importantly, the rise of digital platforms and the Internet, have lowered dramatically the cost of distributing many creative works, thereby lowering the cost of access. But too-strong copyright law is a threat to the increased access that technology would otherwise permit: by raising the price of commercially-valuable new works, copyright threatens to restrict access. This is true in the simple sense of pricing people out of access to works that they would be able to afford in a competitive market. But there is another, very important element to access—i.e., the ability of creators to use pre-existing works as building blocks for new works. Copyright law interferes with this process, and raises the cost of creating new works.

For these reasons, copyright law must seek a balance between private incentives to create new works, and public access to the works created. To pursue balance, copyright law cannot simply provide rightsholders with complete control over their works, and it never has. So, for example, although copyright owners have the right to control the initial distribution of their work, under copyright law's first sale doctrine, owners of copyrighted works are and long have been unable to restrict the price or terms at which their works are re-sold.[8] Similarly, copyright owners' rights have long been limited by a general (albeit narrowly-focused) exemption from liability for uses deemed "fair."[9] And of course copyright rights have always been subject to time limits—the Constitution's "limited Times" proviso proscribes perpetual copyright,[10] and a term that ends, thereby eventually sending works into the public domain where they will forever remain as "free as the air to common use",[11] is an important way in which American law has balanced private

copyright's reformalization – i.e., for the reintroduction into the copyright law of a set of copyright formalities such as registration, notice, and renewal – as a means for focusing copyright on the commercially valuable works for which property rights are salient. *See* Christopher Sprigman, *Reform(aliz)ing Copyright*, 57 STAN. L. REV. 485 (2004).

8. *See* 17 U.S.C. § 109 (2006); Bobbs-Merrill Co. v. Straus, 210 U.S. 339 (1908) (holding that copyright law does not allow rightsholder to control price or terms for resale of his or her work).

9. *See* 17 U.S.C. § 107 (2006); Folsom v. Marsh, 9 F. Cas. 342 (C.C.D. Mass. 1841) (offering early distillation of factors relevant to finding of fair use).

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

11. Int'l News Serv. v. Associated Press, 248 U.S. 215, 250 (1918) (Brandeis, J., dissenting).

incentives with public access. The particular limitations point toward a more general axiom: a copyright law built on utilitarian foundations should be limited to those rights necessary to create and maintain an ample supply of new works. All other uses should be left unregulated.

To achieve this balance, copyright law must be based on a *theory of harm*—i.e., we must understand the kinds of uses that cause significant harm to authors' incentives, and those that do not, so that we might focus on regulating the former and not the latter. And here is where the problem arises: although we understand copyright's concept of harm at an abstract level—i.e., copyright "harm" arises from any use that threatens to suppress author incentives significantly below the optimal level—the theory is exceedingly difficult to apply in many cases.

One frequently-occurring example is the creation of a derivative work—i.e., a work based on a pre-existing copyrighted work that adds significant new creativity. Current copyright law gives rightsholders the exclusive right to make or to authorize the production of all works that use a not-insubstantial amount of material from their work—even where the second creator adds substantial new creativity of his own.[12] What is the justification for this very broad limit on others' creativity? At the very least, the rule seems overbroad, for authors' incentives are not automatically implicated whenever someone makes and distributes an unauthorized derivative work. If a derivative is close enough (both in subject matter and time) to the original that it competes with it, or perhaps preempts a product or licensing market that the author might otherwise plan to enter, then the derivative work at issue may harm the original author. In such an instance, if the copyright law did not reach conduct in this category, we might fear overall harm to author incentives to create. However, if (as is often the case) the derivative is not closely similar to the original, does not compete with it for audience patronage, and does not preempt a market that the original rightsholder realistically is positioned to exploit, then the existence of the derivative is unlikely to harm the original author. In a large number of cases, plaintiffs have claimed, as we would expect, that they are harmed because the defendant

12. *See* 17 U.S.C. § 106(2) (2006). A property right of this sort is far from an inevitable feature of copyright. Indeed, the early U.S. copyright statutes contained no derivative work right, but limited copyright to the "sole right and liberty of printing, reprinting, publishing and vending" – i.e., exclusive rights of reproduction and distribution. Copyright Act of May 31, 1790, ch. 15, 1 Stat. 124 (1790) (repealed 1802). It is also entirely possible, moreover, that a copyright system could give copyright owners a right to get paid for derivatives, but not to control them. The law could do this by, for example, subjecting derivative works to a liability rule – i.e., the second author would be free to make the derivative, but would be required to pay compensation. The current copyright laws already use this intermediate type of right for derivative works based on copyrighted musical compositions– known colloquially as "cover songs"– which may be produced under a compulsory license so long as the basic character of the work is preserved. 17 U.S.C. § 115 (2006).

did not pay for a license to make the derivative. And courts have tended to approve these claims.[13] They do so in part because they lack a theory, or indeed much evidence, to help them distinguish cases in which control is necessary to maintain author incentives from those in which defendants' use does not pose any meaningful chance of harm.

In the absence of such a theory, the courts' tendency has been to give to rightsholders more and more complete control over derivative works. This is true especially because copyright is a strict liability tort. Once a plaintiff proves unauthorized copying of any not-insubstantial protected material, the burden shifts to the defendant to prove that its use is fair.[14] And, importantly, fair use is currently not well designed to produce information about potential harms. Defendants bear the burden of proving fairness, and yet it is plaintiffs who are most often in possession of any evidence relevant to the likelihood that a particular use will harm them. In the absence of evidence, courts are reduced to casual empiricism, or, worse, abstract theorizing about markets and effects. In sum, copyright law's current structure and distribution of proof responsibilities has contributed to copyright's growing disconnect from its utilitarian justification.

It's time to think of a new approach. There has been useful work recently that begins the work of restructuring copyright law along more defensibly utilitarian lines. I refer in particular to important articles by Shyamkrishna Balganesh[15] and Christina Bohannan.[16] Balganesh suggests the addition to the copyright plaintiff's prima facie case of a generally-applicable test of "foreseeable copying"—i.e., that the defendant's copying was of a type reasonably foreseeable as of the time that plaintiff created the work in suit.[17] The aim of this additional element of the plaintiff's case, Balganesh argues, is to limit copyright to the scope of its utilitarian justification: if the copying was not of the type that an author reasonably would have foreseen *ex ante*, it could not have

13. *See, e.g.*, Bridgeport Music, Inc. v. Dimension Films, 410 F.3d 792, 801 (6th Cir. 2005) (holding, in a case involving audio sampling of two-second guitar chord, that defendant must "[g]et a license or . . . not sample"); Rogers v. Koons, 960 F.2d 301 (2d Cir. 1992) (failure to license plaintiff's photograph for use as model for defendant's sculpture preempts licensing transaction and not fair use); Video-Cinema Films, Inc. v. Lloyd E. Rigler-Lawrence E. Deutsch Found., 2005 U.S. Dist. LEXIS 26302 (S.D.N.Y. Nov. 1, 2005) (failure to license found likely to harm value of plaintiff's movie clips).

14. *See, e.g.,* Campbell v. Acuff-Rose Music, 510 U.S. 569, 590 (1994); Harper & Row, Publishers v. Nation Enters., 471 U.S. 539, 561 (1985); H.R. REP. NO. 102-836, at 3 n.3 (1992), *reprinted in* 1992 U.S.C.C.A.N. 2553.

15. Shyamkrishna Balganesh, *Foreseeability and Copyright Incentives*, 122 HARV. L. REV. 1569 (2009), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1117655.

16. Christina Bohannan, *Copyright Harm, Foreseeability, and Fair Use*, 85 WASH. U. L. REV. 969 (2007).

17. Balganesh, *supra* note 15, at 29.

affected (one way or the other) his incentives to create.[18]

Like Balganesh, Bohannan argues for a concept of copyright harm that requires proof that a particular use of a copyrighted work is likely to have a material effect on a reasonable copyright owner's *ex ante* incentive to create a new work.[19] Bohannan would, however, employ a foreseeability standard not as an element of the plaintiff's prima facie case but rather to reform the fair use defense—in fact, Bohannan argues that, properly understood, the Supreme Court's fair use cases already hew to this model.[20]

The Balganesh and Bohannan articles make an important contribution by using foreseeability analysis as a means to re-link copyright law with its utilitarian justification. With that said, in my view, adding a foreseeability requirement, either to plaintiff's prima facie case or to the fair use doctrine, will not in itself supply an administrable theory of copyright harm. The reason for this can be glimpsed in the Supreme Court's recent opinion in *Eldred v. Ashcroft*.[21] In *Eldred*, the Court upheld against Copyright Clause and First Amendment challenges Congress's 20-year extension of subsisting copyrights.[22] The gravamen of the Copyright Clause challenge was a seemingly commonsensical proposition: the Copyright Clause premises Congress's copyright lawmaking on the promotion of progress, but extension of existing copyrights cannot possibly promote progress because authors create any particular work only once, and so extending terms retrospectively is simply a *quid pro nihilo*.[23] In her opinion for the Court, Justice Ginsburg made quick work in dismissing this argument. Noting that Congress consistently has extended subsisting terms each time it extended future terms, Justice Ginsburg wrote, "the author of a work created in the last 170 years would reasonably comprehend, as the [protection] offered her, a copyright not only for the time in place when protection is gained, but also for any renewal or extension legislated during that time."[24]

This is obviously a point about foreseeability, and it illustrates what I take to be an intractable defect that makes the concept difficult to employ as a stand-alone limitation to copyright's theory of harm. What is "foreseeable" in any particular copyright case is uncertain and readily manipulable. If Justice Ginsburg and a majority of the Court believe that the prospect of distant future copyright term extensions helps to shape

18. *Id*. at 34.
19. Bohannan, *supra* note 16, at 970.
20. *Id.* at 991-1002.
21. 537 U.S. 186 (2003).
22. *Id.* at 218-19.
23. *Id.* at 189.
24. *Id.* at 214-15 (footnote omitted).

the incentives of rational authors, then who is to gainsay that the prospect of licensing revenues from the use of plot incidents from a famous television comedy as fodder for a quiz book,[25] or from turning a children's book involving a trouble-making cat into a satirical comment on the O.J. Simpson murder trial,[26] or from the use of a picture postcard photograph as the basis for a modern art sculpture,[27] will foreseeably shape author incentives as well and thus class among the uses subject to copyright law?

Copyright scholars will continue the quest for a useful theory of copyright harm. But in the meantime I suggest we lower our sights. There are other, more modest, ways to nudge copyright law back toward its utilitarian justification. We can rely on a set of indirect strategies to push the incentives of rightsholders in a direction that will helpfully separate unauthorized uses that reduce author incentives from those that do not. I want to briefly suggest and defend two related strategies.

First, we should distinguish between conduct we know will harm author incentives over the run of cases, and conduct with more ambiguous effects. So creation and distribution of exact copies of a work should be treated differently than creation of a derivative work. The first we know will almost always be harmful; whether the second is depends on the facts of a particular case.

Second, we should re-structure copyright's burdens of proof to better filter harmful from harmless uses. This second strategy grows out of and is aimed at implementing the first. For cases involving infringing conduct that is very likely to cause harm, we should preserve copyright's current strict liability rule. Indeed, perhaps we should strengthen it by limiting the availability of the fair use defense in these cases of "per se" copyright liability. But for cases involving infringing conduct in our second category—i.e., where the effect of the infringing conduct is ambiguous—*we should require plaintiffs to prove that they have been harmed in some substantial way*.[28]

There are two principal benefits of such a change. First, by requiring that plaintiffs show substantial actual or likely harm in these "rule of reason" copyright infringement cases, we will encourage plaintiffs who have suffered substantial harm to come forward, while discouraging

25. Castle Rock Entm't, Inc. v. Carol Publ'g Group, 150 F.3d 132 (2d Cir. 1998) (defendant's Seinfeld Aptitude Test, a quiz book based on plot elements of the Seinfeld television comedy show, was infringing and not fair use).

26. Dr. Seuss Enters., L.P. v. Penguin Books USA, Inc., 109 F.3d 1394 (9th Cir. 1997) (holding that a book written in the rhyming style of Dr. Seuss's *The Cat in the Hat* and commenting satirically on the O.J. Simpson trial not protected by fair use).

27. Rogers v. Koons, 960 F. 2d 301 (2d Cir. 1992) (transformative use of postcard image in sculpture not fair use).

28. *See infra* Section II. B.

suits by rightsholders who suffer no harm, or only speculative harm. Second—and perhaps most importantly—altering the plaintiff's prima facie case in this way will produce information about harms and benefits of different uses of copyrighted works. To do this effectively, the law needs to place the burden on the party most likely to have information about the harm—in virtually all cases, that is likely to be the plaintiff. The law as structured now does not reliably produce this information, with the result that copyright litigation does not help us to know more about how creative incentives are or are not harmed. If we hope to improve our understanding over time, we should re-structure the law so that litigation produces the information about harm that we currently lack.

I. COPYRIGHT RULES VS. ANTITRUST STANDARDS

A. *Per Se Copyright vs. Hybrid Antitrust*

Given the costs of copyright regulation, the wide range of conduct regulated, and the differing effects on authors' incentives arising from different forms of infringing conduct, one would expect to find some feature in the copyright law that separates the unauthorized uses of copyrighted works that we suspect will, in almost all cases, strike at authors' incentives from those that have more ambiguous effects. There is, however, no such mechanism in current copyright law. Copyright operates, for all practical purposes, according to a blanket per se rule.

By "per se", I mean that copyright liability is imposed without requiring any showing of actual harm or that the incentives of the plaintiff in an individual case (or of authors generally) have been or would be impaired in any way by the infringing conduct. Instead, current law imposes liability wherever there is conduct that transgresses one of the copyright owner's exclusive rights.[29] Harm to the plaintiff is presumed based on proof of the infringing conduct.

If we decided to re-structure copyright to account for the differing propensities of various types of infringing conduct to harm creative incentives, our first task would be to identify the forms of infringing conduct that we believe are likely in most cases to harm authors' incentives. Conduct in this category should be treated as "per se" copyright infringement and condemned without any need for an individual plaintiff to show that he, or authors in general, have been harmed or are likely to suffer harm in the future. However, for conduct that, although infringing, has ambiguous effects on authors' incentives,

29. 17 U.S.C. § 501(a) (2006) (defining infringement as violation of rightsholders' exclusive rights under 17 U.S.C. § 106).

liability should be reserved for instances in which individual plaintiffs can show that they have in fact been harmed.

We see such a bi-partite liability structure in the law of antitrust, where a small (and shrinking) category of conduct is subject to a per se rule of illegality, but where most forms of potentially anticompetitive conduct are made subject to antitrust's rule of reason. Conduct in this latter category is, unlike that subject to antitrust's per se rule, not the type of conduct that experience or economic theory has identified as reliably anticompetitive across the run of cases. Rather, conduct subject to antitrust's rule of reason has ambiguous welfare effects. Depending on the facts of a particular case, rule of reason conduct may be anticompetitive, or it may be competitively neutral or even pro-competitive. And since we don't know which it will be in advance, plaintiffs in antitrust rule of reason cases are required, as an element of their prima facie case, to demonstrate that the particular conduct at issue harms competition.[30] Copyright law would benefit from the adoption of something approximating antitrust's strategy of calibrating the liability trigger to different forms of potentially harmful conduct. Toward that end, we might consider the importation of a "rule of reason" governing infringing conduct with ambiguous effects on authors' incentives. A copyright law reformed to differentiate between per se and rule of reason infringement would better balance protection of authors' incentives with the widest possible public access to creative works. The introduction into copyright law of an antitrust-like approach could, therefore, yield a substantial efficiency gain. Such a move would also take considerable pressure off of copyright's fair use doctrine, which was never meant to serve—and has in practice never sufficed—as a general mechanism to weed out claims where authors' incentives are not at stake. Finally, and perhaps most importantly over the long term, the introduction of a per se/rule of reason distinction into copyright law would incentivize the production of more information about the incentive effects of a variety of uses of copyrighted works. Progress in our understanding of copyright rules and the effect of unauthorized uses on authors' incentives requires empirical data illuminating the effects of different uses on the markets for the huge variety of creative works that copyright law governs. And yet at the moment, copyright is structured in a way where such information has little salience, and is therefore not produced.

B. *Choosing Between Rules and Standards*

I should pause to make a quick point about terminology. I am

30. *See, e.g.*, Geneva Pharms. Tech. Corp. v. Barr Labs., Inc., 386 F.3d 485, 506-07 (2d Cir. 2004).

arguing for a copyright rule of reason, but the "rule of reason" is not properly a rule at all. It is, rather, a standard.

Rules are legal commands that differentiate between legal and illegal conduct in a way that readily may be determined *ex ante*. The way in which we typically regulate speeding is an example—a speed limit is posted (e.g., 65 mph), and you violate the law if you exceed it. Speeding usually is subject to selective enforcement; it may be that in a particular jurisdiction few drivers are pulled over for doing less than 70 when the posted speed is 65. But selective enforcement does not change the fact that the offense of speeding is defined according to a rule. The proscribed conduct is identified and readily understood in advance.

In contrast, standards are legal commands that differentiate between legal and illegal conduct in a way that requires significant decision-making *ex post*. An example would be a speeding rule that directed drivers to "drive reasonably according to the current weather and traffic conditions." Enforcement of a standard against speeding would require police officers, and then courts, to assess reasonableness under the conditions in each case. This may sound bizarre as a regime to govern speeding. It is, however, the regime that governs the accidents caused in part by speeding, for our conduct as drivers in general is assessed according to a standard. All of us are required when driving to conform to an objectively reasonable standard of conduct, assessed according to circumstances in individual cases.

There is an enormous and well-developed literature setting out the trade-offs between rules and standards in structuring legal commands.[31] Rules promise simplicity, at the cost of accuracy. Staying with the speeding example, 65 mph is not an ideal speed limit for all drivers and all vehicles in all conditions traversing any stretch of road so marked. Speed limits are nonetheless structured as a rule because of the cost and difficulty of enforcing a standard, where neither drivers nor police would know with certainty, until a judge or other fact-finder sorts through the

31. *See, e.g.*, RICHARD A. EPSTEIN, SIMPLE RULES FOR A COMPLEX WORLD, 21-128 (3d ed. 1997) (arguing for simplification of law by favoring formal rules in many different contexts); Colin S. Diver, *The Optimal Precision of Administrative Rules,* 93 YALE L.J. 65 (1983) (illustrating the applications and limitations of precise rules and imprecise standards); Isaac Ehrlich & Richard A. Posner, *An Economic Analysis of Legal Rulemaking,* 3 J. LEGAL STUD. 257 (1974) (arguing that the desire to minimize costs is a dominant consideration in the choice between precision and generality in the formulation of legal rules and standards); Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557 (1992) (concluding that the desirability of rules and standards is most influenced by the frequency of the conduct that will be governed by the law); Carol M. Rose, *Crystals and Mud in Property Law*, 40 STAN. L. REV. 577, 604-10 (1988) (arguing that since "muddy" standards and bright-line "crystal" rules serve different functions, neither is necessarily preferable to the other); Kathleen M. Sullivan, *The Supreme Court 1991 Term: Foreword: The Justices of Rules and Standards*, 106 HARV. L. REV. 22 (1992) (outlining several arguments commonly used in support of both rules and standards).

particular case, where the borderline lies between lawful and unlawful conduct. So in the case of speed limits we favor a rule because of the expense of enforcing a standard through individualized determinations, combined with the expectation that we have little to gain from more precisely tailoring the speed limit via a standard (especially given the availability of selective enforcement of the rule, which also serves a tailoring function, albeit incompletely).

Standards promise the converse: they provide accuracy, at the cost of complexity. An example would be parties in a closely contested tort suit arguing about the reasonableness under the circumstances of specific conduct. The tort system's negligence standard requires substantial investment to determine whether the conduct in question falls below the threshold. Undertaking that investment is worthwhile, however, wherever the possible varieties of negligent conduct are so great that defining a rule in advance is likely to lead to unacceptable over- and under-inclusiveness.

C. *Copyright's Per Se Rule and the Fair Use Standard*

Current copyright law contains both rules and standards, although they are distributed idiosyncratically. The plaintiff's prima facie case in a copyright infringement lawsuit is structured as a per se rule—i.e., if the plaintiff proves infringing conduct (even if undertaken unconsciously), the defendant is liable. This is so whether or not the plaintiff was harmed by the particular infringing conduct. Once the rule is shown to have been breached, harm is presumed.

Copyright law includes a *defense* to infringement liability that is constructed as a standard, or, in Congress's conception, as an "equitable rule of reason."[32] This is the fair use defense, which arose as judge-made doctrine but is now set out in the statute itself:

> § 107. Limitations on exclusive rights: Fair use
>
> Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include–

32. *See* H.R. REP. NO. 94-1476, at 65 (1976), *reprinted in* 1976 U.S.C.C.A.N. 5659, 5679 ("since the [fair use] doctrine is an equitable rule of reason, no generally applicable definition is possible, and each case raising the question must be decided on its own facts.").

> (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
>
> (2) the nature of the copyrighted work;
>
> (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
>
> (4) the effect of the use upon the potential market for or value of the copyrighted work.
>
> The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors.[33]

Is reliance on the fair use defense a workable way to focus copyright on harmful conduct? Judging from the life of the defense thus far, the answer appears to be no. Current law constructs fair use as a defense to infringement, with the burden of proof on the defendant.[34] But in most copyright cases it is *the plaintiff* that has superior access to information about harm—harm to himself, directly, and, by extension, harm to other authors similarly situated. Fair use does not oblige the plaintiff to produce that information. Courts are instead reduced to theorizing about harm, and they do this badly.

There is another reason—probably of greater practical importance—that fair use cannot function as a general mechanism for sorting harmful from harmless uses. This has to do with copyright's remedies provisions. These are constructed in a way that makes it difficult to predict, with any precision, what penalties are likely if an infringement is not excused as a fair use. Under current law a copyright owner who successfully brings a lawsuit for infringement of a registered copyright is entitled to choose between two principal measures of damage. The plaintiff may elect to recover his "actual damages" as well as "any additional profits of the infringer" that are "attributable to the infringement and are not taken into account in computing the actual damages."[35] Under this formulation, the plaintiff is entitled to recover his actual losses (including the value of any license that might have been obtained absent the infringement) plus the disgorgement of any profits realized by the infringer. Alternatively, in instances in which the plaintiff's copyright was timely registered, the plaintiff may choose to recover statutory damages in amounts set out in the copyright law and

33. 17 U.S.C. § 107.
34. *See* sources cited *supra* note 14.
35. 17 U.S.C. § 504(a) (2006).

not linked to any particular showing of harm. Current law permits statutory damages ranging between $750 and $30,000 per work infringed for ordinary infringement,[36] and up to $150,000 per work if the infringement is deemed willful.[37] Additionally, and importantly, the court is permitted, but not required, to award prevailing parties' court costs and attorney's fees[38] and (most likely) prejudgment interest.[39] And injunctive relief is available and is freely granted on both a preliminary and permanent basis.[40]

The wide variation in possible damages, the easy availability of equitable relief, and the possibility that plaintiffs may obtain wide-ranging remedies even in the absence of any showing of harm, means that would-be users are unlikely in many instances to rely on fair use as a sorting mechanism. The risks of failure are too great.

At this point we can assess the merits of the current per se rule in copyright. Our current approach has the benefit of (relative) simplicity—the plaintiff's prima facie case is premised solely on proof of conduct defined in advance and the fair use defense applies not in cases involving "ordinary uses of copyrighted works," but only to those involving "unusual or marginal activities."[41] But reliance on the per se rule produces inaccuracy—current copyright law condemns all infringing conduct according to the same rule, whether or not the particular conduct is likely

36. *Id.* § 504(c)(1). A plaintiff's ability to collect statutory damages is contingent upon registration of the work either prior to its infringement, or within three months of its publication. *Id.* § 412.

37. *Id.* § 504(c)(2). Conversely, in cases of innocent infringement – i.e., infringement unaccompanied by any intent to infringe or knowledge regarding the unlawfulness of the infringing conduct – a court may reduce statutory damages to as little as $200 per work infringed. *Id.* The availability of these special statutory damages is conditioned on the plaintiffs' registration of the infringed work prior to the infringement's commencement. *Id.* § 412. In addition to money damages, courts are permitted to order the seizure and destruction or other disposition – including transfer to the prevailing plaintiff – of infringing articles. *Id.* § 503.

38. *Id.* § 505. Like statutory damages, the availability of attorney's fees is conditioned on timely registration of the work. *Id.* § 412.

39. In Design v. K-Mart Apparel Corp., 13 F.3d 559, 569 (2d Cir. 1994) (noting that availability of pre-judgment interest an unresolved issue in Second Circuit); Kleier Adver., Inc. v. Premier Pontiac, Inc., 921 F.2d 1036, 1040-1042 (10th Cir. 1990) (awarding pre-judgment interest). In addition to these civil remedies, the copyright laws also impose criminal penalties for certain instances of copyright infringement. The federal Copyright Act imposes prison terms of up to ten years and substantial criminal fines for infringement of registered works if the infringement has been undertaken either "for purposes of commercial advantage or private financial gain," *id.* § 506(a)(1), or for reproduction or distribution, within a 180-day period, of copies with a total retail value exceeding $1000, *id.* § 506(a)(2). Under these definitions, the prospect of criminal penalties hangs over many, if not most, instances of infringement. Thus far, however, the federal government has been sparing in its application of the criminal penalties.

40. *See* 17 U.S.C. § 502 (2006).

41. Rebecca Tushnet, *Copy This Essay: How Fair Use Doctrine Harms Free Speech and How Copying Serves It*, 114 YALE L.J. 535, 554 (2004).

to harm authors' incentives.

The costs and benefits of a copyright rule of reason are precisely counter to those inhering in the current rule. Under a copyright rule of reason, liability would be premised not simply on proof of an infringing act, but on proof of the actual or threatened imposition of the type of harm that is anticipated. Employing this liability standard would present greater complexity—the plaintiff's prima facie case in every rule of reason infringement action would include not just proof of infringing conduct, but an assessment of whether harm is likely. The benefit of such an expanded inquiry would be a better fit between means and ends—i.e., between copyright's application and the preservation of authors' incentives.

D. Mixing Rules and Standards in Antitrust

Antitrust, like copyright, is an economic regulatory system focused on avoiding a form of market failure—in the case of antitrust, the possibility that firms or groups of firms with market power will suppress competition, raise prices, and deter innovation and investment.[42] And like copyright, antitrust must perform a balancing act. Most conduct that is potentially anticompetitive is also potentially procompetitive. Which outcome is likely in any particular case depends on the characteristics of the firms, products, and markets at issue. There is the persistent worry, moreover, that the welfare effects of most forms of potentially anticompetitive conduct often is ambiguous even upon close examination, and courts therefore may err in their assessment of these effects. As a result, antitrust liability rules that are too aggressive may deter or preempt procompetitive, as well as anticompetitive conduct.

Unlike copyright, however, antitrust law features a mix of rules and standards tailored to different forms of potentially anticompetitive conduct. For conduct known to harm competition in most cases—e.g., conspiracy to fix prices, rig bids, or divide markets[43]—antitrust imposes liability according to a "per se" rule. For conduct in this category, a plaintiff demonstrates a violation of the antitrust law by showing that the defendant intentionally engaged in the proscribed conduct. The plaintiff is not required to prove that the defendant intended—or even was aware of the possibility of—harm to competition. Indeed, the existence of such

42. *See* Harry First, *Controlling the Intellectual Property Grab: Protect Innovation, Not Innovators*, 38 RUTGERS L.J. 365, 390-91 (2007).

43. N. Pac. Ry. Co. v. United States, 356 U.S. 1, 5 (1958) (identifying price fixing, division of markets, group boycotts, and tying arrangements as unlawful activities "in and of themselves"); *see also* Verizon Commc'ns Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398, 408 (2004) (identifying collusion for purposes such as price fixing as the "supreme evil of antitrust").

harm is irrebuttably presumed.[44]

A very different liability standard applies, however, to conduct with ambiguous welfare effects—i.e., to conduct that may harm competition in some instances, but that may be neutral or even procompetitive in others. In such cases, antitrust law assesses liability under a "rule of reason". Courts employing the rule of reason do not presume harm to competition; rather, the plaintiff must demonstrate it, and show that the competitive harm outweighs any associated and offsetting precompetitive effects ascribable to the conduct at issue.[45]

In order to show harm, plaintiffs in rule of reason cases are typically required to describe the relevant market—i.e., the economic market in which the products or services at issue compete. The likelihood of competitive harm is assessed by analyzing the effect of the conduct at issue on competition within that market or markets. In addition, it often is said in the rule of reason context that plaintiff must show that defendant acted with the intent to harm competition. This does not mean that plaintiff must show that defendant subjectively intended to cause competitive harm, but rather than such an intent can be inferred objectively from the character of the conduct.[46]

Importantly, plaintiffs in a rule of reason antitrust case must demonstrate that the specific conduct at issue threatens, on balance, to

---

44. *N. Pac. Ry. Co.*, 356 U.S. at 5 (

> [T]here are certain agreements or practices which because of their pernicious effect on competition and lack of any redeeming virtue are *conclusively presumed to be unreasonable and therefore illegal* without elaborate inquiry as to the precise harm they have caused or the business excuse for their use. This principle of *per se* unreasonableness not only makes the type of restraints which are proscribed by the Sherman Act more certain to the benefit of everyone concerned, but it also avoids the necessity for an incredibly complicated and prolonged economic investigation into the entire history of the industry involved, as well as related industries, in an effort to determine at large whether a particular restraint has been unreasonable – an inquiry so often wholly fruitless when undertaken.

) (emphasis added).

45. Antitrust courts have been quite adept at using burden-shifting methodologies when plaintiffs have made initial showings that a particular form of conduct is quite likely, in the circumstances to harm competition. *See, e.g.*, United States v. Microsoft Corp., 253 F.3d 34, 59 (D.C. Cir. 2001) (noting that after plaintiffs make a prima facie case demonstrating anticompetitive effects, the burden shifts to the alleged monopolists, who may offer procompetitive justifications for their conduct in order to shift the burden back to the plaintiffs to rebut the claim).

46. *See* Ronald A. Cass & Keith N. Hylton, *Antitrust Intent*, 74 S. CAL. L. REV. 657, 659 (2001). In contrast to antitrust, copyright infringement is a strict liability offense – there is no knowledge or intent requirement, and even accidental or unconscious infringement is actionable. *See, e.g.*, De Acosta v. Brown, 146 F.2d 408 (2d Cir. 1944); Bright Tunes Music Corp. v. Harrisongs Music, Ltd., 420 F.Supp. 177 (S.D.N.Y. 1976), *aff'd*, ABKCO Music, Inc. v. Harrisongs Music, Ltd., 722 F.2d 988 (2d Cir. 1983). In this aspect, copyright's per se rule is even more categorical than antitrust's – no one has ever been held liable for "accidental" or "unconscious" price-fixing, market allocation, or bid-rigging.

harm not just the plaintiff but *competition* in some relevant economic market. This dominant form of antitrust analysis requires case-specific inquiries into market harm.[47]

By varying liability rules in this way, antitrust doctrine tailors the law's application to account for the varying likelihood, given different forms of potentially anticompetitive conduct, of actual threat to competition. Perhaps just as importantly, antitrust law's mix of rules and standards incentivizes the production of more and better information about the market impact of various forms of potentially anticompetitive conduct.

Again, the difference between antitrust's per se and rule of reason methodologies is the difference between rules and standards. Antitrust mixes a select number of per se rules (entering into an agreement to fix prices is unlawful) with a relatively large area where conduct is judged according to a standard (an exclusive dealing contract will be judged unlawful if harm to competition from the arrangement's exclusion of rivals outweighs any efficiencies gained via the arrangement). For the conduct subject to per se rules, antitrust sacrifices some accuracy—even price-fixing does not harm competition in all cases. (Think, for example, of a situation involving an incumbent and a potential entrant, and where entry (absent agreement) is likely to trigger marginal-cost pricing. Entry under these conditions might be unattractive to the potential entrant, especially if the incumbent is a lower-cost producer and would thus be able to underprice the entrant in a competitive market. But if the entrant is able to credibly pre-commit to entry and reaches an imperfect price agreement with the incumbent, the result will be duopoly pricing—pricing above marginal cost, and therefore not as good as full competition, but better for consumers than monopoly pricing. Price-fixing under these conditions makes entry and hence lower-than-

47. Cont'l T.V., Inc. v. GTE Sylvania Inc., 433 U.S. 36, 49-50 (1977) ("Under [the rule of reason], the fact-finding weighs all of the circumstances of a case in deciding whether a restrictive practice should be prohibited as imposing an unreasonable restraint on competition. Per se rules of illegality are appropriate only when they relate to conduct that is manifestly anticompetitive.") (citations omitted); Chicago Bd. of Trade v. United States, 246 U.S. 231, 238 (1918) (

> The true test of legality is whether the restraint imposed is such as merely regulates and perhaps thereby promotes competition or whether it is such as may suppress or even destroy competition. To determine that question the court must ordinarily consider the facts peculiar to the business to which the restraint is applied; its condition before and after the restraint was imposed; the nature of the restraint and its effect, actual or probable. The history of the restraint, the evil believed to exist, the reason for adopting the particular remedy, the purpose or end sought to be attained, are all relevant facts. This is not because a good intention will save an otherwise objectionable regulation or the reverse; but because knowledge of intent may help the court to interpret facts and to predict consequences.

).

monopoly pricing possible). But the law also gains clarity and predictability—if you agree to fix prices, you have violated Section 1 of the Sherman Act, full stop. And the prophylactic rule that the law sets up against per se conduct also conserves prosecutorial and judicial resources: antitrust criminal prosecutors and civil plaintiffs need not offer proof of harm in individual price-fixing cases to establish liability, and courts are spared the task of assessing that evidence.

The situation is, of course, reversed for conduct subject to antitrust's rule of reason. The rule of reason standard sacrifices clarity and predictability by subjecting conduct with uncertain welfare effects to an inquiry that varies significantly in individual cases. (The rule of reason inquiry necessarily varies because both the details of competition in different relevant economic markets and the varieties of competitive harm and procompetitive efficiencies that may arise in these settings are as diverse as the forms of potentially anticompetitive conduct). This loss is—in theory—balanced by increased accuracy. If courts are able reliably to distinguish between procompetitive and anticompetitive conduct, then only the latter is subject to sanction. Plaintiffs are required to shoulder the burden of establishing harm to competition, which suggests a bias toward tolerance of type II errors (i.e., false negatives) over type I errors (i.e., false positives).

One final point: based on the differences in proof requirements set out above, it appears to me that antitrust and copyright are premised on divergent understandings of the workings and potential failures of markets. Copyright is justified, as has been stated, by a fear of market failure created by uncontrolled copying, and resulting in sub-optimal incentives to create new artistic and literary works. Judging from the expansion of copyright scope, duration, and the law's indiscriminate use of a per se liability rule, copyright appears to proceed from a view that market failure is the *ordinary* expectation in the absence of legal intervention. That is a contestable premise, and it is interesting to note that the trend in antitrust law over the past quarter-century is in some tension with copyright's view of markets as fragile. Antitrust—which, because of its dominant rule of reason methodology has developed a much deeper understanding of the likely harmfulness of a wide spectrum of potentially anticompetitive conduct—has migrated toward a concept of markets as robust and not frequently subject to failure, and therefore toward more skeptical treatment overall of claims that particular forms of conduct will suppress competition. Accordingly, the domain of the per se methodology in antitrust has been shrinking. Still in this category is price fixing—the "supreme evil of antitrust"[48]—and related forms of conduct

48. *Trinko*, 540 U.S. at 408.

such as bid-rigging, certain horizontal group boycotts, and market division (whether by territory or customer). But the trend in antitrust is to move other forms of conduct, once carefully fenced within the per se enclosure, toward more circumspect treatment under the rule of reason. Just two years ago, the Supreme Court overturned a century of precedent subjecting minimum resale price maintenance to per se treatment, holding that economic experience had shown that the practice often had procompetitive effects and was therefore appropriately analyzed under the rule of reason.[49] And less dramatically, but no less surely, the Supreme Court and lower federal courts have whittled away at the previous per se test for tying, such that at this point there is little, if anything, left to distinguish the antitrust analysis applying to potentially anticompetitive ties from the ordinary rule of reason.[50]

In sum, the clear trend in antitrust law is to subject the great bulk of potentially anticompetitive conduct to the rule of reason. The per se category remains for the few forms of conduct that are believed reliably to lead to competitive harm. This raises an important question: given antitrust's view that markets are in general robust against a variety of threats to competition, should we be skeptical of copyright's implicit assumption that markets for innovation fail readily in the face of copying—even copying that does not appear to displace demand for the plaintiff's original work? Copyright, as it is structured now, does not produce information useful to address this question. This should concern us.

II. SORTING INFRINGING CONDUCT

A. *Per Se Rule for "Consumptive" Infringement*

At this point, I will provide a quick summary. Technology has increased enormously the potential uses of creative works. Copyright law has expanded to embrace—i.e., to characterize as infringement—almost all of these uses. Some unauthorized uses of copyrighted works strike directly at authors' incentives to create new works. Other unauthorized uses do not threaten the kind of market failure that copyright exists to preempt. Freeing these uses would create welfare gains, measured in terms of access to works that otherwise would be unavailable.

One would expect, given the above, that copyright would sort the

49. Leegin Creative Leather Prods., Inc. v. PSKS, Inc., 551 U.S. 877, 877 (2007).

50. *See, e.g.*, Ill. Tool Works, Inc. v. Indep. Ink, Inc., 547 U.S. 28, 31-43 (2006); Newcal Indus., Inc. v. Ikon Office Solution, 513 F.3d 1038, 1051-52 (9th Cir. 2008); Reifert v. S. Cent. Wis. MLS Corp., 450 F.3d 312, 316-20 (7th Cir. 2006), *cert. denied*, 549 U.S. 1265 (2007); *Microsoft Corp.*, 253 F.3d at 84-94.

harmful from the harmless uses. Unlike in the case of antitrust, copyright lacks an internal mechanism to do so. Copyright relies instead on fair use, which is not up to the job.

What can we do about this? We should consider importing into copyright law an internal sorting mechanism. In designing this mechanism we might take cues from the law of antitrust. If we were to do so, we would define categories of "per se" and "rule of reason" infringement. We would re-structure the copyright law to condemn conduct in the former category based solely on proof that the conduct has occurred. Conduct in the latter category, in contrast, would be condemned only where a plaintiff establishes harm to authors' incentives.

Which brings us to the final question: what kind of conduct belongs in each category? I should preface what follows with an important observation about the stakes—any initial allocation of infringing conduct into per se and rule of reason categories is just that: an initial allocation. One of the signal virtues of an antitrust-style reformulation of copyright liability rules is that the application of rules and standards will be informed over time by information about the harms (or lack thereof) produced by different forms of infringement.

With that in mind, we can identify a very significant category of infringing conduct that belongs in the per se category. I refer to "consumptive" infringement. By "consumptive" I mean forms of infringement involving the reproduction *and* distribution of copies that are either exact or near enough so that they are almost certain to compete with the original work for patronage.

As an example of consumptive infringement, consider the reproduction and distribution of exact copies of a copyrighted song. Although there are cases in which unauthorized copying and distribution of this kind could benefit an author (for example, by creating demand for a previously unknown work, which the author is able later to exploit), this wholesale copying is likely, over the run of cases, to harm authors by diverting to copyists sales that authors otherwise would have made.

The category of consumptive infringement will apply to a large percentage of copyright cases that plaintiffs will wish to bring. Most copyright violations involve the making and distribution of exact or near-exact copies. Virtually all of the infringement via peer-to-peer networks falls into this category. So for all of these instances, the per se rule for consumptive infringement preserves current law.

Let me offer a few caveats. First, this per se category of "consumptive" infringement should apply only if an exact or near-exact copy has been made *and* distributed. Accordingly, the per se liability standard should not apply in cases of copying for personal use, including copying for the purpose of time-shifting (i.e., personal use of a VCR or digital video recorder), or device-shifting (as in copying songs from a person's CD collection for listening on that same person's iPod), or to make a back-up copy. Nor should the per se standard apply in cases of "intermediate" copying, by which I mean any copying done not for the purpose of offering the copy in competition with the original, but rather to create some other product or service. The Google Book Search service—a fully-searchable database that eventually will contain many millions of both public domain and copyrighted books—is made possible by intermediate copying. Google makes wholesale digital reproductions of copyrighted books to construct its search database, but the search service returns only "snippets" showing the occurrence of a searched-for term, and does not distribute the copies themselves (except in the case of public domain books or other books for which the rightsholder has agreed to allow distribution of a full-text or partial copy).[51] Because intermediate copying done to offer a separate, non-competing product or service is not very likely to harm copyright owners and depress incentives to invest in new works, this type of conduct should be actionable under a copyright rule of reason only in cases in which a plaintiff is able to establish harm.

Another important example of intermediate copying is the reverse engineering of software. Most software reverse engineering involves the copying of potentially copyrighted computer object code, and then the reconstruction, using engineering techniques, of the human-readable source code. By reconstructing the source code, rival software firms learn how the targeted software works, and how to provide similar or better functionality in their own products.

Assuming for the moment that the reverse engineering is "clean", the process only involves intermediate copying of object code. That is, the rival firm will eventually market a product that may be based in part on what they have learned from copying the target's object code and reverse engineering it back to a facsimile of the target's source code. The rival's product will not, however, distribute any of the copyrighted object code contained in the targeted software. In this case, what emerges from the process is not a copy, but a separate product.

Reverse engineering has been judged fair use in two important cases in the Ninth Circuit.[52] Under the copyright rule of reason, the same result would be obtained in almost all reverse engineering cases, with the exception that the plaintiff would bear the burden of proving harm.

Second, even in cases of copying and distribution of exact or near-

---

51. *See* About Google Book Search, http://books.google.com/intl/en/googlebooks/history.html (last visited Apr. 5, 2009).

52. *See* Sony Computer Entm't, Inc. v. Connectix Corp., 203 F.3d 596 (9th Cir. 2000); Sega Enters. Ltd. v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1992).

exact copies, the per se rule should not apply unless a copyright in the plaintiff's work has been timely *registered*. For reasons I have explained elsewhere,[53] registration is a signal that the copyright owner places some positive economic value on the property right that the copyright law establishes in a creative work. Only in the instances where a copyright owner expects the work to compete and earn returns can we imagine a copy diverting some of the owner's expected market. Infringement claims involving unregistered works should be treated in all instances under a rule of reason.

Classing unregistered works under a rule of reason would substantially ease the current problem with orphan works—i.e., works that are under copyright, but for which would-be users find it difficult or impossible to identify a rightsholder to ask permission to use. Many orphan works are unregistered, and the use of works that lack either registration or some relatively easily-found data regarding ownership is unlikely, in the run of cases, to produce many instances where authors' incentives are harmed. Why is that? Because under current law pre-infringement registration provides important benefits, including the availability in infringement actions of very powerful statutory damages[54] and attorney's fees.[55] Accordingly, the failure to register a work is a signal that the author does not expect that work to produce a significant return. In such cases, any presumption that unauthorized reproduction and distribution harms the author's incentives to create new work is unjustified, for the copying cannot affect incentives if the expected return from a work is at or near zero.

Of course, harm is possible in such cases—for example, if the copying is undertaken as part of the creation of a derivative work of a type that the author can show that he intended to produce, or to license. The plaintiff should be put to the proof of such harm—i.e., the plaintiff should produce evidence that he was likely to enter such a market, and also that presence of a competing derivative was likely to displace demand rather than create it. We should expect that plaintiffs will rarely be able to make such a showing—as Stewart Sterk has noted, situations where control of derivative works makes a substantial difference in an author's ability to recover his costs are quite rare.[56]

53. *See* Sprigman, *supra* note 7, at 502-14.

54. 17 U.S.C. § 504(c)(1).

55. 17 U.S.C. § 412.

56. Stewart Sterk, *Rhetoric and Reality in Copyright Law*, 94 MICH. L. REV. 1197, 1215-17 (1996) (

> One argument for giving authors copyright in derivative works is that the prospect of profits from derivative works is necessary to create adequate incentives for production of the original. The argument is persuasive only in those situations when (1) the projected returns from the original work are too small to justify the costs of

A third caveat is that the definition of consumptive infringement must be carefully drawn to avoid the temptation to which copyright law's general "substantial similarity" liability standard[57] has already fallen victim—i.e., the condemnation of any non-*de minimis* use of protected material. The gravamen of the consumptive infringement category is our strong intuition that production and distribution of exact or nearly exact copies of commercially valuable works will divert some increment of the potential audience for the original to the copy. That intuition is strong enough to support the application—at least initially—of a per se rule to all such infringing conduct. This is not to say that every instance of distribution of even exact copies will inevitably harm the rightsholder in the original work. One might imagine, for example, instances in which the unauthorized distribution of copies of a song could lead to a bandwagon effect that turns the song into a hit, thereby raising consumption of the original—even in the face of the copying—far beyond what circumstances might otherwise have provided. The question is not whether there will be some inaccuracy—some inaccuracy is in the nature of a per se rule. The question is whether the gain in clear understanding of proscribed conduct, stable expectations, and simplified enforcement that a rule provides are worth the occasional misapplication of the law. For copies that are exact or nearly so, the answer to that question is, at least on current evidence, yes.

B. Rule of Reason Infringement

Now that we have developed, at least in brief, a concept of consumptive infringement subject to a per se liability rule, let us move to the second and in some ways more interesting category. What kinds of infringing conduct belong under the rule of reason? Anything that does not fall under the category of presumptively consumptive use. I have already mentioned personal uses, intermediate copying, and the use of

---

> production, and (2) the projected returns from the derivative work are so large relative to the cost of producing the derivative work that the difference will more than make up the projected deficit on the original work alone. These conditions may apply when the original work is an extraordinarily high-budget movie with the potential for sales of toys, t-shirts, and the like, but they are less likely to apply in more common derivative-works cases.

).

57. "Substantial similarity" has long been the predominant liability standard in copyright law. In assessing whether two works are "substantially similar", a court will first look to whether the defendant in fact copied plaintiff's work. In doing so, the court will assess whether the defendant had access to plaintiff's work, and whether the works are similar enough that an inference of copying is warranted. If the court finds copying in fact, it will then inquire whether that copying went too far – i.e., whether it involved non-de minimis appropriation of protected material – so as to constitute an improper appropriation. *See* Arnstein v. Porter, 154 F.2d 464, 468 (2d Cir. 1946).

unregistered or orphaned works. These are all important, but the most significant cases in this category are likely to involve the creation of derivative works.

Depending on the circumstances in particular cases, a derivative work may, or may not, compete with the original. What, then, should we do with derivative works which use substantial portions of an original work, but are neither exact copies nor nearly so? For these works, we should apply the rule of reason, conditioning liability on the plaintiff's proof of the particular types of harm that we might imagine.

The plaintiff might seek to prove, for instance, that there is significant cross-elasticity of demand between the original and the derivative, and therefore allowing unauthorized reproduction and distribution of the derivative will in fact divert enough of the sales that the original author might otherwise have enjoyed that the court is able reasonably to conclude that the loss, if assessed *ex ante*, would have affected the author's incentive to create. A plaintiff might make such a showing by producing survey evidence of cross-elasticity, in just the way that plaintiffs in antitrust cases are required to support allegations of competitive harm by demonstrating (often via evidence about cross-elasticity of demand) that they compete with the defendant in a relevant product or service market. Another good example is the burden of proof that must be discharged by plaintiffs pressing trademark confusion claims in Lanham Act cases.[58] These plaintiffs are required to introduce evidence that consumers actually are confused when presented with a senior mark and a similar junior mark—i.e., direct evidence of the kind of harm that the trademark law seeks to prevent.[59]

Similarly, in cases where the plaintiff claims preemption of a derivative market or a licensing market, the impugned derivative work may represent a market that the original work's author may wish to enter (as either an author or a licensor), or it may not. In cases involving alleged preemption of licensing opportunities, plaintiffs should be put to proof that they have exploited similar licensing markets in the past, or that similarly-situated rightsholders have done so, or that they stand ready and have taken steps to do so. A bare desire to collect the rents that the defendant has collected is not enough.

Plaintiffs should also be required to demonstrate that the licensing market at issue is one that would be economic for them to enter—i.e., that the transaction costs of licensing would not overwhelm the likely revenues that might be obtained. This is an inversion of Wendy Gordon's suggestion, in a classic article, that the fair use doctrine should

58. *See* 15 U.S.C. § 1125(a) (2006).

59. *See* In re E.I. DuPont DeNemours & Co., 476 F.2d 1357, 1361 (C.C.P.A. 1973) (setting out factors relevant to proof of likelihood of trademark confusion).

free uses from copyright liability where the transaction costs of licensing overwhelm the value of the use.[60] Working within copyright law's current structure, the burden of establishing this market failure rests on the defendant as part of the fair use analysis. Under the revised structure described here, I would place a burden to establish "market viability" on the plaintiff as part of his prima facie case.

Moreover, and importantly, even if the derivative market is one which the author would be likely to enter, and even if licensing is otherwise economic, the plaintiff must show that this market is significant enough that, if viewed from the perspective of the author *ex ante* the creation of the original work, the freedom of another author to occupy that market with an unauthorized derivative would be likely to have an appreciable effect on the author's innovation incentives. For this showing to be made, the harm of market preemption must be substantial relative to the total expected return from the original work. Substantial harms are more likely to be foreseeable—but I should emphasize that the use of foreseeability that I advocate here is not the same as that advanced by Balganesh or Bohannan. I would not have the court determine whether the use was in fact objectively foreseeable *ex ante* the creation of plaintiff's work. I would, instead, focus on an importantly different counterfactual—i.e., whether the use, if in fact foreseen *ex ante*, would have meaningfully affected the author's creative incentives.

C. A Final Note on the Modesty of My Proposals

I should admit here, if it is not already obvious, that none of the changes to copyright law that I have proposed will make the ultimate question of what counts as harm in copyright less difficult or contentious. The modest but, I believe, nonetheless helpful, effect of my suggested changes to the structure of copyright would be to re-distribute the burdens of uncertainty. At present they rest almost entirely on defendants. Under my division and re-ordering of cases, some of the burden of uncertainty in copyright rule of reason cases would be transferred to plaintiffs.

The effect of this shift would be to shape the incentives of those copyright owners who are considering bringing suit. The burden of establishing harm will create expense and uncertainty for potential plaintiffs in copyright disputes classed under the rule of reason. Plaintiffs who believe that they have been substantially harmed, and who consider themselves in a position to offer evidence about harm, will come forward in the expectation that the burden of proving their case ultimately will be

60. 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).

borne by the defendant (under copyright law's provision for shifting costs and attorneys fees).[61] Those who are not substantially harmed, will not file suit—litigation costs and the necessity of proving harm will, for these plaintiffs, serve as a screen.[62] This is a desirable outcome—both limiting litigation and forcing prospective plaintiffs to think hard about their damages before filing suit. Of course we should worry about potential plaintiffs who believe that they have been harmed but cannot precisely quantify such harms. Nothing I have suggested requires precise quantification (or even nearly so). To make out a prima facie case in a copyright dispute governed by the rule of reason, plaintiffs will be obliged to describe their harm, and to show that it is substantial enough that it could have affected incentives if considered *ex ante*. Such evidence will usually be more readily available to plaintiffs than to defendants. It is therefore on plaintiffs—at least in rule of reason cases—that the burden of proof should fall.

These, then, are the structural shifts that should guide an efficient reform of copyright law. The basic principles are, however, friendly to at least two very different implementations. The first, and most direct, I have treated at length. It would be to require, as a predicate to liability in any case involving infringing conduct with ambiguous welfare effects, proof of actual or likely harm to authors' incentives.

The second would be, in some ways, conceptually more modest but nonetheless perhaps more demanding of plaintiffs in actual cases. It would be to limit damages and other remedies for conduct falling within the rule of reason. That is, while plaintiffs in copyright rule of reason cases would not be required to prove harm as a predicate to a liability finding, their recoveries would be limited to proved actual damages. Neither statutory damages nor attorney's fees would be recoverable, and injunctions would be available only where the plaintiff succeeds in showing harm irreparable via an award of damages.

The screening effect of the second implementation would be similar to the first. Only plaintiffs who have suffered substantial harms would bring suit for infringement falling within a rule of reason. Plaintiffs unable to show such harms would find the prospect of filing a lawsuit unattractive. This is precisely the sorting mechanism we would want for an efficient copyright system. This implementation would, however, require plaintiffs to prove their damages with greater precision.

---

61. *See* 17 U.S.C. § 505 (2006).

62. For a valuable explanation of the potential role of litigation costs in efficiently shaping incentives to sue in the case of incomplete contracts, see Albert Choi & George Triantis, *Completing Contracts in the Shadow of Costly Verification*, 37 J. LEG. STUD. 503 (2008).

CONCLUSION

The copyright law should be reformed to differentiate its treatment of conduct likely over the run of cases to lead to harm to authors' incentives from other types of conduct that have more ambiguous and context-dependent effects.

I have identified a category of "consumptive" uses as the type of infringement about which copyright need be most concerned. That conduct should be subject to a strong rule-based proscription. All other conduct should be treated according to a more sensitive standard, where plaintiffs have the opportunity to enjoin the conduct and recover damages when they can show that they have been harmed, but cannot do so otherwise.

These suggestions do not in themselves supply an answer to the most basic and enduring problem in copyright law—i.e., what counts as a relevant harm. They do, however, shape incentives in ways that will teach us more over time about the circumstances in which the many and varying forms of infringing conduct cause harm, and when they do not. This is valuable information the production of which the current structure of copyright law does not encourage.

SPECTRUM AUCTIONS AND THE PUBLIC INTEREST

ELLEN P. GOODMAN*

INTRODUCTION

Last year, the Federal Communications Commission (FCC) held its largest spectrum auction, selling exclusive rights to use coveted wireless frequencies for approximately $20 billion.[1] Not only was this the largest ever auction of spectrum, it was the largest ever single auction of public property in U.S. history.[2] Aside from its sheer magnitude, this auction of frequencies in the 700 MHz band was notable for other reasons, including the federal government's attempt to use the auction as a mechanism to value contested public policy goals.[3] In essence, the FCC

---

* Professor, Rutgers University School of Law – Camden. My thanks to Phil Weiser for encouraging me to present these ideas at Silicon Flatirons and to symposium participants for their comments.

1. Saul Hansell, *Verizon and AT&T Win Big in Auction of Spectrum*, N.Y. TIMES, Mar. 21, 2008, at C3 (reporting generally on results of auction, noting that Verizon Wireless won licenses in an auction that raised $19.12 billion after more than seven weeks of secret bidding).

2. J.H. Snider, *Is the Spectrum Just too Complex for Reporters?*, NEIMAN WATCHDOG, Feb. 21, 2008, http://www.niemanwatchdog.org/index.cfm?fuseaction=ask_this.view&askthisid=00327&stoplayout=true&print=true.

3. Service Rules for the 698-746, 747-762 and 777-792 MHz Bands, *Second Report &*

acknowledged that its policy goals for use of the spectrum involved tradeoffs, and that pursuing one of its principal goals might exact a toll in auction revenue. The FCC set out to ascertain, for the first time ever, just how much a policy goal would cost in foregone auction revenue and vowed to give up the goal if it cost too much.[4]

This use of auctions as a heuristic for valuing public interest goals raises interesting questions about the relationship between markets and policy, and between government as a proprietor of public resources and as a regulator of those resources. In this Article, I argue that it is possible to use auction results to inform the policy process without elevating revenue goals over other public policy objectives. In the 700 MHz auction, however, the FCC misunderstood what information auctions can yield and then failed to design an auction that would supply even that information. Correcting these problems for the spectrum auctions of the future – what may be the last great "land rush" to obtain wireless resources valued at more than $1 trillion – would lead to a more rational, transparent, and equitable communications policy.

Part I below shows how the FCC attempted to use auctions to evaluate communications policy goals in the 700 MHz proceeding. It is sometimes suggested that substantive communications policy goals, such as competition and innovation, should not be permitted to intrude on an otherwise neutral, market-based auction process. This ideal of value-free license assignments is neither possible nor desirable, I argue in Part II. Rather, the assignment process that culminates in auctions is invariably shot through with substantive communications policy goals. Auction results serve both signaling and substantive functions and these should be exploited. Auctions can be structured to reveal private valuations of regulatory burdens that the FCC must otherwise only guess at, thereby aiding in a regulatory process that accounts for policy tradeoffs. Of course auctions can also be structured to yield more or less revenue for the purchase of communications policy objectives. Part III shows the flaws in the FCC's first attempt to use auctions as a sophisticated policy tool and how this attempt is instructive for future policymaking.

I. 700 MHZ AUCTION

The expansion of broadband communications capacity and services

Order, 22 FCC Rcd. 15,289 (2007), http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-07-132A1.pdf [hereinafter *700 MHz Order*].

4. Ultimately, the revenue goal for the auction was met, so the FCC never had to compromise on its policy goal, nor to assess its cost, using the re-auction procedure established in case the auction revenue target was not met. *See* Susan Crawford, *700 MHz C Block Reserve Price Met*, PUBLIC KNOWLEDGE, Jan. 31, 2008, http://www.publicknowledge.org/node/1376.

is among the chief goals of U.S. communications policy.[5] One of the main impediments to better and more plentiful mobile broadband service is a shortage of wireless spectrum.[6] It is thus a significant event whenever the FCC makes additional spectrum available to wireless service providers. No such event has generated more interest than the reallocation of 700 MHz spectrum from television broadcasting to mobile wireless services.[7] This 700 MHz reallocation culminated in a January 2008 spectrum auction – what one FCC Commissioner called the "auction of the century."[8] The FCC auctioned off 1099 licenses covering 62 MHz of what is known as "low band" spectrum – frequencies that are particularly well suited for mobile wireless services.[9]

Because of the importance of the spectrum, interested parties lobbied intensively to get the FCC to structure the licenses and subsequent auctions in ways that would achieve their objectives.[10] As is

---

5. *See generally* JONATHAN E. NEUCHTERLEIN & PHILIP J. WEISER, DIGITAL CROSSROADS 23 (2007) [hereinafter DIGITAL CROSSROADS]; Appropriate Framework for Broadband Access to the Internet Over Wireline Facilities, *Policy Statement*, 20 FCC Rcd. 14,986 (2005), http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-05-151A1.pdf.

6. *See* Thomas W. Hazlett, *The Wireless Craze, the Unlimited Bandwidth Myth, the Spectrum Auction Faux Pas and the Punchline to Ronald Coase's "Big Joke": An Essay on Airwave Allocation Policy*, 14 HARV. J. L. & TECH. 335, 471 (2001) (quoting former FCC official Rudy Baca as forecasting "*chronic spectrum shortages*" that threaten "U.S. leadership in innovation and growth of broadband digital voice, data, and video wireless services." Press Release, Rudy L. Baca, Precursor Group Research, U.S. Disadvantaged by Spectrum Scarcity (July 25, 2000), http://www.precursorgroup.com (emphasis in original)); Comments of ArrayComm, Inc., to the *Request for Comments on Deployment of Broadband Networks & Advanced Telecomms. Servs.*, Dkt. No. 011109273-1263-01, RIN 0660-XX13 (Nov. 19, 2001) ("the most immediate barrier to wireless broadband deployment [is] the lack of available spectrum."), http://www.ntia.doc.gov/ntiahome/broadband/comments/arraycomm.html.

7. Another aspect of the proceeding dealt with a perplexing and critically important communications problem: The lack of a nationwide broadband network over which first responders (e.g., fire and police) can communicate. The FCC allocated spectrum for a public-private partnership whereby a private entity would build out a nationwide network for public safety use and, in return, have access to dedicated public safety spectrum at times when emergency communications were unnecessary. *See 700 MHz Order*, *supra* note 3, at 15,428. The complexity of this proposed partnership, combined with the credit crunch, deterred private entities from bidding for the spectrum. Because the reserve price was not met, the spectrum was not assigned, and the FCC will now have to decide how it will assign rights to the spectrum. Press Release, FCC, FCC Delinks 700 MHz Upper D Block from Other Blocks, Will Release Information on 700 MHz Auction Winning Bidders (Mar. 20, 2008), http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-280948A1.pdf.

8. *700 MHz Order*, *supra* note 3, at 15,571 (statement of Comm'r Robert M. McDowell, approving in part, dissenting in part).

9. *Id.* at 15,316; *see also* Philip J. Weiser & Dale Hatfield, *Spectrum Policy Reform and the Next Frontier of Property Rights*, 15 GEO. MASON L. REV. 549, 577-78 (2007) (describing the propagation characteristics of the 700 MHz band that make it so desirable for wireless communications, namely that signals are able to travel long distances, penetrate walls, and navigate around buildings and other obstructions).

10. *See generally* Susan P. Crawford, *Radio and the Internet,* 23 BERKELEY TECH. L.J. 933 (2008).

usually the case in FCC rulemakings, all parties contended that their favored policies would advance public interest goals – in this case, competition, innovation, and the provision of affordable broadband service.[11] The FCC, for its part, adopted largely compromise positions in designing the licenses to be auctioned and the rules governing wireless operation in the band.[12]

The most controversial rule was a requirement that those winning the largest licenses must abide by "open platform conditions." These conditions grew out of one of the key communications policy quandaries of the digital era, known as the "net neutrality" debate. This debate poses the question of whether broadband network operators, who control cable, fiber, and wireless communications networks, should be required to act as common carriers in transmitting traffic over their networks. That is, should they be required to carry all traffic without discrimination or should they be left alone to control network traffic?[13]

A somewhat different, but related, question is whether consumers should have the "right to attach" devices of their choice to the networks or should operators be able to control what devices their networks will support?[14] Proponents of net neutrality argue that regulation is required to ensure unfettered consumer access to third party applications (such as web-based video services or search functionality) and devices.[15]

---

11. *Compare* Comments of Public Interest Spectrum Coalition to *Google's Motion to Condition Grant*, Report No. AUC-73, File No. 0003382444, May 9, 2008, http://www.newamerica.net/files/PISC-Google-Motion-Comments.pdf (favoring the imposition of open platform conditions on licensees in the name of innovation and the public interest), *with* Comments of Verizon Wireless to the *700MHz Order*, DA 07-3415, AU Docket No. 07-157 (Aug. 31, 2007), http://fjallfoss.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6519721231 (opposing the imposition of open platform conditions on licensees in the name of innovation and the public interest).

12. *See 700 MHz Order*, *supra* note 3, at 15,360.

13. For analyses of the net neutrality debate, see generally Philip J. Weiser, *The Next Frontier for Network Neutrality*, 60 ADMIN. L. REV. 273 (2008); Brett M. Frischmann & Barbara van Schewick, *Network Neutrality and the Economics of an Information Superhighway: A Reply to Professor Yoo*, 47 JURIMETRICS J. 383 (2007); Tim Wu, *Why Have a Telecommunications Law?*, 5 J. ON TELECOMM. & HIGH TECH. L. 15 (2006).

14. *See, e.g.,* Tim Wu, *Wireless Net Neutrality: Cellular Carterfone and Consumer Choice in Mobile Broadband* 5-9 (New Am. Found.: Wireless Future Program, Working Paper No. 17, 2007), http://www.newamerica.net/files/WorkingPaper17_WirelessNetNeutrality_Wu.pdf (advocating the extension to wireless networks of the *Carterfone* rules, which mandated that AT&T permit consumers to attach devices of their choosing to the wired telephone network); Skype Commc'ns. S.A.R.L., Petition to Confirm a Consumer's Right to Use Internet Commc'ns Software and Attach Devices to Wireless Networks, RM-11361 (Feb. 20, 2007), *available at* http://svartifoss2.fcc.gov/prod/ecfs/retrieve.cgni?native_or_pdf=pdf&id_document=6518909730 (proposing the same).

15. *See, e.g.*, *Net Neutrality*: *Hearing Before the S . Comm. on Commerce, Science and Transp.*, 109th Cong. 9-14 (2006) (statement of Vinton G. Cerf, Vice President and Chief Internet Evangelist, Google, Inc.) (

Opponents counter that network operators have no incentives to get in consumers' way unless doing so is necessary to manage network quality or to support investments in network upgrades.[16]

Although in some respects quite technical, the net neutrality debate has profound implications for digital communications. What net neutrality proponents fear is a world in which several powerful companies are able to pick and choose what kinds of services – and therefore what kinds of expression – are privileged on their networks. In such a world, those who create innovative Internet applications and services will have to strike deals with the network operators before being able to reach consumers in a meaningful way. By contrast, what network operators fear is a world in which the government "dumbs down" their networks, preventing them from offering different levels of service or managing their networks efficiently. At stake is who gets to exercise what forms of control over communications traffic in the digital era and what role the government has in framing and securing a healthy communicative sphere.

Prior to the 700 MHz proceeding, the FCC had refrained from imposing net neutrality requirements, although in 2005, it did adopt non-binding net neutrality principles.[17] In the 700 MHz proceeding, the FCC went farther. It was moved by evidence "that wireless service providers are blocking or degrading consumer-chosen hardware and applications without an appropriate justification" and therefore decided "to take a measured step to encourage additional innovation and consumer choice at this critical stage in the evolution of wireless

---

> The Internet's open, neutral architecture has proven to be an enormous engine for market innovation, economic growth, social discourse, and the free flow of ideas. The remarkable success of the Internet can be traced to a few simple network principles – end-to-end design, layered architecture, and open standards – which together give consumers choice and control over their online activities.

).

16. *700 MHz Order*, *supra* note 3, at 15,358; *see also* Christopher S. Yoo, *Network Neutrality and the Economics of Congestion*, 94 GEO. L.J. 1847, 1852-53 (2006) (arguing that networks owners that manage their networks in a way that harms consumers will be at a competitive disadvantage); Letter from John T. Scott, III, Vice President & Deputy General Counsel, Verizon Wireless, to Marlene H. Dortch, Secretary, FCC (July 24, 2007), http://fjallfoss.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6519560209; Reply Comments of Verizon Wireless, RM-11361, Apr. 30, 2007, at 5, http://fjallfoss.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6519411455 (arguing that inter-brand competition prevents wireless providers "engage[ing] in any conduct that would result in the loss of customers.").

17. Michael K. Powell, Chairman, FCC, Preserving Internet Freedom: Guiding Principles for the Industry, Remarks at the University of Colorado Silicon Flatirons Symposium on The Digital Broadband Migration: Toward a Regulatory Regime for the Internet Age (Feb. 8, 2004), http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-243556A1.pdf (stating that broadband users should have the unfettered ability to access content, use applications, attach personal devices, and obtain service plan information).

broadband services, by removing some of the barriers that developers and handset/device manufacturers face in bringing new products to market."[18] In order to "promote innovation" on the largest and most desirable block of spectrum (known as the C Block), the FCC imposed conditions "to provide open platforms for devices and applications."[19]

The open platform conditions the FCC adopted go some distance toward implementing net neutrality mandates. They require a licensee to permit consumers to use any wireless device (e.g., an iPhone) on the network so long as the device causes no harm.[20] Moreover, a licensee may not block consumer access to, or otherwise discriminate against, particular applications (e.g., WiFi) unless it is necessary to do so to manage the network.[21] These conditions were incorporated into the licenses that were auctioned, thus helping to define the set of rights that an entity buys when it is the winning bidder.

By adopting the open platform conditions, the FCC was taking a highly controversial step that was opposed by the incumbent wireless operators thought to be the most likely (and ultimately the actual) winners of the C Block licenses, Verizon Wireless and AT&T.[22] The conditions were supported by entities predicted to be the most serious challengers to the incumbents, especially Google.[23] Indeed, Google informed the FCC that it would commit to bid in the auction and meet the FCC's announced reserve price *only* if the FCC adopted open

---

18. *700 MHz Order*, *supra* note 3, at 15,363.

19. *Id*.

20. Simon Wilkie, Open Access for the 700 MHz Auction: Wholesale Access Licensing and Could Increase Auction Revenue, NEW AM. FOUND., July 23, 2007, http://www.newamerica.net/publications/policy/open_access_700_mhz_auction (a licensee may not impose "prohibitions against devices that may be connected to the network so long as the devices are compatible with, and do not harm, the network").

21. *700 MHz Order*, *supra* note 3, at 15,363 (a licensee "may not block, degrade, or interfere with the ability of end users to download and utilize applications of their choosing on the licensee's C Block network, subject to reasonable network management.").

22. *See, e.g.,* Letter from John T. Scott, III, Vice President & Deputy General Counsel of Regulatory Law, Verizon Wireless, to Marlene H. Dortch, Secretary, FCC, Skype Communication's Petition, RM-11361 (Aug. 28, 2007), http://fjallfoss.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6519708296 (arguing that open platform conditions are unnecessary in a competitive environment, would unduly burden wireless operators, and would depress auction values).

23. *See* Letter from Richard Whitt, Washington Telecom and Media Counsel, Google, Inc., to Marlene H. Dortch, Secretary, FCC, WT Dkt. No. 06-150 (July 9, 2007), *available at* http://fjallfoss.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6519548049. Similar conditions were first proposed by Frontline Wireless, the principal proponent to build the joint public-safety/commercial network. Comments of Frontline Wireless, LLC, PS Dkt. No. 06-229 30 (Feb. 26, 2007), http://fjallfoss.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6518908196 (proposing that operators using a portion of the spectrum be required to support "open devices . . . open services and content, [and] . . . open offerings [wholesale or roaming]").

platform conditions.[24]

The FCC approved the open platform conditions by a vote of 4-1 with obvious trepidation.[25] There were good policy arguments, proffered not only by potential licensees, but also by public interest groups, that the open platform conditions would spur innovation and competition.[26] Google's support of the conditions made them more appealing by promising to enlarge the group of prospective bidders for the spectrum, thereby increasing bidding activity, and raising auction yield.[27] On the other hand, the incumbents submitted credible evidence that the conditions would mire the FCC in continual oversight of a competitive industry and would deter investment in wireless broadband.[28] Furthermore, they argued that open platform conditions would depress auction revenue.[29]

The FCC acknowledged the complexity of the public interest calculation and the statutory requirement that it balance potentially conflicting objectives, including innovation and the recovery of value from the spectrum to be auctioned.[30] The open platform requirements, it

---

24. Letter from Eric Schmidt, Google, Inc., to Kevin Martin, Chairman, FCC, WC Dkt. No. 06-150 (July 20, 2007), http://www.google.com/intl/en/press/pressrel/20070720_wireless.html. In fact, Google had proposed two conditions in addition to the no-locking, no-blocking conditions that the FCC actually adopted, and Google insisted that all four be adopted. *Id.* ("[S]hould the Commission expressly adopt the four license conditions requested in our July 9th letter – with specific, enforceable, and enduring rules – Google intends to commit a minimum of $4.6 billion to bidding in the upcoming auction.").

25. *700 MHz Order*, *supra* note 3, at 15,386 (discussing and rejecting Google's broad open platform proposal).

26. *Id.* at 15,358-61.

27. Cone of Silence (Finally) Lifts on the Spectrum Auction, GOOGLE PUBLIC POLICY BLOG, Apr. 3, 2008, http://googlepublicpolicy.blogspot.com/2008/04/cone-of-silence-finally-lifts-on.html.

28. *See, e.g.,* CTIA, 700MHz Spectrum Auction, http://www.ctia.org/advocacy/policy_topics/topic.cfm/TID/2; 700Mhz Statement, VERIZON POLICY BLOG, July 26, 2007, http://policyblog.verizon.com/PolicyBlog/Blogs/policyblog/DavidFish9/337/700MHz-statement.aspx.

29. Google itself asserted that the open platform conditions would reduce auction revenue. *See* Letter from Richard Whitt, *supra* note 23 (asserting that open platform conditions would drive down the price for the spectrum being auctioned).

30. *700 MHz Order*, *supra* note 3, at 15,368*; see also* U.S. Airwaves, Inc. v. FCC, 232 F.3d 227, 234 (D.C. Cir. 2000) (quoting Fresno Mobile Radio, Inc. v. FCC, 165 F.3d 965, 971 (D.C. Cir. 1999)) (recognizing competing statutory goals contained in the FCC's auction authority, 47 U.S.C. § 309(j)(3) (2006), and that a "regulatory decision in which the Commission must balance competing goals is . . . valid if the agency can show that its resolution 'reasonably advances at least one of those objectives and [that] its decisionmaking process was regular.'"); Melcher v. FCC, 134 F.3d 1143, 1154 (D.C. Cir. 1998) (recognizing that even within one of the 47 U.S.C. § 309(j)(3) objectives – subsection (B) – Congress set forth "a number of potentially conflicting objectives" and that the Commission can choose which to privilege) (citing MobileTel, Inc. v. FCC, 107 F.3d 888, 895 (D.C. Cir. 1997)).

predicted, "may result in a net gain of efficiency, given the potential that it holds for encouraging the development of new and innovative devices and applications in connection with such spectrum use."[31] On balance, it concluded that this potential gain outweighed "whatever possible negative effect [the open platform conditions] have with respect to the other [statutory] objectives" and that the benefits of such conditions justify any "potential for reducing the monetary value and decreasing efficient use of spectrum in some respects…."[32]

Despite this confident conclusion, the FCC was concerned that the incumbent wireless providers would be proved right and the open platform conditions would result in an auction that underperformed, leaving the FCC open to charges that it had mismanaged the sale of a great public asset. It therefore decided to hedge its bets by doing something it had never done in the past. It announced that if bidders for the C Block licenses failed to meet the FCC-adopted reserve price of $4.6 billion, the agency would remove the open platform conditions and re-auction the licenses.[33] In addition, the FCC vowed to change the geographic area and channel size of the licenses in any re-auction to make the licenses cheaper and thus more attractive to a larger set of potential bidders.[34]

It is difficult to know exactly what the agency's reasoning was in adopting this novel procedure because, disturbingly, there was no notice and comment on it.[35] In its order, the FCC merely stated that the re-auction procedure would "address the possibilities that license conditions adopted today significantly reduce values bidders ascribe to those licenses and/or have unanticipated negative consequences."[36] The Order said very little about how the FCC conceived of the relationship between the competitive bidding process and the valuation of license conditions. One can only conclude that the FCC wanted open platform conditions so long as they did not cost too much. In other words, the agency was acknowledging that the innovation and competition to be gained through open platform conditions might exact a price in auction revenues that was too great to bear.

Before addressing the FCC's use of the auction mechanism in 700

---

31. *700 MHz Order*, *supra* note 3, at 15,368 (discussing the requirements of 47 U.S.C. § 309(j)(3)(D) that the FCC foster the most efficient and intensive use of the spectrum).

32. *Id*. at 15,368.

33. *Id.* at 15,399. For the FCC's statutory authority to adopt reserve prices for spectrum licenses, see 47 U.S.C. § 309(j)(4)(F); 47 C.F.R. § 1.2104(c) (2006).

34. *700 MHz Order*, *supra* note 3, at 15,402-03.

35. The Notice of Proposed Rulemaking did not propose the re-auction procedure. At some point during the administrative deliberations, the FCC did let parties know about its intent to adopt the procedure and there were *ex parte* comments on the matter. *See, e.g.*, *700 MHz Order*, *supra* note 3; Letter from Richard Whitt, *supra* note 23 .

36. *700 MHz Order*, *supra* note 3, at 15,402.

MHz policymaking, there is a more basic question to consider: The proper relationship between auction revenue and the public interest.

II. AUCTIONS AND THE PUBLIC INTEREST

After the 700 MHz auctions concluded in March 2008, Congress held oversight hearings to investigate their outcome.[37] Members criticized the FCC for having sold the C Block spectrum at a bargain price – a discount they attributed to the open platform conditions.[38] In addition to the C Block, the FCC had also auctioned smaller licenses of reduced frequency size and geographic scope in other blocks of the 700 MHz spectrum. Because of the difference in license size and composition, it is hard to make an apples-to-apples comparison of the prices paid. But according to the conventional measurements of spectrum value, the C Block spectrum sold for about one-third of the unit price of the spectrum that was not burdened by open platform conditions.[39]

Should policymakers care about the prices that spectrum fetches? The Communications Act forbids the FCC from regulating spectrum licenses, and designing auctions, for the purpose of maximizing auction revenue, rather than pursuing non-fiscal public interest goals.[40] At the same time, the law mandates that the FCC pursue as one public interest objective "recovery for the public of a portion of the value of the public spectrum resource made available for commercial use…."[41] As many commentators have recognized, the single-minded pursuit of revenue

37. For general information regarding the hearings, see Committee on Energy and Commerce, Oversight of the Federal Communications Commission – The 700 MHz Auction, http://energycommerce.house.gov/index.php?option=com_content&task=view&id=237&Itemid=106; *see also Oversight of the Federal Communications Commission – The 7 [sic] MHz Auction: Hearing Before the H. Subcomm. on Telecomm. & the Internet*, 110th Cong. Rec. D437 (daily ed. Apr. 15, 2008) (statement before Committee on Energy and Commerce) (Witness prepared testimony is available at http://energycommerce.house.gov/index.php?option=com_content&task=view&id=237&Itemid=106) [hereinafter *Auction Hearings*].

38. *Auction Hearings, supra* note 37 (statement of Rep. Stearns, Ranking Member, House Comm. on Energy and Commerce) (citing "other studies" valuing the C Block "anywhere up to $30 billion" and Commissioner McDowell stated that the auction "could have done better.").

39. *Auction Hearings, supra* note 37 (statement of Harold Feld, Senior Vice President, Media Access Group, on behalf of the Public Interest Spectrum Coalition) (citing AT&T claims that it paid roughly $2.68 MHz/Pop for B Block licenses rather than the roughly $0.76 MHz/Pop that Verizon Wireless paid for C Block licenses to avoid the open platform conditions – a reduction of nearly $1.90 MHz/Pop).

40. 47 U.S.C. § 309(j)(7)(A) ("[T]he Commission may not base a finding of public interest, convenience, and necessity on the expectation of Federal revenues from the use of a system of competitive bidding . . . .").

41. *Id.* § 309(j)(3)(c).

maximization would result in poor communications policy.[42] What has been less commented upon is how the FCC should balance among competing public interest values in the use of spectrum, including value recovery.

This Section argues that the process of defining spectrum rights is never neutral, nor is the process of auctioning them. Auction and license design share responsibility with the "market" for picking "winners" and "losers" in the contest for spectrum rights.[43] Given this, auctioning licenses with an eye towards revenue capture as one of many goals is not a distortion of neutral licensing practices, but of a piece with policymaking. In addition, consideration of revenue generation as one among competing policy values is appropriate because auction results can supply useful information about the costs of regulatory burdens. This information can then be fed into the public interest balance, making it more efficient and transparent.

A. Auctions as a Market Allocation Tool

For nearly fifteen years, auctions have been the principal mechanism used to assign exclusive rights to use the spectrum for wireless communications.[44] There is almost universal agreement that auctions are superior to other FCC methods of license assignment.[45] Before Congress granted the FCC auction authority, the agency had assigned licenses by lottery and by comparative hearing.[46] Throughout the second half of the twentieth century, economists urged the FCC to abandon these methods in favor of auctions.[47] The argument was that competitive bidding would

---

42. *See, e.g.,* Thomas W. Hazlett, *Liberalizing US spectrum allocation*, 27 TELECOMMS. POL'Y 485, 492 (2003) ("a pre-occupation with government revenue extraction leads to anti-consumer policies."); Glen O. Robinson, *Spectrum Property Law 101*, 41 J. L. & ECON. 609, 621 (1998) (using auctions "as a means of filling a depleted treasury . . . has the effect of making communications policy a simple tool of fiscal policy, probably to the detriment of both.").

43. *See, e.g.*, *Auction Hearings, supra* note 37 (statement of Harold Feld) (criticizing the FCC for its refusal to "pick winners" in designing the 700 MHz auction and, as a result, allowing the best-capitalized incumbents to win).

44. *See* FCC, About Auctions (Aug. 8, 2006), http://wireless.fcc.gov/auctions/default.htm?job=about_auctions.

45. *See, e.g.*, Peter Cramton, *The Efficiency of the FCC Spectrum Auctions,* 41 J.L. & ECON. 727, 728 (1998).

46. *See* STUART MINOR BENJAMIN ET AL., TELECOMMUNICATIONS LAW AND POLICY 144-46 (2001) [hereinafter TELECOMMUNICATIONS LAW AND POLICY].

47. *See, e.g.,* Ronald H. Coase, *The Federal Communications Commission*, 2 J. L. & ECON. 1, 30 (1959) (urging the FCC "to dispose of the use of a frequency to the highest bidder, thus leaving the subdivision of the use of the frequency to subsequent market transactions."); David Porter & Vernon Smith, *FCC License Auction Design: A 12-Year Experiment*, 3 J.L. ECON. & POL'Y 63 (2006) ("Economists have long argued that auctions would promote efficiency in various ways, including the reduction of rent seeking and the avoidance of transaction costs used to reassign licenses in secondary markets."); *see also* Thomas W. Hazlett, *Assigning*

deliver initial entitlements to use the spectrum to those who valued them most.[48] Even in the absence of auctions, transactions in secondary markets for wireless assets could transfer licenses to their highest valued use, but these transactions entailed significant costs and sacrificed desirable efficiencies.[49]

In 1993, seeking to ensure that the FCC allocated spectrum efficiently for its most productive uses,[50] Congress gave the agency the authority to auction spectrum licenses in cases of mutually exclusive license applications.[51] Four years later, Congress made auctions obligatory for most commercial services.[52]

What is auctioned is a license to transmit an electrical signal over a particular frequency band at a particular power in a certain geographic area. Auctions determine who gets the initial entitlements to use spectrum, which may be sold thereafter subject to continued regulatory oversight. Licenses are for a limited period of time, typically 10-15 years, but are usually subject to renewal and function effectively as permanent entitlements.[53] As the law stands today, the FCC *must* auction spectrum when there are mutually exclusive applications for any initial license to provide a commercial service, unless the spectrum use is one of several

---

Property Rights to Radio Spectrum Users: Why Did FCC License Auctions Take 67 Years?, 41 J. L. & ECON. 529 (1998); Gregory L. Rosston & Jeffrey S. Steinberg, *Using Market-Based Spectrum Policy to Promote the Public Interest*, 50 FED. COMM. L. J. 87 (1997).

48. *See* Clayton P. Gillette & Thomas D. Hopkins, *Federal User Fees: A Legal and Economic Analysis*, 67 B.U. L. REV. 795, 805 (1987) (An alternative to auctions would have been spectrum user fees that "can successfully ration limited supplies of currently available goods and services to more highly valued uses, signal whether particular output levels should increase or decrease, avert wasteful usage, and encourage use of more suitable substitutes" as "an alternative to first-come, first-served, to lotteries, and to administrative judgment."); *see generally* EVAN KWEREL & ALEX FELKER, FCC OFF. FOR PLANS & POL'Y, USING AUCTIONS TO SELECT FCC LICENSEES (1985), http://www.fcc.gov/Bureaus/OPP/working_papers/oppwp16.pdf..

49. *See* Coase, *supra* note 47.

50. H.R. REP. NO. 103-111, at 253 (1993), *reprinted in* 1993 U.S.C.C.A.N. 378, 580 ("a carefully designed system to obtain competitive bids from competing qualified applicants can speed delivery of services, [and] promote efficient and intensive use of the electromagnetic spectrum").

51. Omnibus Budget Reconciliation Act of 1993, Pub. L. No. 103-66, § 6002, 107 Stat. 312, 388 (1993) ("If mutually exclusive applications are accepted for filing for any initial license or construction permit which will involve a use of the electromagnetic spectrum. . . then the Commission shall have the authority . . . to grant such license or permit to a qualified applicant through the use of a system of competitive bidding that meets the requirements of this subsection.").

52. Balanced Budget Act of 1997, Pub. L. No. 105-33, § 3002, 111 Stat. 251, 258 (1997) (codified at 47 U.S.C. § 309(j)) ("If . . . mutually exclusive applications are accepted for any initial license or construction permit, then, except as provided [herein], the Commission shall grant the license or permit to a qualified applicant through a system of competitive bidding that meets the requirements of this subsection.").

53. *See* Eli Noam, *Spectrum Auctions: Yesterday's Heresy, Today's Orthodoxy, Tomorrow's Anachronism. Taking the Next Step to Open Spectrum Access*, 41 J.L. & ECON. 765 (1998).

enumerated exceptions.[54]

B. Auction Revenue as Compensation

As important as efficiency was in the adoption of spectrum auctions, another motivation was equally powerful: the desire to capture revenue from commercial use of the spectrum. One of the problems with the private markets for spectrum licenses that were initially assigned by lottery or by hearing was that the revenue went to private parties, not the government.[55] By holding auctions for initial licenses, the government could capture this "windfall" for the public at the same time that it facilitated an efficient allocation of the spectrum resource. It was no accident that Congress gave the FCC auction authority during the budget crisis of the early 1990s when the desire to monetize public assets was particularly keen.[56] The legislative history of the auction legislation makes clear that the efficiency gains associated with spectrum auctions were of no more importance than the distributional effects, namely that the Treasury was compensated for licensee use of the spectrum.[57]

The law does not say exactly how much the pursuit of auction revenue should influence federal spectrum policy. In deciding when to use auctions, who is eligible to bid in them, and what the characteristics of the auctioned licenses should be, the FCC is instructed to pursue public interest objectives.[58] It is forbidden from merely equating the public interest with auction revenue.[59] And yet, one of the public interest objectives it must pursue is "recovery for the public of a portion of the value of the public spectrum resource made available for commercial

---

54. The FCC is not permitted to auction licenses for public safety radio services, for noncommercial educational or public broadcast stations, or for digital television service provided by incumbent television broadcast licensees. 47 U.S.C. § 309(j)(1)-(2). The FCC is also prohibited from auctioning licenses for satellite orbital slots or to provide international or global satellite communications services. *Id.* § 765f.

55. *See* Harold J. Krent & Nicholas S. Zeppos, *Monitoring Governmental Disposition of Assets: Fashioning Regulatory Substitutes for Market Controls*, 52 VAND. L. REV. 1703, 1735-36 (1999) ("lotteries drew fire for precipitating a secondary auction" in which licensees could sell their spectrum usage rights "in the open market, reaping windfalls at the expense of the public at large.").

56. *See* Noam, *supra* note 53. Although communications legislation ordinarily comes out of the Congressional Commerce Committees, the auction legislation was the product of the Budget Committees whose main interests lie in the management of money, not communications.

57. H.R. REP. NO. 103-111, at 253 (1993), *reprinted in* 1993 U.S.C.C.A.N. 378, 580 (spectrum auctions would "prevent unjust enrichment, and produce revenues to compensate the public for the use of the public airwaves").

58. 47 U.S.C. § 309(j)(3).

59. *Id.* § 309(j)(7)(A) (FCC "may not base a finding of public interest, convenience, and necessity on the expectation of Federal revenues").

use".[60] In other words, the FCC should seek to collect a fair return on use of the public airwaves.[61] It is fair to say that, when it comes to auctions, the FCC must operate between two extremes, neither ignoring auction revenue in the design of spectrum policy nor focusing exclusively on revenue generation.

The FCC has had little success in assessing the relative importance of monetary recovery as a public interest goal. In one of the rare cases in which it has addressed the question, it suggested that monetary recovery should take a back seat to other goals: The "most basic spectrum-management power is to assign spectrum to achieve public interest benefits *other than* monetary recovery."[62] Two sentences later, however, the FCC suggested that monetary recovery is one among equally important public interest "factors the Commission must consider in establishing bidding qualifications and license conditions."[63]

However important the FCC believes monetary recovery to be, the fact is that the agency is motivated by money. Whenever Congress orders the FCC to assign licenses in a particular block of spectrum, it also commissions a spectrum valuation from the Congressional Budget Office (CBO). For example, the CBO estimated the value of the auctionable 700 MHz spectrum to be $12.5 billion.[64] Although the FCC is not required to achieve the "score" that Congress has given a particular spectrum asset, this number becomes a target that the FCC tries to hit when it conducts its auction.[65] It is the rare FCC Chairman who can resist reveling in the delivery of a large check to the Treasury, or who wants to bear the obloquy of delivering a small one. As an independent agency, the FCC is sensitive to Congressional criticism and the legislature's members have roundly criticized the agency when spectrum

---

60. *Id.* § 309(j)(3)(C).

61. To this end, Congress eliminated the "pioneer's preference" program which had allowed the FCC to bypass the auction process by awarding a license to an especially innovative technological pioneer. *See, e.g., id.* § 309(j)(13)(F); Qualcomm, Inc. v. FCC, 181 F.3d 1370, 1380-81 (D.C. Cir. 1999) (when Congress withdrew the FCC's authority to grant pioneer's preferences, "its focus was on increasing federal revenues" by requiring the FCC to recover for the public a portion of the value of the spectrum); *FCC Pioneer Preference Policy: Hearing Before the H. Comm. on Energy and Commerce*, 103d Cong. 24 (1994) (statement of Rep. Edward Markey, Chairman, House Subcomm. on Telecomm. and Finance).

62. Improving Public Safety Communications in the 800 MHz Band, *Report & Order*, 19 FCC Rcd. 14,969, 15,019 (2004) (emphasis in original).

63. *Id.*

64. Letter from Douglas Holtz-Eakin, Director, Cong. Budget Off., to Thad Cochran, Chairman, Sen. Comm. on Appropriations (Dec. 20, 2005), at 3, http://www.cbo.gov/ftpdocs/69xx/doc6990/hr2863.pdf (providing the cost estimate for H.R. 2863, Department of Defense Appropriations Act, 2006).

65. *See, e.g., Auction Hearings, supra* note 37 (statement of Deborah Taylor Tate, Comm'r, FCC) ("Given that the Congressional Budget Office estimated auction receipts of $10 billion to perhaps as much as $15 billion, the [700 MHz] auction was clearly a financial success").

auctions have failed to produce as much revenue as expected or desired.[66] Charges of spectrum "giveaways" abound whenever the FCC distributes spectrum usage rights at zero or reduced price.[67] To some extent, these complaints are rooted in efficiency concerns about non-market allocations of resources. But they also reflect a more basic insistence on the equitable distribution of public resources and on public compensation for their benefits.[68]

In addition to its distributional value, auction revenue is an appropriate consideration in shaping spectrum policy because of what it can signal.

C. *Auction Revenue as a Signal*

In spectrum auctions, the FCC does not act solely as auctioneer. Its more important function is to define the property rights that are auctioned off. The "metes and bounds" of the spectrum right are identified by the frequency range that the license covers, the geographic scope of the license and the conditions under which the licensee can operate.[69] Under the Communications Act, the FCC must define these license terms in accordance with the "public interest."[70] It is well known

---

66. *See, e.g.,* Ellen P. Goodman, *Spectrum Rights in the Telecosm to Come*, 41 SAN DIEGO L. REV. 269, 300 (2004) (describing the perceived under-performance of an early auction and the Congressional response). When the FCC auctioned several television channels in 2002, Representative John Dingell, currently Chair of the House Commerce Committee, jeered that a "jackass out of a barn lot could have done a better job of selling this public property" and chastised the FCC for "a gross mismanagement of the spectrum." J.H. Snider, *The Art of Spectrum Lobbying: America's $480 Billion Spectrum Giveaway, How it Happened, and How to Prevent It From Recurring* 12 (New Am. Found.: Wireless Future Program, Working Paper No. 19, 2007), http://www.newamerica.net/files/art_of_spectrum_lobbying.pdf (citing Molly M. Peterson's account of the proceedings, in *House Panel Votes to Kill Deadline for Airwaves Auction*, TECHNOLOGY DAILY, May 2, 2002).

67. Snider, *supra* note 66, at 26-27.

68. *See generally* Ellen P. Goodman, *Spectrum Equity*, 4 J. Telecomm. & High Tech L. 217, 227-31 (2005) (describing the role of fairness in spectrum allocation and access).

69. As Phil Weiser and Dale Hatfield have pointed out, the spectrum right is not nearly as clearly defined as a real property right. Weiser & Hatfield, *supra* note 9, at 587.

70. Congress commanded federal communications regulators from the earliest days of radio regulation to administer "[r]adio [c]ommunication" as "a public utility… in the public interest." THOMAS G. KRATTENMAKER & LUCAS A. POWE, JR., REGULATING BROADCAST PROGRAMMING 9 (1994) (quoting *To Amend the Radio Act of 1912: Hearings on H.R. 11,964 Before the House Comm. on the Merchant Marine and Fisheries*, 67th Cong., 4th Sess. 32 (1926) (statement of Hon. Herbert Hoover, Secretary of Commerce)); *see also* 47 U.S.C. § 302a (2000) (requiring that the FCC rulemaking power over broadcasting must be exercised in "the public interest, convenience, and necessity"); *id.* § 303 (2000) (requiring that the FCC power to classify, license, and regulate radio must be "as public convenience, interest, or necessity requires"); *id.* § 303(g) (requiring that the FCC study new uses for radio that are "in the public interest"); *id.* § 307(a) (2000) (requiring that the FCC grant radio broadcast licenses "if public convenience, interest, or necessity will be served"); *id.* § 307(e)(1) (2000) (providing that the FCC may authorize certain types of radio broadcasting without a license if it "serves the public interest, convenience, and necessity").

and has been repeatedly shown that the "public interest" standard is highly malleable and has yielded few satisfying rules of decision in communications policy.[71] Because the public is interested in competing social and economic goals, including efficiency, competition, innovation, universal service, public safety, diverse programming, *and* auction revenue, the process of constructing a license involves policy tradeoffs from beginning to end.

Once the government articulates a public interest goal in connection with licensed spectrum, it can "pay for" that goal in one of three ways (1) impose requirements on licenses to be auctioned, presumably at a discounted price that reflects the costs of the requirements; (2) impose requirements on licenses in lieu of auction payments, thereby effectively discounting the licenses 100%;[72] or (3) auction licenses without public interest requirements and reinvest the proceeds to achieve the same goals. Suppose, for example, that the government was considering auctioning off broadcast licenses conditioned on the provision of at least one hour per day of local political programming. The FCC could mandate the programming as a license condition in lieu of an auction. It could impose the mandate and auction the licenses at a discount, or it could auction the licenses and pay out of pocket (or through tax subsidies) for the desired programming.

It is difficult for a government agency to assess the relative merits of these options without attaching a price to the public interest requirements imposed – the price of compliance for the regulated entities plus the indirect costs that the regulations might impose on third parties or on the economy as a whole. As a general matter, federal administrative agencies use cost-benefit analysis in decisionmaking because it guides, and makes more transparent, the selection of regulatory options.[73]

---

71. *See* Erwin G. Krasnow & Jack N. Goodman, *The "Public Interest" Standard: The Search for the Holy Grail*, 50 FED. COMM. L.J. 605, 606–08 (1998) (criticizing the public interest standard as too vague and fluid); Randolph J. May, *The Public Interest Standard: Is It Too Indeterminate to Be Constitutional?*, 53 FED. COMM. L.J. 427, 428–29 (2001) (arguing that the public interest standard violates the nondelegation doctrine); *see also* Hazlett, *supra* note 6, at 401-05 (criticizing the public interest standard in the spectrum allocation context).

72. One way to look at public interest broadcast regulation is as a quid pro quo for foregone auction revenue. An early proponent of auctions criticized this regulation in lieu of auctions as a "tax" on the public for a government "purchase" of public interest benefits. Thomas W. Hazlett, *The Rationality of U.S. Regulation of the Broadcast Spectrum*, 33 J.L. & ECON. 133, 170 (1990).

73. *See* Jennifer Nou, *Regulating the Rulemakers: A Proposal for Deliberative Cost-Benefit Analysis*, 26 YALE L. & POL'Y REV. 601, 613-15 (2008) (describing the strong commitment to cost-benefit analysis in all three branches of the federal government). President George W. Bush has gone so far as to issue an Executive Order requiring federal agencies to measure total annual costs and benefits for every proposed regulation. Exec. Order No. 13,422, 72 Fed. Reg. 2,763 (Jan. 18, 2007). For a comprehensive overview of cost-benefit analysis in policymaking, see generally CASS SUNSTEIN, THE COST-BENEFIT STATE: THE FUTURE OF

Indeed, it is for this reason that the FCC is required by law to consider the costs of its regulations where "small [business] entities" are affected.[74]

Because it lacks the wherewithal and resources to conduct its own research and because so many of the judgments that it makes are predictive, the FCC must rely on prospective licensees' assessments of the costs of public interest requirements. Often, regulated entities will forecast their own economic ruin in the event that proposed public interest requirements are adopted.[75] At other times, as in the 700 MHz proceeding, the predictions of doom will be more muted and vague, such as Verizon Wireless' prediction that open access requirements would exact a toll on wireless innovation and service.[76] These cost predictions may reflect honest assessments or they may be rent-seeking attempts to reduce encumbrances on spectrum entitlements.

The auction process can be helpful in flushing out parties' actual valuations of the costs (to them) of public interest requirements. The simultaneous auctions described in Part IV below would simulate a market for spectrum with assorted regulatory requirements. Provided that the FCC has developed a record on the appropriate price to be paid for a public interest goal, these valuations would provide useful input into the particular proceeding in which they are revealed. But even in the absence of such a record leading up to the auction, the data would improve future spectrum policy decisions by telling us, for example, whether and to what extent open platform conditions are likely to shackle wireless operations. This information about the actual cost to the bidders of policy choices would improve subsequent asset sales and public

---

REGULATORY PROTECTION (2002) [hereinafter THE COST-BENEFIT STATE].

74. The Regulatory Flexibility Act requires federal agencies to assess the economic impact of rules on "small entities." 5 U.S.C. § 603(b)(4) (2006). The analysis should consider alternatives "which accomplish the stated objectives of applicable statutes and which minimize any significant economic impact of the proposed rule on small entities." *Id.* § 603(c). No cost-benefit analysis is required so long as agencies investigate least cost alternatives in regulation. "[A]n agency may provide either a quantifiable or numerical description of the effects of a proposed rule or alternatives to the proposed rule, or more general descriptive statements if quantification is not practicable or reliable." *Id.* § 607. *See also* Alenco Commc'ns., Inc. v. FCC, 201 F.3d 608, 625 (5th Cir. 2000) (noting that the Regulatory Flexibility Act, 5 U.S.C. § 604(a)(5), "specifically requires 'a statement of the factual, policy, and legal reasons for selecting the alternative adopted in the final rule'" but does not require "cost–benefit analysis or economic modeling.").

75. *See, e.g.*, Seth Grossman, *Creating Competitive and Informative Campaigns: A Comprehensive Approach to "Free Air Time" for Political Candidates*, 22 YALE L. & POL'Y REV. 351, 376 n.110 (2004) (citing broadcaster testimony in response to 2001 legislative proposal that would require free advertising time for federal candidates that requirement would "'severely injur[e] a television station's ability to raise revenue'" and a National Association of Broadcasters' claim that free time would "'pose substantial financial burdens to the industry, and could even result in lay-offs of employees'").

76. Reply Comments of Verizon Wireless, WT Dkt. No. 06-150, June 4, 2007, http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6519516267.

interest debates.

D. *Auction Revenue as Substance*

Beyond its value as a source of public compensation and policy information, the most obvious public interest benefit of auction revenue is that it can buy communication services of interest to the public. Under current law, spectrum auction revenue is deposited in the federal Treasury and cannot be allocated to communications projects in the absence of special legislative authority.[77] For example, Congress has authorized the use of auction revenues to fund the provision of digital converter boxes to facilitate the transition from analog to digital broadcast television, and the purchase of public safety communications equipment by police and fire departments. [78]

Many more public interest objectives might be purchased with auction revenues. One purpose of the 700 MHz proceeding was to auction spectrum to a private entity that would subsidize the construction of a nationwide public safety network for interoperable emergency service communication.[79] Unfortunately, the spectrum block that was to be auctioned for this purpose – the D Block – failed to attract a bid over the reserve price and the licenses were not assigned. Prior to the 700 MHz auction, there had been calls for the federal government to use spectrum auction revenue to provide for a public safety network, rather than relying on license design.[80]

Of course, Congress could simply appropriate funds for a public safety network, or for any other communications policy objective, rather than relying on earmarks from auction revenue. Indeed, in the aftermath of the failed D Block auction, the then-Chairman of the FCC himself urged Congress to allocate funds for the network.[81] It turns out, however, that such appropriations have been difficult to come by.[82] Even for the

---

77. *See*, *e.g.*, 47 U.S.C. § 309(j)(8)(A) (directing 700 MHz auction revenue to be deposited in the federal Treasury).

78. Digital Television Transition and Public Safety Act of 2005, Title III of the Deficit Reduction Act of 2005, Pub. L. No. 109-171, 120 Stat. 4 (2006) (codified at 47 U.S.C. §§ 309, 337); *see also* LENNARD G. KRUGER & LINDA K. MOORE, CONG. RESEARCH SERV., THE DTV TRANSITION 6 (2005), http://digital.library.unt.edu/govdocs/crs/permalink/meta-crs-7682:1.

79. *700 MHz Order*, *supra* note 3, at 15,386.

80. *See* Jon. M. Peha, *The Digital TV Transition: A Chance to Enhance Public Safety and Improve Spectrum Auctions*, IEEE COMMS. MAGAZINE, June 2006, at 22, 23-4, http://www.ece.cmu.edu/~peha/DTV.pdf (proposing that 700 MHz auction revenue be used to fund a national public safety network).

81. *Auction Hearings, supra* note 37 (statement of Kevin J. Martin, Chairman, FCC).

82. Congress recently appropriated approximately $4.7 billion for broadband communications development in the American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-16, §§ 6001(b)(4), 6001(k)(2)(D) (2009). Public safety networks are eligible to receive grants under the programs created by this law, but there is no mandate that they be

provision of a public safety network – a relatively appealing and uncontroversial goal – the political will has been lacking. This is so notwithstanding the well-publicized public safety communications disasters of September 11 at the World Trade Center and the Katrina hurricane in New Orleans.[83] Auction revenue earmarks provide a more politically palatable way to funnel federal funds to communications projects. Public interest goals that reduce auction revenues, while perhaps ultimately worthwhile, should be assessed with the opportunity costs in mind.

E. Skewing the Auction Results

It can be argued that if regulators take expected revenue into account in structuring a spectrum auction, they will skew the auction towards certain outcomes and away from the auction's natural, market-determined course. This would indeed be problematic if revenue maximization became the be-all-and-end-all of communications policy, but it is much less worrisome if revenue is only one factor in what is necessarily a value-laden process of spectrum allocation. Indeed, revenue consideration in auctions is consistent with, not a deviation from, the public interest balancing that takes place throughout the process of spectrum allocation.

Spectrum policy involves the government in two functions – public interest regulation and the disposition of a public asset. As has been recognized in other contexts, "government property dispositions" are a form of "regulatory policymaking."[84] This is nowhere truer than in the spectrum context. The very existence of an exclusive right to use the spectrum – the existence of an asset that can be auctioned – is a government creation that embodies public interest judgments about industry structure and the public good. Moreover, the value of the spectrum entitlements when they are auctioned, unlike the value of physical assets like timber on federal lands, is entirely dependent on government choices about how the entitlement should be defined.[85]

---

principal recipients or receive any funding. *Id.*

83. Philip J. Weiser, *Communicating During Emergencies: Toward Interoperability and Effective Information Management*, 59 FED. COMM. L.J. 547, 547-48 (2007).

84. Krent & Zeppos, *supra* note 55, at 1747.

85. Property rights are always structured by government policies that create and enforce legal entitlements. Because of the nature of spectrum rights, however, government's role is especially intensive. The government does not define the properties of land it might auction, or other physical assets like timber and oil. The assets themselves have characteristics that pre-exist the regulatory structure. Zoning rules, or limitations on the extraction of natural resources, are imposed on extant assets. While regulatory interventions affect the value of the physical assets, the regulations are layered atop assets that have an independent existence and value.

The first step in any spectrum policy decision is the allocation of spectrum for a particular purpose.[86] Spectrum is divided into blocks, which are then further divided into channels of varying bandwidths. Spectrum allocation is the process of defining the bands that may be used for particular services. These allocations are often referred to as spectrum zoning.[87] Because many kinds of spectrum uses are incompatible with each other, the FCC must privilege some uses over others for each spectrum band: satellite in one, broadcasting in another, mobile wireless in a third, "mixed use" in a fourth.[88] Since the commercially usable spectrum in the United States has already been allocated for some purpose, this process is really a process of *re*-allocation, typically leading to contests among rival claimants for the spectrum.[89]

The FCC seeks to resolve these contests in the public interest. How the FCC frames competing public interest objectives will determine eligibility for spectrum entitlements. If the FCC concludes, for example, that the public interest in competition is paramount, it will allocate spectrum for services that it thinks will provide a competitive balance to incumbents.[90] If the agency is taken with the public interest in diverse speech, it might allocate spectrum for additional broadcast stations.[91] As potential uses of the frequencies change, the very entities that were given entitlements for one reason (e.g., to increase broadcast speech) become an obstacle to distributing entitlements for another reason (e.g., to increase broadband competition by freeing spectrum for new entrants).[92] It is the government, based on public interest considerations, that makes

---

86. 47 U.S.C. § 303(b)-(c) (authorizing the FCC to "[p]rescribe the nature of the service to be rendered by each class of licensed stations" and to "[a]ssign bands of frequencies to the various classes of stations"); *see also* TELECOMMUNICATIONS LAW AND POLICY, *supra* note 46, at 62.

87. *See, e.g.,* DIGITAL CROSSROADS, *supra* note 5, at 267; Goodman, *supra* note 66, at 282.

88. U.S. GEN. ACCT. OFF., BETTER COORDINATION AND ENHANCED ACCOUNTABILITY NEEDED TO IMPROVE SPECTRUM MANAGEMENT 3 (2002), http://www.gao.gov/new.items/d02906.pdf.

89. *Id.*

90. *See* Michael K. Powell, Chairman, FCC, Broadband Migration III: New Directions in Wireless Policy, Remarks at the University of Colorado Silicon Flatirons Symposium on Digital Broadband Migration (Oct. 30, 2002), http://www.fcc.gov/Speeches/Powell/2002/spmkp212.html (noting that the FCC's current conception of "the public interest must reflect the realities of the marketplace").

91. For example, the FCC may decide to allocate more spectrum for low power FM radio stations dedicated to commercial-free educational programming. *See generally* Creation of a Low Power Radio Service, *Third Report and Order*, 22 FCC Rcd. 21,912 (2007) (establishing rules and policies designed to foster growth in LPFM, especially within local groups such as schools, churches, and other community-based organizations).

92. *See, e.g.,* Unlicensed Operation in the TV Broadcast Bands, *Second Report and Order*, 46 Comm. Reg. (P&F) 940 (2008), http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-08-260A1.pdf (dealing with the obstacles television broadcasting poses to use of broadcast spectrum for unlicensed wireless broadband services).

the market for spectrum assigned by auction.

Allocation decisions not only *reflect* public interest determinations, but will *influence* public interest considerations "downstream" in the administrative process. How the FCC allocates spectrum will determine, for example, whether spectrum users can share the frequencies in a non-rivalrous fashion or whether they need exclusive licenses to operate. This allocation choice will determine whether or not spectrum rights are auctioned at all, or are given away. This is because, under the law, only mutually exclusive (or rivalrous) spectrum usage rights may be auctioned.[93]

The second step in the rulemaking process is the establishment of "service rules" for spectrum use. The limitations and obligations contained in these rules are incorporated into the spectrum licenses. Service rules may impose construction deadlines and service requirements to keep licensees from warehousing the spectrum. They may require licensees to interconnect with competitors or to provide emergency services, and impose other kinds of public interest service obligations.[94] Each of these mandates entails public interest tradeoffs, such as rural service at the expense of urban service or interconnection at the expense of network investment.

The service rule that has the most impact on a spectrum auction – the one that most directly defines the "metes and bounds" of the license – is the definition of the license size. In any given spectrum proceeding, the FCC chooses to create a few nationwide licenses, dozens of regional licenses, or thousands of smaller licenses. The decision about license characteristics reflects public interest choices and industry predictions. Smaller licenses (in terms of geography and bandwidth) tend to favor smaller players and/or new entrants as well as local services over national ones.[95] Larger licenses can be expected to have the opposite effect.

Given the pervasiveness of the regulatory power in the construction of spectrum licenses, it should be understood that there is nothing inevitable or "neutral" about a particular auction result.[96] All spectrum

---

93. *See* KWEREL & FELKER, *supra* note 48, at 2.

94. *See* TELECOMMUNICATIONS LAW AND POLICY, *supra* note 46, at 63.

95. *See* Porter & Smith, *supra* note 47, at 67 (describing the first major auction for PCS licenses in 1994 in which there were more than 2000 licenses auctioned off in some of the PCS bands).

96. This is even before one considers the ways in which spectrum design might bias the outcome in favor of certain bidders. *See, e.g.*, GREGORY ROSE, SPECTRUM AUCTION BREAKDOWN: HOW INCUMBENTS MANIPULATE FCC AUCTION RULES TO BLOCK BROADBAND COMPETITION 18-19 (New Am. Found., Wireless Future Program, 2007), http://www.newamerica.net/files/WorkingPaper18_FCCAuctionRules_Rose_FINAL.pdf; *see also* SIMON WILKIE, CTR. FOR COMM. L. & POL'Y, U. OF SO. CAL., SPECTRUM AUCTIONS ARE NOT A PANACEA: THEORY AND EVIDENCE OF ANTI-COMPETITIVE AND RENT-SEEKING BEHAVIOR IN FCC RULEMAKINGS AND AUCTION DESIGN 7-10 (2007),

auctions reflect a chain of decisions that, from beginning to end, incorporate regulatory values and priorities. At no point is this process value-free or driven purely by efficiency concerns. At every stage, the government makes decisions that will favor certain bidders over others and certain spectrum applications over others in what it claims is a vindication of the public interest.

III. THE AUCTION HEURISTIC

I have argued that it makes sense to use the auction process to produce information about the cost of public interest objectives, and that such information can ultimately be used in the policymaking process to assess tradeoffs among communications policy goals. This seems to be what the FCC wanted to accomplish in the 700 MHz auction. Unfortunately, the re-auction technique it created to trade off policy goals could not have accomplished the agency's stated goal. This Section shows why and concludes with some thoughts on how auctions might be structured to provide more useful inputs into the policymaking process.

A. The Problem With the 700 MHz Re-Auction Concept

Section I above described the FCC's assumption that the open platform conditions on C Block licenses would reduce the high bids for those licenses.[97] The FCC asserted that if bidders failed to meet the reserve price of $4.6 billion for the licenses, it would be because the conditions imposed "a greater negative impact on network operations" than the agency had predicted.[98] In that event, the FCC would have to change its "assessment of the net public interest benefit of imposing these requirements (*i.e.*, the benefit of fostering the development of innovative devices and applications vs. the potential negative effects on network operations)...."[99] The FCC would then immediately re-auction the C Block licenses without the open platform conditions, presumably to achieve a better balance between the benefits of innovation and the negative effects on network operations.

This procedure suffers from two fundamental errors (1) Auction results say very little about the benefits of a public interest condition and therefore cannot reveal the net benefits of the associated policy choice; and (2) auction results *can* say something about the costs of a public interest condition, but only if there are simultaneous auctions that

http://www.m2znetworks.com/xres/uploads/documents/Wilkie%202%20Auctions%20No%20Panacea%20Wilkie.pdf.

97. *See supra* Section I.

98. *700 MHz Order*, *supra* note 3, at 15,403.

99. *Id.*

control for a single variable and are able to measure the magnitude of devalued bids, as the FCC's reserve price trigger could not.

1. Measuring Benefits

For an auction to reveal the "net public interest benefit" of the open platform conditions, as the FCC asserted it could,[100] the auction process would have to quantify both the benefits and costs of imposing the open platform conditions.

One basic problem with this approach is that there is simply no way to calculate the actual or even putative value of open platform conditions.[101] By its nature, the fostering of innovative devices and applications produces widespread benefits that accrue to numerous players.[102] Many of these benefits are positive externalities to the auction transaction. Beneficiaries include the developers of applications and the manufacturers of devices that would have nondiscriminatory access to the C Block networks. In addition, proponents predict that the consumer liberty to attach devices to, and freely use applications on, the network will produce a consumer surplus.[103] These benefits, which we can call "X," are unquantifiable, at least *ex ante*. Google, or some other bidder, might well capture some of the value of X indirectly through greater broadband penetration or device usage, but not all of it.

Commentators have recognized the limitations of any cost-benefit analysis when it comes to quantifying the benefits of regulation.[104] Indeed, some of the most influential proponents of cost-benefit analysis concede that the analysis cannot identify the benefit-maximizing rule. Rather, it is a tool for generating information about some of the likely consequences of a proposal.[105] In other words, it is a heuristic for

---

100. *Id.*

101. *See generally* Susan Crawford, *The Internet and the Project of Communications Law*, 55 UCLA L. REV. 359, 391 (2007) (discussing the inability of net neutrality proponents to quantify the benefits of openness).

102. *See* Wu, *supra* note 14, at 2 (advocating, *inter alia*, open platform and open application requirements within the wireless industry to stimulate the development of new hardware and software).

103. *See, e.g.,* PETER CRAMTON ET AL., SUMMARY: REVENUES IN THE 700 MHZ SPECTRUM AUCTION 11 (2007), http://www.cramton.umd.edu/papers2005-2009/cramton-skrzypacz-wilson-e-block-plan-increases-revenues.pdf ("Essentially, competition [in the device and application markets] transfers existing profits from firms to consumers, and yields overall efficiency gains from expanded demand due to lower prices.").

104. *See, e.g.*, Matthew Adler & Eric Posner, *Rethinking Cost-Benefit Analysis*, 109 YALE L.J. 165, 245-46 (1999) (cost-benefit analysis in rulemakings usefully incorporates a wide range of values into administrative decisions, although it cannot capture social welfare effects of regulation); Robert Frank & Cass Sunstein, *Cost-Benefit Analysis and Relative Position*, 68 U. CHI. L. REV. 323, 374 (2001) (endorsing cost-benefit analysis, but criticizing techniques that underestimate the benefits from regulation).

105. *See, e.g.*, Nou, *supra* note 73, at 604 n.13 (quoting THE COST-BENEFIT STATE,

evaluating contested regulatory options along one dimension.

2. Measuring Costs

In theory, an auction could reveal the costs of open platform conditions to the prospective licensees. In the 700 MHz proceeding, the FCC plausibly defined the costs as the "potential negative effects [of the conditions] on network operations. . . ."[106] Assuming rational bidding, these costs will equal the premium a network operator would pay for unconditioned spectrum over what it would pay for the conditioned spectrum. If Verizon Wireless, for example, would pay $4.6 billion for the conditioned spectrum, but $5.6 billion for the unconditioned spectrum, then the cost of the conditions to the network operator is $1 billion. What needs to be measured then, and what the FCC said it was measuring, is "the magnitude of the devalued bids."[107]

The FCC asserted that it would know this number was too great to support its presumption that the open platform conditions produce net public interest benefits if bidders failed to meet the reserve price it set for the C Block licenses in the initial auction. [108] The problem with this logic is that the reserve price mechanism says nothing about the magnitude of auction revenue under different regulatory conditions. The reserve price mechanism is Boolean. If the reserve price is not met, at which point the FCC automatically drops the open platform conditions and re-auctions the spectrum, the agency actually knows nothing about the difference in value between the conditioned and unconditioned licenses. It is only after the conditions have been removed and the spectrum re-auctioned that there is any useful information on the magnitude of the devalued bids. By this time, however, the information is irrelevant to the policy choice about whether or not to impose the conditions. In other words, this information cannot be brought to bear on the fundamental question of whether the open platform conditions are worth their costs.

Let us suppose, for example, that the high bid in the initial auction is $4.5 billion, just $100 million shy of the $4.6 billion reserve price. When re-auctioned, the licenses fetch $4.7 billion. The magnitude of the devalued bids is thus rather small – only $200 million. Was it worth dropping the open platform conditions for a mere $200 million? This is a discussion that *should* be conducted before the conditions are lifted, of course, but can only be conducted as a hypothetical after the decision has

supra note 73) ("[Cost benefit analysis] can be seen . . . not as an endorsement of the economic approach to valuation, but as a real-world instrument, designed to ensure that the consequences of regulation are placed before relevant officials and the public as a whole.").

106. *700 MHz Order*, *supra* note 3, at 15,403.

107. *Id*.

108. See previous discussion of Measuring Benefits *supra*.

been made. Now, suppose that the high bid in the initial auction is $4.6 billion. Because the reserve price *has* been met, there will be no re-auction. If there had been a re-auction, suppose that the high bid would have been $5.6 billion. In this example, the magnitude of the devalued bid is $1 billion, but because the reserve price was met, we will never know that the open platform conditions (which we were willing to give up for a mere $200 million if the reserve price was not met) actually cost $1 billion.

Another problem with the FCC's re-auction procedure is that it assumed, but did not implement, a controlled experiment. If the only thing that changes between the auction with conditions and the auction without conditions is the presence of conditions, then we can learn what the conditions cost. This was not the case in the 700 MHz proceeding. The FCC announced that in addition to dropping the conditions from the licenses in a re-auction, it would also disaggregate the licenses into smaller blocks of spectrum covering smaller geographic areas.[109] Given that the re-auction would involve an entirely different package of assets, it cannot be said that the difference in price says anything in particular about the open platform conditions.

B. Towards a Valid Auction Heuristic

The only way to accurately price a particular public interest obligation, like the open platform conditions, is to hold simultaneous auctions of conditioned and unconditioned licenses. A simultaneous auction of two assets, different in only one respect, should tell us the magnitude of the devalued bids for conditioned spectrum. There is much that such an auction would not tell us. It could not answer the normative policy decision about whether the costs of the public interest requirement are worth bearing. It could not determine whether a proposed public interest requirement produced net benefits, at least not where there are positive externalities.

The question an auction heuristic could help answer is whether the costs of a public interest requirement, as measured by foregone auction revenues, is acceptable given all of the competing considerations. Where the public interest obligation involves public health and safety, such as an obligation to provide E911 services, the answer might well be that no price is too high and there is nothing that auction results can teach us. There will undoubtedly be greater ambivalence about other public interest obligations, particularly those that seek to structure economic markets.

To make the auction heuristic useful, the rulemaking that precedes

109. *700 MHz Order*, *supra* note 3, at 15,402-03.

the auction would have to produce two prices. The first would be an ordinary reserve price. The second would be the maximum price the public is willing to pay for a particular policy goal. If it is $1 billion for open platform conditions, then the winner of the auction for the conditioned licenses would win the licenses so long as its bid was no more than $1 billion less than the highest bid for the unconditioned licenses in a simultaneous auction.

There are undoubtedly practical challenges in the construction of such simultaneous auctions. All auction design is complex and susceptible to gaming by bidders. Whether or not game theorists could surmount these auction design challenges, I cannot say. What does seem clear is that a sequential auction of the kind envisioned in the 700 MHz proceeding would have told us very little about the actual costs of the open platform conditions. A comparison between the C Block license prices and those of unencumbered licenses in other 700 MHz bands is similarly uninformative because of the variables of frequency, license size, and other license conditions.

CONCLUSION

Thoughtful critics of the FCC are calling for a new approach to telecommunications policymaking that is more transparent and fact-based.[110]

Fact-based decisionmaking can be difficult when parties to spectrum proceedings throw around conflicting, often detailed, claims that any given policy choice will confer dramatic benefits or equally dramatic losses on the public, and the FCC lacks the means to independently evaluate the claims. In some cases, the claimed benefits and losses will be captured by private assessments of the value of spectrum licenses that have been encumbered by public interest conditions or otherwise crafted to advance specific public policies. In these cases, fact-based decisionmaking would be aided by using the auction process to flush out the parties' actual (rather than claimed) assessments of the benefits and burdens of the FCC's tentative policy choices.

Ultimately, the choice to forgo auction revenue to achieve specific public interests must be a normative one based on telecommunications policy objectives. For sound reasons, the FCC is not permitted under law to maximize auction revenue at all costs, lest fiscal policy usurp telecommunications policy responsibilities. And yet, particularly in times

110. *See, e.g.*, Philip J. Weiser, *FCC Reform and the Future of Telecommunications Policy* (forthcoming 2009) (manuscript at 13, http://fcc-reform.org/paper/fcc-reform-and-future-telecommunications-policy).

of federal budget deficits, there is pressure on the FCC, codified in law, to obtain fair value for wireless licensee use of the spectrum. What fair value is and what kinds of benefits the public is receiving for spectrum use are questions that are, to some extent, empirical inquiries. They can be advanced by well-designed auction procedures. Use of an auction heuristic along the lines that the FCC developed in the 700 MHz proceeding, but corrected to function properly, would provide important feedback for future spectrum policy decisions. It would also allow the FCC to defend policy choices that reduce auction revenue or, indeed, to abandon them.

THE POISON FRUIT: HAS APPLE FINALLY SEWN THE SEED OF ITS OWN DESTRUCTION?

DANA P. JOZEFCZYK*

* Dana Jozefczyk is a student at University of Colorado Law School. Thanks to my Article Editors, Daniel Estes and Gil Selinger, for taking the time to work with me on this article. Also, thanks to Hiwot Molla, John Bergmayer, Kyaw Tin, and the other editors on JTHTL for their time and efforts in making this article publishable. Finally, thanks to Bryan Shannon for his input and insight.

INTRODUCTION

On June 3, 2002, BBC News announced that "Napster, the California company that pioneered the mass-market swapping of music online, has filed for bankruptcy protection from its creditors."[1] Napster was a well known Peer-to-Peer (P2P) network that allowed users to upload copyrighted music files without charge. Napster's legal troubles began in 1999 when the Recording Industry Association of America (RIAA) filed suit against Napster for copyright infringement.[2] Napster's troubles continued when Metallica discovered the unauthorized circulation of its demo "I Disappear" on the Napster network.[3] Unlike other artists in similar situations at that time, Metallica did not passively allow the illegal file sharing to continue and filed suit for copyright infringement in violation of the Digital Media Copyright Act (DMCA, or "the Act").[4] At the time these suits were filed, the Act was relatively new and the legal community perceived the suit against Napster as the Act's first real test.[5] The Act prevailed. On July 26, 2000, the District Court of Northern California issued an injunction against Napster, which was later upheld by the Ninth Circuit Court of Appeals.[6] The

1. *Napster Files for Bankruptcy*, BBCNEWS, June 3, 2002, http://news.bbc.co.uk/1/hi/business/2023201.stm.

2. *See* Matt Richtel, *Napster Charts a New Course After Ruling*, N.Y. TIMES, Feb. 14, 2001, at C5.

3. GABRIEL ALATORRE, ET AL., MASS. INST. OF TECH., COPYRIGHT INFRINGEMENT DUE TO ONLINE FILE SHARING (2005), http://ocw.mit.edu/NR/rdonlyres/Electrical-Engineering-and-Computer-Science/6-901Fall-2005/B28F8F46-AE8B-4323-ACB0-D99937779637/0/online_fileshrng.pdf.

4. Christopher Jones, *Metallica Rips Napster*, WIRED, Apr. 13, 2000, http://www.wired.com/politics/law/news/2000/04/35670.

5. Patricia Jacobus, *Napster Suit Tests New Copyright Law*, CNET NEWS, Apr. 11, 2000, http://news.cnet.com/2100-1023-239092.html.

6. A&M Records, Inc. v. Napster, Inc., 114 F. Supp. 2d 896 (N.D. Cal. 2000), *aff'd in part by* A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004 (9th Cir. 2001). For an overview of the Napster case, see *Appeals Court Upholds Napster Injunction*, TIDBITS, Feb. 15, 2001,

injunction barred Napster from "causing, assisting, facilitating, copying, or otherwise distributing all copyrighted songs or musical compositions."[7] The decision was met with concern and criticism from major technology and Internet companies. These companies filed lengthy legal briefs, stating that the decision "could threaten the future of much of the technology industry."[8] Industry-wide developments in the wake of the Napster decision have confirmed these companies' greatest concerns.

The Napster decision has come to serve as a primer in a chain of events leading up to the current antitrust suits filed against Apple Computer.[9] This article will investigate this causal chain and its effect on the current lawsuits pending against Apple. Part I of the article will provide an introduction to the basics of copyright law. Part II will explore the implementation of the DMCA and what it forbids. Part III will examine Apple's digital rights management (DRM) encryption schematic, FairPlay, in detail and explain what makes FairPlay different from other DRM encryption systems. Part IV will parse out different perspectives on inoperability and how they apply to Apple. Part V will scrutinize the facts and legal arguments of the current U.S. cases, *Tucker v. Apple Computer Inc.* and *Somers v. Apple Computer, Inc*. Part VI of the article will examine how Apple's release of DRM-free music will affect the *Tucker* and *Somers* plaintiffs' anti-tying claims. A thorough investigation of the history and legal claims surrounding DRM-protected music yields the conclusion that Apple, by releasing DRM-free music, could have implemented less restrictive measures of music protection than FairPlay to protect its property rights. This conclusion reinforces Tucker's claim that Apple engaged in anticompetitive behavior in violation of antitrust laws.

I. COPYRIGHT LAW

When a musician creates or produces a song, that work of art is afforded the protections inherent under copyright law. However, recent digital technology advances have made enforcing these protections much more difficult. In particular, the advent of P2P networks in the early 1990s significantly undermined the rights guaranteed under copyright

http://db.tidbits.com/article/6295.

7. *Appeals Court Upholds Napster Injunction*, *supra* note 6.

8. John Borland, *Tech Giants Slam Napster Injunction*, CNET NEWS, Aug. 25, 2000, http://news.cnet.com/2100-1023-244976.html.

9. Tucker v. Apple Computer, Inc., 493 F. Supp. 2d 1090 (N.D. Cal. 2006) (order denying defendant's motion to dismiss); Complaint, Somers v. Apple Computer, Inc., No. 5:2007CV06507 (N.D. Cal. Dec. 31, 2007), *available at* http://docs.justia.com/cases/federal/district-courts/california/candce/5:2007cv06507/198939/1 [hereinafter Somers' Complaint}.

law by facilitating the online transfer of free bootlegged music. The proliferation of P2P network users resulted in an enormous number of file downloads in violation of copyright law. In order to understand the violations that occurred, it is first necessary to understand the basics and boundaries of copyright law.

A copyright is a form of legal protection provided to authors of "original works of authorship" by the Copyright Act, found in Title 17 of the U.S. Code.[10] An "original work of authorship" can include literary, dramatic, musical, artistic and intellectual works.[11] The 1976 Copyright Act gives the owner of a copyright the rights to reproduce the work in copies or phonorecords, prepare derivative works, distribute copies of the work, perform the work publicly, display the work publicly, and in the case of sound recordings, to perform the work publicly by means of audio transmission.[12] It is illegal for anyone to violate these rights but exceptions do exist such as "fair use" and "compulsory license" under which certain uses of copyrighted works are permitted.[13]

Copyrights are secured when a work is created, and creation occurs when the work is fixed in a copy or phonorecord for the first time.[14] Once a copyright is created, it generally lasts for the lifetime of the author plus seventy years after the author's death, although exceptions do exist.[15] Notice of copyright to others is not required, but it is often beneficial to the copyright owner because effective notice informs the public that the work is protected by copyright.[16] Since notice is not required under the Copyright Act, it can be difficult for potential infringers to know whether a work is copyrighted or not.[17] Even an individual who does not profit from infringing a copyright can be held liable for violation of the Copyright Act if that individual simply distributes a copyrighted work.[18] This is not to say that all copyright infringers are held criminally responsible. Copyright is mostly civil law,[19] but the Copyright Act does have a provision for criminal offenses.[20] The Copyright Act states that a person can be held liable for criminal infringement of a copyright when that individual willfully infringes for

10. *See* 17 U.S.C. § 102(a) (2006); United States Copyright Office, Copyright Office Basics, http://www.copyright.gov/circs/circ1.html#wci.

11. 17 U.S.C. § 102(a).

12. *Id*.

13. *Id*. §§ 107, 115.

14. *Id*. § 101.

15. *See id*. § 302(c) (listing exceptions for anonymous and pseudonymous works).

16. *Id*. § 401(a).

17. *See* Brad Templeton, Founder of ClariNet Communication Corp., Ten Big Myths About Copyright Explained (Oct. 2008), http://www.templetons.com/brad/copymyths.html.

18. 17 U.S.C. § 501 (2006).

19. *Id*. §§ 502-05.

20. *Id*. § 506.

either 1) purposes of commercial advantage or personal financial gain, or 2) distributes or copies one or more copies or phonographs or their equivalent, if such copies are worth over $1,000.[21]

With the possibility of a civil fine or even a felony charge lurking overhead, one might wonder why a music lover would choose to administer or even download from a P2P network. The answer is that the Copyright Act, as it existed in 1998, did not actively fight music piracy or enforce punishments, thus allowing P2P users to enjoy unfettered use of copyrighted works.[22] The record and movie companies, displeased with this trend, pushed for reform and convinced Congress to pass the Digital Millennium Copyright Act of 1998.[23]

II. THE DIGITAL MILLENNIUM COPYRIGHT ACT

The DMCA, commonly known for its crippling effect on free P2P networks such as Napster, was Congress' solution to online music piracy. The Act was passed by Congress on October 12, 1998[24] and was signed into law by President Clinton sixteen days later.[25] The purpose of the DMCA is to update U.S. copyright laws for the digital age and impose criminal sanctions on those who infringe copyrights.[26] One aspect of copyright law regulated by the DMCA is the circumvention of technological measures used to protect copyrighted works. In particular, Section 1201 of the DMCA establishes that contracting parties have a responsibility to provide effective protection against circumvention of technological measures used to protect copyrighted works.[27] The Act defines a technological measure as one that "effectively controls access to a work" and "requires the application of information, or a process or a treatment, with the authority of the copyright owner, to gain access to the work."[28]

In order to comply with Section 1201 of the DMCA, companies selling licensed music over the Internet were compelled to develop adequate "technological measures" to protect copyrighted works.[29] The common response was the development of DRM. DRM is a comprehensive term for access control technology that limits the use of

21. *Id.*
22. Fred von Lohmann & Wendy Seltzer, *Death by DMCA*, IEEE SPECTRUM ONLINE, June 2006, http://www.spectrum.ieee.org/jun06/3673.
23. *Id.*
24. 144 CONG. REC. S12375 (daily ed. Oct. 12, 1998) (statement of Sen. Hatch).
25. Statement on Signing the Digital Millennium Copyright Act, 2 PUB. PAPERS 1902-03 (Oct. 29, 1998).
26. S. REP. NO. 105-190, at 2 (1998).
27. 17 U.S.C. § 1201 (2006).
28. *Id.* § 1201(a)(E)(3)(B).
29. *Id.*

copyrighted digital media. On the most basic level, DRM allows a user to download music legally while at the same time restricts the user's ability to share the legally-downloaded music over P2P networks by encrypting the media file.[30] In response to the DMCA provision, Microsoft developed Windows Media DRM in 1999.[31] Microsoft licensed its Windows Media DRM to online music stores such as SpiralFrog, Directsong, Napster To Go, PassAlong, and Rhapsody to Go.[32] Although Windows Media DRM was available for licensure, Apple decided to develop its own DRM called FairPlay to protect music sold through its own online store.

III. FAIRPLAY

Like Windows Media DRM, FairPlay encrypts legally purchased music files from the iTunes Music Store to prevent infringement and protect the copyrighted works. Apple first introduced the iTunes Music Store on April 29, 2003.[33] Music sold through iTunes Music Store differs from music sold by other online media stores in two ways: music files sold through iTunes are 1) AAC files and 2) encrypted with FairPlay DRM.

First, instead of using Windows Media Audio (WMA) or MP3 formats, Apple chose to employ Advanced Audio Coding, or AAC, to encode its iTunes digital audio collection. All three technologies, WMA, MP3, and AAC, are compressed audio files designed to greatly reduce the amount of data necessary to reproduce high-quality versions of the original uncompressed recording.[34] The reduced size of compressed audio files allows users to easily download and play the files on their portable music players.[35]

Audio experts generally agree that WMA and AAC deliver higher quality sound than MP3. WMA was developed by Microsoft to compete with the original MP3[36] while AAC was developed by the same audio

30. MICHAEL A. EINHORN & BILL ROSENBLATT, PEER-TO-PEER NETWORKING AND DIGITAL RIGHTS MANAGEMENT: HOW MARKET TOOLS CAN SOLVE COPYRIGHT PROBLEMS 1 (Cato Institute 2005), http://www.cato.org/pubs/pas/pa534.pdf.

31. Microsoft, Windows Media DRM FAQ, http://www.microsoft.com/windows/windowsmedia/forpros/drm/faq.aspx.

32. For a comprehensive list of online music stores offering files protected by Windows Media DRM, see Microsoft, Online Stores in Windows Media Player, http://www.microsoft.com/windows/windowsmedia/player/stores.aspx [hereinafter Microsoft Online Store].

33. PQDVD.com, iPod History and Design, http://www.pqdvd.com/ipod-software-detail.html.

34. Marc Saltzman, *Acronym Soup: A Quick Guide to MP3, WMA, and AAC*, SYNC, June 29, 2007, http://www.sync-blog.com/sync/2007/06/acronym-soup-a-.html.

35. *Id*.

36. *See* Real.com, Analysis of the Microsoft Audio Codec,

experts who created the original MP3.[37] Apple's iPod can only play files that are in MP3 or AAC format. However, unprotected (DRM-free) WMA files are automatically converted to MP3s when imported into iTunes and can, therefore, be played on the iPod.

The second and more controversial way music sold through iTunes differs from music sold through other online media stores is its employment of FairPlay DRM encryption. While iTunes files can play on five computers and an unlimited number of iPods, FairPlay DRM renders music purchased from iTunes inoperable with digital music players that are not iPods.[38] In order to understand why use of FairPlay DRM promotes inoperability two questions must be addressed. First, how does FairPlay work? And second, precisely how are iTunes files inoperable with non-iPod digital music players?

A. How does FairPlay work?[39]

FairPlay DRM encrypts every song purchased from the iTunes store through a series of complex coding processes. Before purchasing media from iTunes, a user has to create an account with Apple's servers. Creation of an account not only registers the user with iTunes but also authorizes the user's PC or Mac with the iTunes Store by creating a user ID for that computer. When a user purchases a song from the iTunes Store using any of five authorized computers, iTunes creates a user key for the purchased song. The purchased track is then scrambled by a master key and the master key is encrypted with the user key. The user key is held by the purchaser's iTunes account and is also sent to the iTunes store. This process results in a system that does not require iTunes to interact with Apple every time it plays a song—the necessary information to play a track is stored in that song's user key within iTunes.

The FairPlay encryption system is capable of authorizing up to five different computers using a single iTunes account. When a new computer is authorized, the Apple server sends the machine all of the user keys for the songs purchased by the user's account. This enables the

---

http://www.real.com/msaudio/.

37. Daniel Eran Dilger, *How FairPlay Works: Apple's iTunes DRM Dilemma*, ROUGHLY DRAFTED, Feb. 26, 2007, http://www.roughlydrafted.com/RD/RDM.Tech.Q1.07/2A351C60-A4E5-4764-A083-FF8610E66A46.html; *see also* iTuneSupport.com, How does FairPlay work?, http://www.itunesupport.com/node/177.

38. Dilger, *supra* note 37.

39. This section is based primarily on research conducted by Daniel Eran Dilger and summarized in his article. *Id.* It is modified here with permission from the author. For additional information, see Howard Wen, *JHymn Goes Behind Atoms and Apple to Bring DRM-Free Music*, OSDIR, Jan. 27, 2005, http://osdir.com/Article3823.phtml.

user to play previously purchased music on a newly authorized machine.

Once authorization is complete, a user can connect his iPod to any of the five authorized machines and transfer music. However, the complex process of encryption works to protect the purchased tracks from being played by any portable music player that does not have access to the information stored in the user key. By connecting an iPod to an authorized computer, the user is not only transferring music but is also giving the iPod access to the user keys stored on that computer.

An iPod can only play music purchased from the iTunes store once it has decrypted it and can only decrypt the music by accessing the necessary user keys. This arrangement places several limitations on a user's ability to transfer music to an iPod from a computer that is not authorized on the users' iTunes account. First, if a user has music in his iTunes library that lacks a user key—either from another computer or from a computer that has been deauthorized—that music will not be copied to the iPod because the iPod lacks the user keys to decrypt it.

Second, a user cannot dock his iPod to an unauthorized computer and transfer music while at the same time retaining the music already on the iPod. If a user transfers music from an unauthorized computer onto his iPod, the new music will replace his existing music as the new user keys will replace existing user keys and the old music will no longer be playable.

Third, the iPod will not play music protected with DRM other than FairPlay (i.e., Windows Media DRM) because the music lacks the necessary user key. These three aspects of the user key system exemplify how FairPlay's design results in inoperability between the iPod and improperly encoded music. The inoperability between FairPlay encrypted music and non-iPod portable music players, however, is accomplished through a process by which FairPlay DRM actually alters the underlying music file, and renders the altered file unplayable on non-iPod portable music players.

B. *Why iTunes music files are inoperable with non-iPod digital music players*

Music files purchased from the iTunes store are inoperable with digital music players other than the iPod. Inoperability occurs because FairPlay encryption alters the structure of the typical AAC file so that it can no longer be converted to MP3 format and is not compatible with other players. FairPlay alters the structure of the typical AAC file by replacing a standard element of the file called the "mp4a atom" with a non-standard proprietary "drms atom."[40] Every four letter sequence at the

40. Wen, *supra* note 39.

beginning of a line of code is an atom, and the replacement drms atom contains the same basic song information as the mp4a atom plus information about the identity of the song purchaser as well as some cryptographic information necessary in ultimately decrypting the music.[41] Because the atom is altered and encrypted, a digital music player that does not have the decryption codes (i.e., one that is not an iPod) cannot unlock the file.[42] Thus, by altering the structure of an AAC music file, the FairPlay DRM causes that file to be inoperable with other digital music players.

IV. INOPERABILITY

As stated above, Apple's FairPlay DRM creates an overall system of inoperability between iTunes and other portable digital music players as well as protected music purchased from other online digital music stores and the iPod. While critics of Apple have faulted the company for this arrangement, proponents of Apple argue that inoperability is the only way to ensure that music is protected from piracy.[43] Counter to this argument, there exist at least three other measures Apple could have taken to protect music from piracy other than creating a system of inoperability: 1) Apple could have licensed WMA from Microsoft; 2) Apple could have licensed FairPlay to other online music stores; or 3) Apple could have allowed other companies to develop DRM encryption that would be playable on an iPod. These alternatives raise an important question—is operating an inoperable DRM system the least restrictive way by which Apple can protect its music?

A. Inoperability Alternative Number 1: Apple could have licensed WMA from Microsoft

Instead of creating FairPlay DRM when opening its iTunes store, Apple could have licensed Windows Media DRM from Microsoft. The first version of Windows Media DRM was released in April 1999, almost two years before Apple announced the release of iTunes and FairPlay DRM.[44] This early version of Windows Media DRM was compatible with both early versions of Windows (Windows 95 and later) and with Mac O.S. 8.1.[45] Numerous online music stores chose to license

41. *Id*.
42. *Id*.
43. Dilger, *supra* note 37.
44. Microsoft, Microsoft Windows Media – SDKs and Versions of Windows Media DRM, http://www.microsoft.com/windows/windowsmedia/forpros/drm/sdksandversions.aspx#version.
45. *Id*.

Windows Media DRM, including AOL MusicNow, FYE, Musicmatch Jukebox, Napster, Yahoo Music, PassAlong, CinemaNow, and Wal-Mart music downloads.[46] Not only have online music stores chosen to license Windows Media DRM, but many companies that manufacture portable digital music players have ensured that their products will be compatible with files encrypted with Windows Media DRM. These portable digital music companies include Archos, Cingular, Cowon, Creative Labs, Denon, Digitrex, D-Link, Ericsson, Insignia, iRiver (a Samsung product), Motorola, Nokia, Palm, Pioneer, Phillips, Roku, RCA, Samsung, SanDisk, Sony, and Toshiba.[47] However, since Microsoft's DRM can only protect WMA files and not AAC files, by choosing to use the AAC encryption, Apple precluded the possibility of Windows Media DRM licensure and failed to join the army of Windows Media DRM compatible companies.

Even though Windows Media DRM is unable to protect AAC files, it remains possible for Apple to license Windows Media DRM if it desires. In order to protect its files with Windows Media DRM, Apple would first have to switch new files to a WMA format and convert old AAC files to WMA format before the files could be protected. Once the files were successfully converted to WMA format, Apple could then license Windows Media DRM to protect its music. Since Microsoft charges royalty fees of $0.20 per unit or an annual maximum fee of $800,000 dollars to license Windows Media DRM, using 2005 sales figures, Apple could license Windows Media DRM at a cost of about $0.03 per iPod sold.[48]

Despite the obvious upside of interoperable DRM, two arguments support Apple's decision not to license Windows Media DRM from Microsoft. The first is an argument against monopolies. Economists since Adam Smith have documented the issues surrounding un- or under-regulated monopolies and cautioned the public against them. If Apple licenses Windows Media DRM from Microsoft, it will contribute to a Microsoft monopoly over DRM protection. As a competitor, Apple has no incentive to pass business over to Microsoft because it has already created DRM protection on its own.

The second argument against licensure is rooted in the fact that Microsoft's DRM code is frequently cracked and Apple could not

46. Microsoft Online Store, *supra* note 32.

47. Microsoft, Top Portable Media Devices, http://www.microsoft.com/windows/windowsmedia/devices/topdevicepicks.mspx.

48. *See, e.g.*, *Tucker*, 493 F. Supp. 2d at 1094; Microsoft, Windows Media Licensing Royalties for Final Products, http://www.microsoft.com/windows/windowsmedia/licensing/final.aspx#WindowsMediaDRM10_Final.

rationally license a mediocre product.[49] In 2006, a program was released called FairUse4WM that removed Windows Media DRM encryption from WMA files.[50] Windows has since secured the breach, but to those who argue against licensure, programs such as FairUse4WM strengthen an anti-licensure argument. This is not to say that Apple's FairPlay code has been free from attack by hackers. Programs such as QTFairuse, PlayFair, and PyMusique are all hacker designed programs that have exposed Apple's DRM to security breaches in the past.[51] However, since Microsoft has not developed a truly superior DRM, Apple does not have an incentive to license it.

B. Inoperability Alternative Number Two: Apple could license FairPlay to other online music stores

With the seemingly common aim of interoperability, one would think that if Apple is unwilling to license Windows Media DRM from Microsoft, it would be willing to license FairPlay so that consumers could achieve interoperability between all portable digital music players and all music sold by online music stores. This thought pattern, however, does not reflect Apple's demonstrated business decisions. Indeed, Apple's unwillingness to license FairPlay has been a central issue in lawsuits and legislation arising out of the European Union.[52] The French legislature recently passed a bill regarding inoperable DRM that attempted to force Apple to license FairPlay.[53] Apple, nevertheless, was saved from the dictates of the French legislature when the French Constitutional Council held the bill unconstitutional in part.[54] France is not the only country interested in mandatory licensure of FairPlay. The Danish Minister of Cultures warned that forthcoming legislation regarding Apple's DRM schematic would be adopted in 2007.[55] Norway took its

---

49. Jeremy Reimer, *Windows Media DRM Cracked*, ARS TECHNICA, Aug. 28, 2006, http://arstechnica.com/news.ars/post/20060828-7607.html.

50. *Id*.

51. Nate Anderson, *Hacking Digital Rights Management*, ARS TECHNICA, July 18, 2006, http://arstechnica.com/articles/culture/drmhacks.ars.

52. Thomas Crampton, *Key Parts of 'iPod Law' Struck Down in France*, INT'L HERALD TRIBUNE, Sept. 12, 2006, (Finance) at 13 (citing the 1789 Declaration on Human Rights, the French Constitutional Council declared major portions of the 'iPod law' unconstitutional. In particular, the Council found that "companies could not be forced, without compensation, to make music sold online compatible with any music device.").

53. Charles Jade, *Parts of French "iPod Law" Ruled Unconstitutional*, ARS TECHNICA, July 29, 2006, http://arstechnica.com/news.ars/post/20060729-7380.html.

54. *Id*.

55. Ken Fisher, *Denmark Next in Line to Challenge Apple, DRM*, ARS TECHNICA, Mar. 26, 2006, http://arstechnica.com/news.ars/post/20060326-6463.html (expressing optimism that "DRM interoperability would be backed by the various record labels who are eager to see legal alternatives to piracy flourish online.").

qualms against Apple one step further. The Consumer Council in Norway lodged a complaint charging Apple with violation of Norwegian contract law. The complaint alleged that Apple breached a contract between itself and Norwegian consumers when it restricted consumer use of music by employing FairPlay DRM.[56] These international legal matters reflect the views of many U.S. consumers—that Apple should license FairPlay to make the digital music industry as interoperable as commercially feasible.

On its face, licensing FairPlay appears to be a win-win scenario. The consumer wins because licensing FairPlay will create an interoperable DRM, allowing the consumer to play music purchased from any licensing music store on an iPod. Apple would also seemingly benefit from licensing FairPlay by collecting royalties from use of its license. Although Apple concedes that licensure royalties would provide a slight benefit, it argues that the potential detriments of such a licensure would outweigh the benefits of royalty payments.[57]

These potential detriments were described by Apple President Steve Jobs in an open letter dated February 7, 2006.[58] In this letter, Jobs lists two primary reasons why Apple refuses to license FairPlay DRM. First, licensing a DRM entails disclosing some of the DRM's secrets to many people in many different companies, and with widespread exposure, these secrets may leak.[59] If these secrets leak, then the DRM could be disabled and legally purchased music could be used in an illegal fashion.[60] Second, fixing a leak would be more difficult if it had to be conducted by many licensees instead of just one company.[61] If leaks such as this were to occur, Apple would no longer be able to guarantee DRM protection, and the "Big Four" music companies—EMI, Sony BMG, Universal, and Warner—may not sell as much of their music catalog to Apple.[62] The potential of a leak, coupled with the difficulty of fixing such a leak, led Mr. Jobs to categorically reject any proposals to license FairPlay.[63]

---

56. *Apple DRM is Illegal in Norway, Says Ombudsman*, OUT-LAW, Jan. 24, 2007, http://www.out-law.com/page-7691 (quoting Ombudsman Torgeir Waterhouse that the Norwegian Marketing Control Act requires "balanced and fair rights to the consumer when they purchase music form [sic] iTunes Music Store and similar download services.").

57. Open Letter from Steve Jobs, CEO of Apple Inc., "Thoughts on Music" (Feb. 6, 2007), http://www.apple.com/hotnews/thoughtsonmusic/ [hereinafter Jobs' Letter].

58. *Id*.

59. *Id*.

60. *Id*.

61. *Id*.

62. *Id*.

63. Jobs' Letter, *supra* note 57; *see also* Louis Hau, *Apple To Big Music: Set It Free*, FORBES.COM, Feb. 6, 2007, http://www.forbes.com/2007/02/06/jobs-apple-drm-tech-media-cx_lh_0206apple.html (listing the "big four" music companies and giving EMI's response to Jobs' letter).

C. Inoperability Alternative Number Three: Apple could allow other companies to develop DRM encryption that would be playable on an iPod

Not only will Apple neither license Windows Media DRM nor offer FairPlay for license, it also has thus far declined to allow other companies to sell music protected by a DRM schematic other than FairPlay, which could still be played on an iPod. In 2004, RealNetworks attempted to sell its own DRM music protection called Harmony.[64] Unlike other DRM music at the time, Harmony-encrypted music could be played on the iPod.[65] Apple lashed back, stating "we are stunned RealNetworks has adopted the tactics and ethics of a hacker to break into the iPod, and we are investigating the implications of their actions under the DMCA and other laws."[66] Despite this strong language, RealNetworks did not back down. Rather, RealNetworks stated that it never broke the DMCA because the statute "explicitly allows the creation of interoperable software."[67] In response to RealNetwork's refusal to discontinue Harmony, Apple altered the iPod's design rendering Harmony-encrypted music unplayable on the iPod.[68] Although the matter was never litigated, the clash between Apple and RealNetworks demonstrates Apple's commitment to being the exclusive company able to manufacture DRM music playable on the iPod.

V. INOPERABILITY AND TYING

Although Apple successfully defended its inoperable FairPlay DRM from the attacks explored above, two recent pending U.S. cases, *Tucker v. Apple Computer, Inc.* and *Somers v. Apple Computer Inc.*, have paved another avenue upon which inoperability can be challenged: antitrust tying.[69]

A. Tucker and Somers Background

The allegations brought forth against Apple in *Tucker* and *Somers* are markedly similar. Therefore, to simplify discussion concerning the cases at hand, this article will only thoroughly discuss the more advanced

---

64. Ryan Naraine, *Apple: RealNetworks Hacked iPod*, INTERNETNEWS.COM, July 29, 2004, http://www.internetnews.com/bus-news/article.php/3387871.

65. *Id*.

66. *Id*.

67. *Id*.

68. John Borland, *Apple Fights RealNetwork's "Hacker Tactics"*, CNET NEWS, Dec. 14, 2004, http://news.cnet.com/Apple-fights-RealNetworks-hacker-tactics/2100-1027_3-5490604.html.

69. *See Tucker*, 493 F. Supp. 2d 1090; Somers' Complaint, *supra* note 9.

case, *Tucker,* as the facts and legal arguments in *Somers* practically mirror the *Tucker* case.

Plaintiff Melanie Tucker brought an antitrust class action suit against Apple Computer on seven counts, three counts alleged violation of the Sherman Act and the remaining four counts alleged violation of related state law.[70] Specifically, under the Sherman Act, the plaintiff alleged Apple engaged in "(1) unlawful tying or bundling of Online Video and FairPlay music files to the iPod; (2) unlawful acquisition or maintenance of monopoly power in the digital music player market; and (3) attempted monopolization of the online music and video markets."[71]

Tucker filed the suit after she purchased an iPod from Apple and, periodically thereafter, purchased music from iTunes Music Store and transferred her downloaded music files to her iPod.[72] In her complaint, Tucker identified three separate markets in the U.S. in which Apple conducts business.[73] The first of these markets is the "Online Music Market."[74] "The 'Online Music Market' is the market for digital music delivered to the consumer by way of Internet download."[75] At the time of the hearing, Apple's share of the Online Music Market was 83 percent.[76] The second market identified by the plaintiff is the "Online Video Market."[77] The Online Video Market is the market for downloading digital video files that can be played on a computer or a video enabled digital music player.[78] Apple's share of the Online Video Market was found to be "at least 75 percent."[79] The third market identified by plaintiff is the "Digital Music Player Market."[80] The Digital Music Player Market is the market for portable battery powered devises that can store and play a large number of music files.[81] The Digital Music Player Market is comprised of two different types of portable music players: hard drive based and flash drive based.[82] Apple's share in these markets is about 90% and 70% respectively.[83]

Tucker alleged that Apple deliberately made music from the iTunes Music Store inoperable with other digital music players and that Apple

70. *Tucker*, 493 F. Supp. 2d at 1093.
71. *Id*. at 1095.
72. *Id*.
73. *Id.* at 1094.
74. *Id*.
75. *Id*.
76. *Tucker*, 493 F. Supp. 2d at 1094.
77. *Id*.
78. *Id*.
79. *Id*.
80. *Id*.
81. *Id*.
82. *Tucker*, 493 F. Supp. 2d at 1094.
83. *Id*.

deliberately manufactured the iPod so it is unable to play music purchased from competitors' online music stores.[84] According to the complaint, inoperability is achieved through both the construction of the iPod and the workings of FairPlay DRM.[85] Tucker examined each obstruction to inoperability in turn.

First, Tucker examined the structure of the iPod. Apple's core processor is the "Portal Player System-On-A-Chip."[86] This processor naturally supports WMA files, but Apple changed the configuration of the processor so that it only plays FairPlay protected AAC files.[87] Deliberately disabling a feature of a computer, as Apple has allegedly done, is known as "crippling" the product.[88] Software that has been altered in such a manner is known as "crippleware."[89] Since iPod is "crippled" from using any DRM other than FairPlay, iPod owners' only option for transferring music to an iPod is to do so through iTunes.[90]

Second, the plaintiff addressed Apple's FairPlay DRM, and argued that it renders music purchased on iTunes incapable of playback on any digital music player other than iPod and, thus, obstructs interoperability.[91]

Last, Tucker contended that these features of iTunes and iPod allowed Apple to charge a supracompetitive price.[92] Stemming from the above contentions, the plaintiff alleged three antitrust claims against Apple, the first of which, unlawful tying, is at the heart of the discussion below.[93]

B. *Unlawful Tying or Bundling of Online Video and FairPlay Music Files to the iPod*

Section 1 of the Sherman Act proscribes contracts, conspiracies, and combinations that restrain trade.[94] "Tying in violation of Section 1 can either be a per se violation or a violation of the rule of reason."[95] Tucker's complaint alleged that Apple participated in both per se tying and tying in violation of the rule of reason. A plaintiff seeking to establish a per se

84. *Id*.
85. *Id*.
86. *Id*.
87. *Id*.
88. *Tucker*, 493 F. Supp. 2d at 1094.
89. *Id*.
90. *Id*. at 1094-95.
91. *Id.* at 1094.
92. *Id*.
93. *Id*. at 1095.
94. 15 U.S.C. § 1 (2000).
95. *Tucker*, 493 F. Supp. 2d at 1096 (citing County of Tuolumne v. Sonora Cmty. Hosp., 236 F.3d 1148, 1157-58 (9th Cir. 2001)).

tying arrangement must establish that there exists "1) a tie between two separate products or services sold in separate markets; 2) sufficient economic power in the tying product market to affect the tied market; and 3) an effect on a substantial volume of commerce in the tied product market."[96] Implicit in these three elements is the requirement that the seller of the tying product "force[s] the buyer into the purchase of the tied product that the buyer did not want at all, or might have preferred to purchase elsewhere on different terms."[97] This final element, in effect, is an element of coercion and needs to be included in a plaintiff's complaint if the same is to survive a motion to dismiss.[98]

In its response to Tucker's claim, Apple contended that the plaintiff failed to state a claim upon which relief can be granted and moved to dismiss the tying claims pursuant to Federal Rule of Civil Procedure 12(b)(6). In accordance with this Rule, the court may dismiss a complaint for failure to state a claim based on either lack of cognizable legal theory or absence of sufficient facts to support such a legal theory.[99] When analyzing a motion to dismiss, "the court must presume that all factual allegations of the complaint are true and draw reasonable inferences from those factual allegations in favor of the non-moving party."[100] In the present case, Apple argued that the plaintiff failed to state a claim in three different legal theories of Tucker's complaint, namely: failure to allege an act of individual coercion; failure to allege a per se tying violation; and failure to allege an antitrust violation under a rule of reason analysis.[101] Each allegation is discussed below.

1. Individual Coercion

Apple first moved to dismiss the tying claim on the ground that the plaintiff failed to allege any "individual coercion."[102] To succeed in a tying claim, a plaintiff must allege some "modicum" of coercion.[103] In other words, a plaintiff cannot allege an antitrust tying violation if she acted out of free will and was not compelled to act by the defendant. The *Tucker* court acknowledged this necessity, but stated that the law did not require an allegation of coercion at the individual level and that an allegation at the market level was all that was needed for the action to

96. *Id*.
97. *Id*. (quoting Jefferson Parish Hosp. Dist. No. 2 v. Hyde, 466 U.S. 2, 12 (1984)).
98. *Id*. at 1097.
99. *Id*. at 1096.
100. *Id*.
101. *Tucker*, 493 F. Supp. 2d at 1096.
102. *Id.* at 1096-97.
103. *Id*. at 1097.

survive a motion to dismiss.[104] In analyzing Tucker's coercion assertion, the court looked to the language of the Ninth Circuit, quoting that "the essence of an antitrust tying violation is not the seller's unilateral refusal to deal with a buyer who refuses to buy the tied product, but the use by the seller of its 'leverage' to force a purchaser to do something he would not do in a competitive market."[105] While Tucker's complaint did not allege individual coercion, it did allege coercion in the general sense (i.e., that iPod owners are coerced into purchasing music from iTunes because Apple placed technological restraints on the iPod).[106] Applying the standards set forth by the Ninth Circuit, the court held that the plaintiff sufficiently alleged coercion and denied Apple's motion to dismiss this count.[107]

2. Per Se Tying

In her complaint, Tucker posited two theories of antitrust tying. First, Apple deliberately used technological restrictions to force purchasers of Apple's iPod (tying product) to purchase online music and online video only from iTunes Music Store (tied product).[108] And second, Apple deliberately used technological restrictions to force purchases of online music and online video from iTunes.[109] Through these two theories, the plaintiff alleged that Apple's conduct constituted a per se tying violation.[110] A tying claim consists of three elements "1) a tie between two separate products or services sold in separate markets; 2) sufficient economic power in the tying product market to affect the tied market;" and 3) as a result of these first two elements, there has been an effect on the substantial volume of commerce in the tied product.[111] Tucker's claim could only survive a motion to dismiss if it sufficiently alleged all three of these elements.

Apple contested the first element of the per se tying claim—namely by refuting the claim that iTunes and iPod are actually tied.[112] In defense of this argument, Apple stated that "[s]ome people buy iPods and never buy music from iTMS. That some people, like Tucker, choose to buy

---

104. *Id*.
105. *Id.* (quoting Murphy v. Bus. Cards Tomorrow, Inc., 854 F.2d 1202, 1204 (9th Cir. 1988), *partially overruled on other grounds,* Townsend v. Holman Consulting Corp., 929 F.2d 1358 (9th Cir. 1990)).
106. *Id*.
107. *Tucker*, 493 F. Supp. 2d at 1097.
108. *Id.*
109. *Id*.
110. *Id*.
111. *Id*. at 1096.
112. *Id*. at 1097.

both does not constitute unlawful tying."[113] Addressing Apple's contention, the court examined whether the complaint sufficiently alleged all three elements of the tying claim.[114] First, the court acknowledged that the plaintiff alleged a tie between separate products sold in two separate markets, which the complaint described as the Digital Music Player Market and the Online Video and Music Markets.[115] Next, Tucker alleged that each product, iPod and iTunes, is both a tied and tying product.[116] The plaintiff then alleged significant market power in each of the tying markets with 83% market share of the Online Music Market, 75% market share of the Online Video Market, and 90% market share in the Digital Music Player Market.[117] Last, the court found that the plaintiff alleged that Apple's conduct sufficiently affected the tied product markets. Since Tucker alleged all three of elements of a per se tying claim, the court found Apple's argument that some consumers do not buy products in both markets unavailing.[118] As a result, the court denied Apple's motion to dismiss this claim.[119]

3. Rule of Reason

The rule of reason requires a fact finder to weigh the anti-competitive and pro-competitive effects of defendant's business practices to determine whether the practice is unreasonable on balance.[120] To meet this burden, a plaintiff must show that "the activity is the type that restrains trade and the restraint is likely to be of significant magnitude."[121] Since the court found that the plaintiff properly alleged these elements in her per se tying claim, it did not find a reason to explore the rule of reason and denied Apple's motion to dismiss.[122]

C. Court's Holding

The court held that the plaintiff alleged illegal tying claims sufficient to withstand a motion to dismiss. *Tucker* is still awaiting trial, as is the factually similar *Somers*.

113. *Tucker*, 493 F. Supp. 2d at 1097.
114. *Id.* at 1097-98.
115. *Id*. at 1097.
116. *Id*.
117. *Id*. at 1098.
118. *Id*.
119. *Tucker*, 493 F. Supp. 2d at 1098.
120. *Id.*; *see also* Bhan v. NME Hosps., Inc., 929 F.2d 1404, 1413 (9th Cir. 1991).
121. *Bhan*, 929 F.2d at 1413.
122. *Tucker*, 493 F. Supp. 2d at 1098.

VI. DRM-FREE ITUNES AND ITS EFFECT ON *TUCKER* AND *SOMERS*

Since the court in Tucker denied Apple's motion to dismiss in December 2006, two major changes occurred in Apple's online music store. First, iTunes began offering some of its catalog as DRM-free music.[123] Beginning in May 2007, Apple and EMI (one of the four biggest record labels) teamed up and offered EMI's entire catalogue DRM-free through iTunes Music Store.[124] This new DRM-free service is called "iTunes Plus" and when first launched, sold DRM-free music on iTunes at $1.29 per song while the DRM music was still offered for sale at $0.99 per song.[125] Apple attributed the increased price of its iTunes Plus music not only to the DRM-free status of the music, but also to the enhanced sound quality.[126] Apple, however, is not the only online music retailer bitten by the DRM-free bug. In the fall of 2007, both Amazon and RealNetworks began offering DRM-free music but only charged between $0.89 and $0.99 per song, respectively.[127] Unable to justify the increased song price for DRM-free music when the same music was being offered for the market price of DRM music by competing online music retailers, Apple announced in October 2007 that it was ready to lower the price of its iTunes Plus music to $0.99 per song.[128] Even though Apple began offering EMI's entire catalogue as DRM-free music for the same price of music that has DRM, EMI music does not constitute the entirety of the iTunes library, and as a result, many music files are still only available with DRM encryption.

That was all about to change. In January 2009, Apple announced that it expected to offer its entire catalog DRM-free by the end of the first quarter of 2009.[129] These successive changes beg the question: Will recent developments in Apple's iTunes music store will affect Tucker's and Somers' tying claims?[130] First both plaintiffs need to argue illegal

---

123. Kristen Nicole, *iTunes DRM-Free Music Now Available*, MASHABLE, May 30, 2007, http://mashable.com/2007/05/30/itunes-upgrade/.

124. *Id*.

125. *Id*.

126. *Id*.

127. Press Release, RealNetworks, Rhapsody Teams with Universal Music Group for DRM-Free Music Test (Aug. 10, 2007), http://www.realnetworks.com/company/press/releases/2007/rhap_umg.html; Joshua Topolsky, *Amazon Launches DRM-Free "Amazon MP3" Music Downloads*, ENDGAGDET.COM, Sept. 25, 2007, http://www.engadget.com/2007/09/25/amazon-launches-drm-free-amazon-mp3-music-downloads/.

128. Posting of Scott McNulty to The Unofficial Apple Weblog, http://www.tuaw.com/2007/10/16/itunes-plus-price-drop-today-or-tomorrow/ (Oct. 16, 2007) [hereinafter McNulty Post].

129. Press Release, Apple, Changes Coming to the iTunes Store (Jan. 6, 2009), http://www.apple.com/pr/library/2009/01/06itunes.html.

130. Even though DRM-free music was released before Somers filed suit, her complaint

tying by demonstrating "a tie-in between two distinct products or services."[131] By offering music as DRM-free and thus eliminating the FairPlay tying mechanism, does Apple render these plaintiffs' arguments moot or strengthen their underlying allegations? This article addresses several potential answers below.

A. Apple's Potential Argument: Apple's actions are justified by the "business justification" defense, a defense that was only strengthened when Apple began offering DRM-free music in response to the record companies' licensure policies.

Under antitrust law, if the elements of an illegal tying claim are established and the products are deemed to be illegally tied, the defendant is liable for violation of the Sherman Antitrust Act. However, tying law does recognize certain per se tying defenses to an otherwise illegal tying practice.[132] One of these defenses is known as the "business justification" defense.[133] In cases involving the business justification defense, the court may find ample evidence of a tying violation, either per se or through a rule of reason analysis, but hold that the defendant is not guilty of an otherwise illegal tying arrangement because there are sound business interests that justify the conduct.[134] A common illustration used by courts when determining whether conduct is necessary under the business justification defense is the "one legged man" scenario.[135] As one court put it, it does not seem necessary that a seller sell only one shoe (out of a pair) to a one legged man when business interests would require that the shoes be sold as a pair.[136] As demonstrated by the one legged man illustration, the business justification must be compelling, and if a

---

reaches back to December of 2005 when Apple did not offer DRM-free music and thus Apple's release of DRM-free music will affect the *Somers* suit as well. *See* Somers' Complaint, *supra* note 9, at 3.

131. Mozart Co. v. Mercedes-Benz of N. Am., 833 F.2d 1342, 1345 (9th Cir. 1987) (quoting Robert's Waikiki U-Drive, Inc. v. Budget Rent-A-Car Sys., 732 F.2d 1403, 1407 (9th Cir. 1984)).

132. *See* Les Shockley Racing Inc. v. Nat'l Hot Rod Ass'n, 884 F.2d 504, 507 (9th Cir. 1989); Int'l Norcent Tech. v. Koninkligke Philips Elec. N.V., No. CV 01-00043 MMM (SSx), 2007 WL 4976364, at *6-7 (C.D. Cal. Oct. 29, 2007); *see also* Arik Johnson, *Tying Arrangements: Illegal Tying is One of the Most Common Antitrust Claims*, AURORA WDC, Oct. 30, 2007, http://www.aurorawdc.com/arj_cics_tying_arrangements.htm.

133. *See, e.g.,* Carpa, Inc. v. Ward Foods, Inc., 536 F.2d 39, 46 (5th Cir. 1976) ("The burden of supplying a business justification for what otherwise would be an illegal tie rests with the party asserting the defense.... Such a limited defense traditionally has been allowed in tie-in cases despite the *per se* characterization."); Ciminelli v. Cablevision, 583 F.Supp. 158, 162 (E.D.N.Y. 1984) ("[I]t is well settled that business justification may serve as a defense to a *per se* violation of the antitrust laws, as, for example, an illegal tying arrangement.").

134. *See* Johnson, *supra* note 132.

135. *Id*.

136. Dehydrating Process Co. v. A.O. Smith Corp., 292 F.2d 653, 655 (1st Cir. 1961).

defendant can establish such a necessity, then otherwise illegal tying actions will not be held as antitrust violations.

Apple has long emphasized that employing FairPlay DRM was not a choice, but a necessary measure mandated by the Big Four record companies to enforce compliance with the DMCA. According to Steve Jobs, "when Apple approached these companies to license their music to distribute legally over the Internet, they were extremely cautious and required Apple to protect their music from being illegally copied."[137] These four record companies control over 70% of the world's music, and FairPlay is Apple's response to demands these companies place on online music stores.[138] At the end of his letter, Jobs claimed that "if the Big Four Music companies would license Apple their music without the requirement that it be protected by DRM, we would switch to selling only DRM-free music in our iTunes Store."[139]

Shifting the blame for inoperability of FairPlay DRM from Apple to the record companies has been Apple's illegal tying counterargument since its DRM dilemmas began. Now that EMI is willing to license its catalogue to Apple as DRM-free music, Apple may argue that it is no longer necessary for them to employ FairPlay DRM to protect the music for EMI.[140] Furthermore, because Apple no longer uses FairPlay DRM on music from EMI's catalogue, Apple could argue that iPod and iTunes are no longer tied because DRM-free music sold can be played on every digital music player and DRM-free music sold from other online music stores can play on the iPod. By arguing that FairPlay DRM was a necessary measure for music licensure mandated by the record companies, Apple may be able to demonstrate a necessary business justification and escape liability for antitrust violation.

B. *Plaintiff's Potential Response: Apple's actions have demonstrated that alternatives exist to protect licensed music and, thus, DRM encryption was not a necessity and cannot support the "business justification" defense*

The business justification defense, while seemingly an ideal exit route for Apple, is not an easy hurdle to clear. In order to establish that the actions it took were justified, Apple must demonstrate that its actions were implemented for a legitimate purpose and no less restrictive alternative was available.[141] On first blush, it may appear that this was the

137. Jobs' Letter, *supra* note 57.
138. *Id.*
139. *Id.*
140. McNulty Post, *supra* note 128.
141. *See Mozart Co.*, 833 F.2d at 1349; Phonotele, Inc. v. AT&T, 664 F.2d 716, 738-39 (9th Cir. 1981).

case; Apple's only choice was to use DRM, and now that EMI is licensing its music DRM-free, Apple is free of that burden. This observation, however, is misconceived. Apple still needs to comply with the DMCA and the music sold on iTunes must still be protected against piracy. Instead of using DRM, Apple implemented a new form of protection: digital watermarking coupled with other forms of data storage.

1. Watermarking and iTunes

The phrase "digital watermarking" was first coined in 1992 by Andrew Tirkel and Charles Osborne.[142] Watermarking is the process of imbedding information into a digital file and is primarily used for copyright protections.[143] For example, a downloaded file may contain the downloader's personal information, including IP address, credit card number, and other private data. Apple does exactly this.

Apple is using watermarking to hide an iTunes music purchaser's personal information, such as full name and account email, in the purchased iTunes track.[144] This personal encryption is not unique to Apple's DRM-free music.[145] As discussed above, DRM music also has the user's personal information stored within it. But watermarking is not the only form of data storage used by Apple in its iTunes tracks. Music files purchased on "iTunes Plus" also contain large amounts of information in what seems to be a table format.[146] The information in these tables has not yet been decrypted, but through comparison to other AAC files, researchers have determined that the information stored on iTunes tracks is not only massive in size but detailed in nature.[147] Although Apple has not officially commented on the purpose of the controversial watermarking and information storage on its DRM-free files, it suggested that the obvious reason these measures were enacted was so Apple can track users who try to distribute DRM-free files over P2P networks and thus still protect the music licensed to them by

142. A.Z. Tirkel et al., *Electronic Water Mark*, *in* DIGITAL IMAGE COMPUTING: TECHNIQUES AND APPLICATIONS 666-73 (1993).

143. Digital Watermarking World, Digital Watermarking Frequently Asked Questions, http://www.watermarkingworld.org/faq.html#SECTION00022000000000000000 (Aug. 26, 2005).

144. Nick Farrell, *Apple's DRM-free Music has Poison Tip: Tells Everyone About You*, THE INQUIRER, May 31, 2007, http://www.theinquirer.net/en/inquirer/news/1036750/apples-drm-free-music-has-poison-tip.

145. *See* Tirkel, *supra* note 142.

146. Peter Eckersley, *Apple's DRM-Free AAC Files Contain More Than Just Names and E-Mail Addresses*, ELECTRONIC FRONTIER FOUNDATION, May 30, 2007, http://www.eff.org/deeplinks/archives/005282.php.

147. *Id.*

EMI.[148]

2. Will Plaintiff's tying claim survive Apple's release of DRM-free music?

Tucker's and Somers' complaints alleged that Apple intentionally used inoperable DRM; while Apple countered that it had no other choice.[149] Apple's defense can be opposed on two grounds: behavioral and, the more legally compelling, technological. As discussed above, Apple demonstrated through its past actions, including its disagreement with RealNetworks, that it would rather choose inoperability over interoperability.[150] While this refusal does not amount to legal evidence against Apple, it does speak to Apple's attitude regarding its tied product. The second, and more legally compelling, counter to Apple's defense is that Apple could have used watermarking technology to protect its music instead of DRM. Since watermarking technology existed prior to the release of iTunes and was actually built into iTunes files before they were offered as DRM-free, implementing a system of inoperable DRM was not the only way that Apple could protect its music. Apple, therefore, had access to a "less restrictive alternative," which it could have used instead of FairPlay DRM. Since such an alternative existed that did not inhibit interoperability, Tucker may be able to argue that Apple cannot prevail with the business justification defense. With no defense to justify its actions, Apple could be held liable for an illegal tying arrangement in violation of antitrust law and, thus, Tucker's and Somers' suits may well prevail.

CONCLUSION

The controversy concerning Apple's FairPlay DRM schematic has affected consumers across the globe. The exodus of DRM from some of the most popular online music stores saves consumers new to the online music community from having to suffer the ill effects of DRM encrypted files. Furthermore, those consumers who have purchased Apple's DRM music, which is now available in a DRM-free form, can update their files and rid themselves of the burdensome encryption. Unfortunately, fairly large populations of music purchasers remain who still experience the tying effects of Apple's products. These consumers remain tied because the DRM music that they purchased does not have a DRM-free update. They are left in a situation similar to those experienced by Tucker and

148. Scott Shuey, *Apple's Latest Trick to Enforce Digital Rights*, GULFNEWS.COM, Oct. 29, 2008, http://archive.gulfnews.com/articles/07/06/09/10131156.html.

149. *Tucker*, 493 F. Supp. 2d at 1096.

150. *Id.*

Somers in the pending suits against Apple. Fortunately for consumers in this situation, Apple might have sewn the seed of its own destruction by releasing iTunes Plus. Since Apple demonstrated that it could employ a less restrictive means to protect its music than FairPlay, it should not prevail on a business justification defense to an illegal tying claim. Assuming that these plaintiffs can establish the other elements of the per se claim, Apple will be held liable for its anticompetitive conduct. Consumers around the world can only hope that the monetary and punitive damages imposed on Apple after such a result would change Apple's business practices and perhaps then Apple would finally start playing fair.

DON'T SEND A LAWYER TO DO AN ENGINEER'S JOB: HOW ADVANCING TECHNOLOGY CHANGES THE SOFTWARE MONOCULTURE DEBATE

JOHN BERGMAYER*

INTRODUCTION AND SUMMARY

Modern computer and telecommunications technologies are particularly susceptible to network effects, where the value of a technology increases the more that people use it. Network effects combine with related phenomena, such as the drive toward technological standardization, to create markets that are often dominated by one

* John Bergmayer is a law student at the University of Colorado. He would like to thank Paul Ohm, Scott Challinor, and Shanelle Kindel for their feedback, and Susie Baker, for her patience.

technology. In personal computer software, and by analogy with the agricultural phenomenon, the dominance of Microsoft technology has been called a "software monoculture." In addition to its software being pervasive, it has been argued that Microsoft's engineering practices result in its software being overly complicated and insecure. Because of the widespread dependence on Microsoft's software, these insecurities are then argued to have widespread negative repercussions for the economy and national security. Various proposals, such as requirements that Microsoft share its technology, an expansion of tort liability principles, or merely isolating high-value computer systems from the Internet, have been advanced to deal with this problem.

This Note neither seeks to defend Microsoft's engineering practices, nor to argue that the dominance of its software is a good thing from an economic or engineering perspective. It only notes that the problem of software monoculture is largely a problem with technology, and that technological developments alone, without any legal or policy impetus, may be sufficient to deal with the problem. It also notes that the experience of a particular company following particular engineering principles at a particular time should not be extrapolated to general policy prescriptions. Because the evidence of the negative consequences of software monocultures is usually related to Microsoft products, the case against "monoculture" is really a case against one company.

The analogy to the dangers of excess homogeneity in biological systems is instructive when thinking about technology and software, and many of the same principles that explain the rise of an agricultural monoculture also explain the rise of a software monoculture. But measures that seek to improve diversity, while perhaps appropriate to agriculture, may not be applicable to the more malleable domain of computer software. Even if there are valid policy justifications for some intervention to increase technological diversity, countering the security effects of software monocultures is not among them. Legal reforms should be approached warily, because the risk of unintended consequences that could follow from an improperly calibrated liability regime is very great. Although the picture may have looked different only a few years ago, recent experience shows that such reforms are not justified as a means of counteracting negative security consequences of software monocultures.

I. BACKGROUND

A. Network Effects and Related Phenomena

Microsoft has undoubtedly attained a dominant, near-monopoly position in some software markets.[1] Unlike accounts that attribute Microsoft's success solely to its business and technological acumen, or that paint it as a company perpetually engaging in abusive, anti-competitive behavior, this Note will argue that its success is at least in part attributable to a number of economic and technological phenomena that have amplified and sustained its successes. In order to understand why a software monoculture might arise, it will be helpful to explain these phenomena and see how they apply to software monocultures and to Microsoft. To that end, this section will discuss network effects, indirect network effects such as the "applications barrier to entry," standardization, and path dependence. Often, the same phenomenon can be viewed as an example of more than one of these concepts. The purpose of this section is not to hold up aspects of Microsoft or of other software monocultures as exemplars of particular concepts. Rather, it is to help demonstrate that the rise of a software monoculture is at least partly the result of a complicated interplay of related economic, technical, and market forces, and not necessarily solely the result of improper behavior or an abuse of a dominant market position. Viewing the dominance of Microsoft in this light, it may be easier to accept that the similar forces can be relied on to counteract social costs that may have been caused by its success.

1. Network Effects

Markets for communications technology are particularly susceptible to network effects. Products that feature network effects become more valuable to a person the more that other people use them. While a hammer is just as valuable to a carpenter no matter how many other carpenters use the same kind of hammer, a fax machine is valuable only if there are other fax machines to send faxes to. The more people who have fax machines, the more valuable all fax machines become. Robert

1. Much has been written on whether Microsoft is a monopoly, and whether markets for software products can be considered natural monopolies. *See, e.g.*, COMPETITION, INNOVATION, AND THE MICROSOFT MONOPOLY: ANTITRUST IN THE DIGITAL MARKETPLACE (Jeffrey A. Eisenach & Thomas M. Lenard, eds., 1999). This discussion is largely irrelevant to this Note as whether or not a software company is a monopoly in the sense that it is able to take monopoly profits, it may still control a monoculture by having a large enough base of installed users. Additionally, monopolies may be distinguished from technology monocultures in that it is possible for one company to support multiple technologies, and for multiple companies to contribute to a software monoculture.

Metcalfe, co-inventor of Ethernet, stated the value of a telecommunications network is proportional to the square of the number of the users of the system.[2] While "Metcalfe's Law" has been criticized as inaccurate and overly simplistic,[3] it helps to keep in mind, as a rule of thumb, that communication networks obtain their value not just from the quality of their technology, but from the number of people who use them. Telephone networks offer a key historical example of network effects in a communication system. Although at first, there were several competing telephone networks that did not interconnect with one another, one network, the Bell System, soon drew enough users to make its offering significantly more valuable than that of its competitors.[4] Network effects are pervasive in newer electronic communications networks, as well. Even social networking sites such as MySpace benefit from network effects.[5]

Many computer technologies, notably operating systems,[6] are subject to network effects. The dominance of Microsoft Windows makes it vital that all PCs, even those running Linux and Macintoshes, be able to communicate with Windows PCs.[7] Therefore, even as Apple has

---

2. *See* Simeon Simeonov, *Metcalfe's Law: More Misunderstood Than Wrong?*, HIGH CONTRAST, July 26, 2006, http://simeons.wordpress.com/2006/07/26/metcalfes-law-more-misunderstood-than-wrong.

3. *See* ANDREW ODLYZKO & BENJAMIM TILLY, DIGITAL TECH. CTR., UNIV. OF MINN., A REFUTATION OF METCALFE'S LAW AND A BETTER ESTIMATE FOR THE VALUE OF NETWORKS AND NETWORK INTERCONNECTIONS (2005), http://www.dtc.umn.edu/~odlyzko/doc/metcalfe.pdf.

4. JONATHAN E. NUECHTERLEIN & PHILIP J. WEISER, DIGITAL CROSSROADS 5-6 (paperback ed. 2007).

5. Dion Hinchcliffe, *Web 2.0's Real Secret Sauce: Network Effects*, July 15, 2006, SOCIAL COMPUTING MAGAZINE, http://web2.socialcomputingmagazine.com/web_20s_real_secret_sauce_network_effects.htm.

6. Roman Beck has written on this topic, observing that

> Like other computing technologies, competing standards battle acrimoniously on the market to read a critical mass to take over the marker. Once established, a dominant standard becomes even stronger due to positive feedback effects while the "outgunned" standards lose even more market share. In extreme cases, a monopoly can be established, also known as the "winner takes it all" in increasing returns networks. Despite strong positive feedback effects accelerating the diffusion of dominating standards, stable equilibriums with several coexisting standards can also emerge. A prominent example of a stable oligopoly is the operating system software market for computers with Microsoft Windows as the dominant standard and Linux, as well as Mac OS for Apple Macintosh as sturdy clusters. Although Microsoft extended the positive feedback effects of its standards by adding complementary applications (e.g., by integrating Windows Internet Explorer), it was not able to displace its competitors completely. The former example indicates that standards on network effect markets can tend to lead to natural monopolies.

ROMAN BECK, THE NETWORK(ED) ECONOMY: THE NATURE, ADOPTION AND DIFFUSION OF COMMUNICATION STANDARDS 60 (2006).

7. Compatibility in computer systems is analogous to interconnection in

moved away from its proprietary AppleTalk networking technology in favor of the open standard TCP/IP,[8] it has continued to improve its support of SMB,[9] a technology used by Windows. The widespread support for SMB on non-Microsoft platforms is an example of SMB benefitting from a direct network effect.[10] Network effects also account for the rise of the Internet itself.[11] Although most software applications do not benefit from this kind of network effect,[12] applications that engage in any form of communication do. For instance, Microsoft Office benefits from the network effect of large numbers of computers being able to view and edit its documents.[13] In the 1980s, when word

---

communications systems. *See* BECK, *supra* note 6, at 55-56 (describing different forms of "compatibility" and noting that in some instances the dominant platform (Windows) is compatible with the smaller one (Macintosh)); KLAUS W. GREWLICH, GOVERNANCE IN "CYBERSPACE": ACCESS AND PUBLIC INTEREST IN GLOBAL COMMUNICATIONS 148 (1999) ("Without interconnection a small network would be severely disadvantaged relative to a large one").

8. Shelly Brisbin, *All Roads Lead to Rendezvous*, MACWORLD, July 2, 2003, http://www.macworld.com/article/26841/2003/07/allroadstorendevous.html (

> AppleTalk is a Mac-only technology in a cross-platform world. These days, most network hardware, PCs, and printers–as well as other devices don't support AppleTalk. They use TCP/IP, the language of the Internet. Universal TCP/IP support provides both seamless communication with the Internet and a single networking medium that all computer makers, software vendors, and users can agree on. As a result of this, Apple–while continuing to support AppleTalk in OS X–has started to focus on TCP/IP.

).

9. Daniel Drew Turner, *Apple Preps Early Release for Jaguar*, EWEEK, July 3, 2002, http://www.eweek.com/c/a/Past-News/Apple-Preps-Early-Release-for-Jaguar ("Jaguar will include new support for cross-platform standards such as . . . SMB . . . .").

10. *See generally* Chris Hertel, *Samba: An Introduction*, Nov. 27, 2001, http://us3.samba.org/samba/docs/SambaIntro.html.

11. B.G. KUTAIS, INTERNET POLICIES AND ISSUES 224-25 (2002). The Internet itself has also been described as a monoculture. *See* WILLIAM R. CHESWICK, STEVEN M. BELLOVIN & AVIEL D. RUBIN, FIREWALLS AND INTERNET SECURITY: REPELLING THE WILY HACKER 112 (2003).

12. For instance, a non-networked computer game does not derive its primary value from a large installed base. By contrast, game consoles themselves (as platforms) are subject to various kinds of network effects. *See* David S. Evans, *The Antitrust Economics of Multi-Sided Platform Markets*, 20 YALE J. ON REG. 325, 364 (2003).

13. STAN J. LIEBOWITZ & STEPHEN E. MARGOLIS, WINNERS, LOSERS, AND MICROSOFT: COMPETITION AND ANTITRUST IN HIGH TECHNOLOGY 181 (Independence Institute 2001); Knowledge@Wharton, *Rivals Set Their Sights on Microsoft Office: Can They Topple the Giant?*, http://knowledge.wharton.upenn.edu/article.cfm?articleid=1795 (quoting Kendall Whitehouse as saying

> [Y]ou bought Word because people send you Word files and you need to edit them… The one thing that's critical [in a competing product] is the ability to read and write those files. If you have a Mac [using iWork] that can read and write Word and PowerPoint files, then your ability to switch [away from Office] becomes a lot easier. The differentiator becomes user interface, speed and stability.

(bracketed text in the original.)) To be sure, Microsoft probably gained an initial edge due to its having a superior product. LIEBOWITZ & MARGOLIS, *supra*, at 180-200 (arguing that processors were used primarily to create printed documents, there was much more competition among different technologies than exists today. Today, it is just as important that people be able to email each other documents and be sure that they were able to be viewed.[14]

2. Indirect Network Effects

Although technologies such as Java that have sought to lessen the importance of desktop applications have not lived up to expectations,[15] the recent rise of web-based applications has been very rapid. For some users, web-based applications offer advantages over desktop software in categories such as email.[16] In the near term, however, desktop software is likely to remain important.[17] The overwhelmingly popular choice for desktop operating systems is Microsoft's Windows, and the majority of new applications are written for that platform.[18] This remains a significant advantage for the Windows platform, which continues to benefit from indirect network effects that reinforce its popularity.

Microsoft's dominance in word processors came about through the quality of their products, and containing charts showing the reduction in the number of major players in the word processing market). *But see* Ed Foster, *The Gripelog: How Did Word Perfect Go Wrong?*, Dec. 27, 2007, INFOWORLD.COM, http://weblog.infoworld.com/gripeline/archives/2007/12/how_did_wordper.html (quoting one reader as writing that "MS Office crushed its competition for one reason and one reason ONLY—undocumented application programming interfaces".).

14. LIEBOWITZ & MARGOLIS, *supra* note 13, at 181.

15. Andy Johnson-Laird, *Looking Forward, Legislating Backward?*, 4 J. SMALL & EMERGING BUS. L. 95, 115 (2000) (

> Created by Sun Microsystems . . . Java is an object-oriented programming language with goals that include the ability to write programs that will run on many different computers. This goal was dubbed "Write Once, Run Anywhere," and while being an admirable goal, it has not been attained yet--the epithet "Write Once, Debug Everywhere" is unfortunately more appropriate.

); RUBICON CONSULTING, GROWTH OF WEB APPLICATIONS IN THE US: RAPID ADOPTION, BUT ONLY WHEN THERE'S A REAL BENEFIT (2007), http://www.rubiconconsulting.com/downloads/whitepapers/Rubicon_-_Rising_adoption_of_web_applications.pdf (adoption of web applications is very rapid, although use varies among kind of application).

16. Brad Stone, *Firms Fret as Office E-Mail Jumps Security Walls*, N.Y. TIMES, Jan. 11, 2007, § A, at 1.

17. BOB BAXLEY, MAKING THE WEB WORK: DESIGNING EFFECTIVE WEB APPLICATIONS 28 (2002); Martin Lamonica, *Ray Ozzie's Quiet Revolution at Microsoft*, ZDNET, May 1, 2007, http://news.zdnet.com/2100-3513_22-6180539.html (Microsoft's Ray Ozzie on the continuing importance of the desktop in an era of web applications).

18. The total market share of operating systems for personal computers of various kinds of Windows has been estimated to be around 90%, with the Macintosh at around 7.3%, and Linux and "other" rounding things out. NEOWIN.NET, OPERATING SYSTEM MARKET SHARE (2007), http://staff.neowin.net/slimy/dec2007.pdf. On February 23, 2008, the website Versiontracker.com registered 52 updates its directory of Windows applications, and only 18 updates for its directory of Mac OS X applications. *See* Versiontracker, http://www.versiontracker.com (snapshot of site from Feb. 23, 2008 on file with author).

Indirect network effects are advantages popular platforms enjoy other than those directly related to interconnection. A computer platform becomes more valuable if many third-party applications are written for it, but developers will only create applications for platforms that have many users. A self-reinforcing cycle can develop as users gravitate towards platforms with the most applications, and application developers gravitate towards platforms with the most users. As this cycle makes it difficult for a new entrant to create a software platform, it has been called "the applications barrier to entry."[19] The past success of a software platform therefore contributes to its future success, in an example of "path dependence."[20]

Switching costs also contribute to the continued popularity of a platform.[21] Once a user has invested time and money in a particular platform, the costs of switching to a new platform may outweigh any gain to be had from adopting a new platform.[22]

A popular platform enjoys a few other advantages besides software availability and direct switching costs. For example, it is easier to obtain support and assistance for technologies that are widely used,[23] and new employees may need less training.[24] Companies that provide popular

---

19. Kenneth G. Elzinga, David S. Evans, & Albert L. Nichols, *U.S. v. Microsoft Corp.: Remedy or Malady?*, *in* MICROSOFT, ANTITRUST, AND THE NEW ECONOMY 154 (David S. Evans, ed., 2002). Note that the current prevalence of three video game consoles (the Wii, XBox 360, and Playstation 3), each of which has a software library incompatible with the others, demonstrates that the desire for software compatibility is not sufficient by itself enough to create a platform monopoly. *Cf.* BECK, *supra* note 6.

20. *See* PAUL J. EDWARDS ET. AL, UNDERSTANDING INFRASTRUCTURE: DYNAMICS, TENSIONS, AND DESIGN 17 ("Path dependence refers to the 'lock-in' effects of choices among competing technologies. It is possible, following widespread adoption, for inferior technologies to become so dominant that superior technologies cannot unseat them in the marketplace.").

21. *See generally* Michael L. Katz & Carl Shapiro, *Antitrust in Software Markets*, *in* COMPETITION, INNOVATION AND THE MICROSOFT MONOPOLY: ANTITRUST IN THE DIGITAL MARKETPLACE 29, 31-34 (Jeffrey A. Eisenach & Thomas M. Lenard eds., 1999).

22. Daryl Lim, *Copyright Under Siege: An Economic Analysis of the Essential Facilities Doctrine and the Compulsory Licensing of Copyrighted Works*, 17 ALB. L.J. SCI. & TECH. 481, 506 fig.2 (2007). It has recently been suggested that part of Apple's strategy with the iTunes App Store (which sells applications that can run on the iPhone and the iPod Touch) is to create switching costs for its users. Sean Devine, *Inconsequential Apps Used by Many People Increase Stickiness*, DEAL RANGE, Jan. 4, 2009, http://dealrange.typepad.com/deal_range/2009/01/inconsequential-apps-increase-stickiness-if-everyone-uses-them.html; Sean Devine, *The App Store: First Comes Power*, DEAL RANGE, Jan. 3, 2009, http://dealrange.typepad.com/deal_range/2009/01/the-apple-app-store-and-pricing-power.html.

23. For example, of the 48 businesses listed in the 2008 Yellow Pages for Boulder, Colorado, under "Computers–Service & Repair," only seven advertise expertise in Macintosh computers, and none in Linux. DEX: OFFICIAL DIRECTORY–BOULDER 226-31 (2007).

24. ICT Hub Knowledgebase, Software Standardization, http://www.icthubknowledgebase.org.uk/softwarestandardisation ("If you standardise [sic] on software which is widely used in the outside world, it will make sharing information easier and

technologies are unlikely to go out of business, leaving their customers "orphaned."[25] Furthermore, it is easier for IT purchasers to justify investments in widely-used platforms.[26]

3. Standardization

Communication technologies and computer software are also subject to pressures of standardization. Among other things, a "standard" is a technology or method that is selected because there is an advantage to picking just one.[27] There may not be any advantage at all to a standard, other than the fact that it *is* a standard. The metric system is a standard. So is the gauge of railroad tracks,[28] household electric voltage, the custom of driving on the right-hand side of the street in most countries, the use of certain formulations of gasoline, and AM radio. Standards may exist because of law, a dominant marketplace position, or habit. By virtue of its dominant marketplace penetration, Microsoft's software has become a de facto standard for home and business use. Standards have wide-ranging benefits. Railcars can move easily between different railroad tracks that share a standard gauge.[29] Different manufacturers create AK-47 rifles, and different manufacturers produce the 7.62 mm ammunition they use. Because people in a country all drive on the same side of the street, accidents are reduced.[30] There are human

---

reduce training needs of new staff. Many organisations [sic] have standardised [sic] on Microsoft Office for this reason.").

25. *See* About.com: Desktop Publishing, Where Are They Now? Finding Software Orphans, http://desktoppub.about.com/library/weekly/aa033199.htm.

26. *See* LOIS KELLY, BEYOND BUZZ: THE NEXT GENERATION OF WORD-OF-MOUTH MARKETING 115 (2007) ("The classic anecdote, 'You'll never get fired for buying IBM,' was based on anxieties. If I buy a little-known technology and it bombs, I'll be fired for it. If I hire IBM and the technology fails, IBM will be blamed, not me.")

27. *See generally* Yesha Y. Sivan, *Knowledge Age Standards: A Brief Introduction to Their Dimensions*, *in* INFORMATION TECHNOLOGY STANDARDS AND STANDARDIZATION 1-18 (2000).

28. For an informed discussion of the forces at work in settling on a standard gauge for railroads, see GEORGE HILTON, AMERICAN NARROW GAUGE RAILROADS (1990).

29. In order to move freight from Russia into Germany, Hitler's Germany had to offload freight from cars using one railroad gauge onto cars of another gauge. ALBERT L. WEEKS, RUSSIA'S LIFE-SAVER: LEND-LEASE AID TO THE U.S.S.R. IN WORLD WAR II 91 (2004).

30. Sometimes, countries switch their traffic directionality from one side of the street to the other. *See* Paul Friedlander, *H-Day is Coming to Sweden*, N.Y. TIMES, Aug. 20, 1967, § 10, at 1 (describing the transition from left side of the road driving to right side of the road driving that was to take place in Sweden on September 3, 1967); *see also* Scott Berinato, When Voice Becomes Data, CSO ONLINE, Sept. 21, 2006, http://www.csoonline.com.au/index.php/id;924061898;fp;16;fpid;0 (switch from driving on one side of the road to another in Sweden had no measurable effect on the accident rate in the long term). The arbitrariness of the choice of which side of the road to drive on can be seen by the importance placed on it by the "xenophobic, capricious, [and] superstitious" General Ne Win, former president of Burma (now Myanmar). *General Ne Win*, DAILY TELEGRAPH, Dec. 6, 2002, at 31. In addition to having been observed "in the middle of the night, dressed as a

benefits, as well. When a standard exists and is widely adopted, there is a greater pool of human knowledge to draw on regarding that standard. A Microsoft-certified engineer has better employment prospects than a computer specialist who knows only IBM System z servers.[31]

There are many reasons why markets for computer operating systems are subject to pressures of standardization. Purchasers' lives are made easier, because they don't have to worry about picking the "right" system. The old saying that "you never get fired for buying an IBM" today applies to Microsoft.[32] Users only have to be trained on one kind of system, and there are fewer worries about compatibility. Nevertheless, as discussed below, the homogenizing pressures of technology standards can have negative consequences.

B. "Monocultures" in Agriculture and Technology

The analogy between agriculture (and biology generally) and software is pervasive in discussions of computer security. The term "monoculture" itself has an agricultural origin, and many computer security threats have biological names, like "worms" or "viruses." The perceived threat from excess homogeneity in software is likened to the threat to crops and species from insufficient genetic diversity.[33] Therefore, it will be helpful in the understanding of the above-described economic effects to understand how they might apply to agriculture, as well as to complex technologies and computer networks.

The economic and cultural pressures on agricultural tend to create

---

king, walking backwards over a bridge in Rangoon, apparently on the advice of his soothsayers[,]" he directed his nation to begin driving on the right hand side of the road, instead of the left. *Id*.

31. On February 23, 2008, there were eight job postings in the "Computer/Software" category with "System z" as a keyword, but 152 job postings with "MSCE" as a keyword. Monster.com, Job Search, http://www.monster.com (Feb. 23, 2008) (on file with author).

32. BETH FOSS ET AL., IS CORPORATE AMERICA READY FOR OPEN SOURCE SOFTWARE? (2002), http://www.danmccreary.com/Open_Source_Report.pdf.

33. Jim Chen, *Webs of Life: Biodiversity Conservation as a Species of Information Policy*, 89 IOWA L. REV. 495, 505 (2004) ("

> Though the biosphere and the world of human-generated information teem with diversity, both are slouching toward uniformity. Driven by the value that inheres in networks and in the cost-reducing benefits of uniform operating standards, the quest for univeral [sic] interoperability in electronic communication and commerce has come close to realization. This quest has come dangerously close, in fact, for uniformity carries a cost of its own, in the natural realm as well as the electronic. "Never before in human history have there been comparable monocultures … of billions of genetically similar plants covering millions of acres across whole continents."

) (citing H. Garrison Wilkes, *Plant Genetic Resources over Ten Thousand Years: From a Handful of Seed to the Crop-Specific Mega-Gene Banks*, *in* SEEDS AND SOVEREIGNTY: THE USE AND CONTROL OF PLANT GENETIC RESOURCES 67, 73 (Jack R. Kloppenburg, Jr. ed., 1988)).

food economies heavily dependent on particular crops.[34] In the first instance, only certain plants are suitable for domestication.[35] However, among the possible candidates, species that were domesticated in one place were not domesticated in another.[36] In many instances, one plant is domesticated while its equally suitable near relatives are not.[37] The choice of exactly which plant to domesticate may therefore be seen as somewhat arbitrary—a standard, like what side of the road to drive on or the length of a meter.

While the reasons one plant species may be domesticated and another not may be complex, it is easier to continue to grow already-domesticated crops than to domesticate new ones. This demonstrates path dependence. Crops can also be viewed as being subject to indirect network effects, because many agricultural "applications," from particular formulations of pesticide to planting cycles, run on top of agricultural "platforms." Technological advances have greatly increased the pressure to rely on only a few crops.[38] While historically, farmers had to grow a wider variety of crops in order to effectively exploit their soil, modern fertilizers have limited the need for that kind of crop rotation.[39] Furthermore, while genetic diversity within a species was once the norm, commercial seed distribution has homogenized crops to an unprecedented degree. As a result of these pressures, currently, "[a] mere dozen species account for over 80 percent of the modern world's annual tonnage of all crops."[40] Michael Pollan has described how commercial pressures create incentives for farmers to rely heavily on monocultures,[41]

---

34. The same pressures apply to the agricultural use of animals. However, I will limit my discussion to plant-based agriculture.

35. JARED DIAMOND, GUNS, GERMS, AND STEEL 132-133 (2nd ed. 1999).

36. *Id.* at 133.

37. *Id.* at 134.

38. Kyuma, *infra* note 50, at 68 (

> Three technological factors pushed farmers toward monoculture. The first is mechanization, which enabled farmers to expand their farms. ... With a heavy investment in large, specialized machinery, the farmer has a strong incentive to grow only the crop for which the machinery was designed.
>
> The improvement of crop varieties is the second force pushing farmers toward monoculture…. By concentrating on a single, improved crop, the farmer can exploit its traits to the utmost.
>
> The third technological factor underlying the shift toward monoculture is the development of chemicals, i.e., fertilizers and pesticides, which have made it possible to grow a single crop year after year….

).

39. GUNS, GERMS, AND STEEL, supra note 35, at 134.

40. *Id*. at 132.

41. MICHAEL POLLAN, THE BOTANY OF DESIRE: A PLANT'S-EYE VIEW OF THE WORLD 231 (2002) ("Monoculture is where the logic of nature collides with the logic of

and his thinking on agricultural issues generally has been extremely influential.[42]

C. *Negative Side Effects*

Above, this Note briefly covered how different economic and technological phenomena work together to create a standard technology or product. In most cases, this standardization leads to great economic efficiency. However, the homogenizing results of these phenomena cause negative "externalities."

An externality is a cost or benefit to a party not involved in a transaction that is caused by the transaction.[43] Economic regulation is often focused on limiting externalities that have a negative social impact. For instance, A and B may enter into a transaction that is mutually beneficial. But that transaction may impose costs on C that exceed the benefit to A and B together. In those instances, the transaction is a net loss to society, and government regulation may seek to modify or prevent it. Alternatively, the transaction may be a net benefit to society, but equity concerns may motivate the government to limit the costs borne by third parties such as C. Markets that prominently feature negative externalities justify regulatory intervention.[44] There may be compelling reasons that lead to the creation of technology standards and software monocultures, but there may also be negative externalities and costs associated with those processes must be acknowledged.

The creation or maintenance of a monopoly may be an example of a negative externality caused by transactions that create technology standards. Monopolies and monocultures are different creatures, but they may contribute to one another. A firm that manages to have one of its technologies become a standard may have monopoly control of that technology, which it can maintain through patent, copyright, or otherwise. A monopoly is a firm that has little or no competitive pressure, and is able to charge high prices for a product it has no incentive to improve.[45] Public utilities and communications regulation

---

economics . . . .").

42. Pollan's thinking has started to shape the national debate over food policy. For one prominent example, then-candidate Obama at one point remarked that he had just read an article by Pollan, and went on to say that our agriculture system is "creating monocultures that are vulnerable to national security threats, are now vulnerable to sky-high food prices or crashes in food prices, huge swings in commodity prices, and are partly responsible [for various health care problems]." *The Full Obama Interview*, TIME, Oct. 23, 2008, http://swampland.blogs.time.com/2008/10/23/the_full_obama_interview.

43. John A. Rothchild, *The Social Cost of Technological Protection Measures*, 34 FLA. ST. U. L. REV. 1181, 1198 (2007).

44. *Id.* at 1204-05.

45. Some markets, such as utilities or telecommunications, can be described as "natural

has traditionally been premised on the presumed negative consequences of allowing a monopolist to control a network that many depend on without check.[46] Despite some similarities, however, monopolies are a different concept than monocultures. A monopoly is a single firm that has excess market power—but it may provide many different, robust technologies. It is at its core an argument about a lack of diversity among firms. By contrast, a monoculture may be supported by a variety of firms—it is an argument about a lack of technological diversity. Nevertheless, technology monocultures may *tend* to produce business monopolies.[47] Therefore, to the extent that monopolies are undesirable for their own set of reasons, it may be desirable to limit the technology monocultures that may contribute to them.

The heavy reliance on any single commodity can have wide-ranging economic repercussions, as it creates a "bottleneck" and a single point of failure.[48] One key example is the 1970s oil crises, where the world's dependence on a single commodity for much of its energy needs showed that even a modern industrial economy could be surprisingly fragile.[49] A farmer's reliance on a single crop can also cause him problems.[50] While agricultural monocultures carry certain economic benefits, there are also attendant risks. Large-scale agricultural monocultures, though efficient,

---

monopolies" that exhibit traits that limit competition. *See generally* NUECHTERLEIN & WEISER, *supra* note 4, at 12-15. Once a physical network of wires or pipes is built, it may be uneconomical for a new entrant to build a duplicate network to compete with the first one—even though the new entrant may have a more efficient technology. In the case of operating systems or other software, the high cost of building a user base may be such a high initial, fixed cost that the current dominant player can be seen as having a natural monopoly. As mentioned *supra*, note 1, consideration of monopoly issues is beyond the scope of this Note.

46. *See* ALFRED E. KAHN, THE ECONOMICS OF REGULATION 28 (1988).

47. For example, business monopolies can be created if the vendor behind the technology monoculture is insulated from competition through patent or copyright law.

48. The fear that our complex economy can be brought down by a single weak point is widespread. *See* Frank J. Cilluffo et al., *Bad Guys and Good Stuff: When and Where Will the Cyber Threats Converge?*, 12 DEPAUL BUS. L.J. 131, 141-42 (1999/2000) (

> Modern societies are dependent upon critical infrastructures, such as telecommunications, electric power, health services, banking and finance, transportation, and defense systems, as they provide a comfortable standard of living. These systems are increasingly interdependent on one another and damage to one can potentially cascade and impact others - with single point failures being of great concern.

).

49. MICHAEL CARR, NEW PATTERNS: PROCESS AND CHANGE IN HUMAN GEOGRAPHY 367 (1997).

50. K. Kyuma, *Protection of the Environment: Sustained Agriculture, Sustained Ecosystems*, *in* PHOSPHORUS REQUIREMENTS FOR SUSTAINABLE AGRICULTURE IN ASIA AND OCEANIA 57, 68 (1990) (It is seen as a problem that "monoculture, which is widely practiced in the United States as an efficient means to attain high crop productivity, may not be compatible with the other goal of a good farming system, i.e., sustained production through protection of the environment.")

require more modern agricultural products such as chemical pesticides than do mixed plantings.[51] Not only is the farmer more economically vulnerable to swings in the price of the crop, but everything he grows becomes susceptible to the same pests and diseases. Because 19th century Ireland depended on the potato for much of its nutrition, when the potato blight struck Ireland in 1845, mass starvation resulted.[52] Indeed, Pollan writes that "it was not the potato so much as potato monoculture that sowed the seeds of Ireland's disaster."[53] Genetic homogeneity carries risks outside of agriculture, as well. A genetically homogenous human population is more susceptible to endemic disease, and an ecosystem with many different species is considered more robust than a simpler ecosystem.[54]

Monocultures in crops, commodities, or technologies create economic fragility. They may contribute to the creation of a monopoly that is able to charge higher prices for its products, and when an economy depends heavily on a monocultural bottleneck for an important activity, threats to that single item can bring an entire economy to its knees.

D. *Software Monoculture*

While it is hard to argue against the virtues of standardization when it comes to light bulbs or soda can sizes, as with agriculture, the standardization of desktop operating systems has had certain negative side effects. By analogy with agriculture, the prevalence of Microsoft's products has been called a "software monoculture."[55] Just as large-scale plantings of single crops are susceptible to being wiped out by a single disease, a software monoculture can lead to a majority of the world's computers simultaneously becoming susceptible to the same security vulnerability; and just as biodiversity contributes to an ecosystem's robustness,[56] a more diverse software "ecosystem" may be less susceptible to security flaws.

It is important to bear in mind that Microsoft is neither the only company that has had a dominant position in a software market, nor the only example of such dominance leading to widespread security incident.

51. *See* POLLAN, *supra* note 41, at 225-26.

52. CORMAC Ó GRÁDA, BLACK '47 AND BEYOND: THE GREAT IRISH POTATO FAMINE IN HISTORY, ECONOMY, AND MEMORY 13, 203 (1999).

53. *See* POLLAN, *supra* note 41, at 230.

54. Charles C. Mann, 1491: NEW REVELATIONS OF THE AMERICAS BEFORE COLUMBUS 112-18 (Vintage 2006) (2005).

55. The term is certainly loaded. Describing Microsoft Windows as a "standard" has neutral or even positive connotations, while the term "monoculture" is extraordinarily negative. Nevertheless, it is the usual term used when discussing this issue.

56. Chen, *supra* note 33, at 549-50.

The Morris worm of 1988, the first computer worm to propagate itself over the Internet, took advantage of security vulnerabilities in sendmail and other programs to hobble the Internet.[57] It infected 6,000 Unix computers, crashed 10% of the Internet, and caused $100 million in damage.[58] The Morris Worm incident led to the creation of the Computer Emergency Response Team Coordination Center.[59] The widespread use of MoveableTypes's blogging software has made it easier for spammers to take advantage of weakness and post spam comments on blogs.[60] The Internet's dominant web serving software,[61] the open source program Apache, has occasionally been subject to a security vulnerability that put the majority of the world's web sites in danger simultaneously.[62] As John Quarterman writes, "[m]onoculture is not limited to operating systems or application software, nor even to application servers. Monoculture can exist in network routers as well. And if an exploit becomes widely known for a widely used router, big problems could result."[63] These examples show that the issue of monoculture is pervasive in computer technology. Nevertheless, most of the attention given to software monocultures has focused on Microsoft,[64] and the majority of the most widespread and severe security incidents have affected Microsoft products.[65]

One of the first Microsoft vulnerabilities to receive widespread attention was the "I love you" virus, which in late May 2000 spread rapidly throughout the world by taking advantage of flaws in VBScript, a simple programming language included in all versions of Microsoft Windows since 1998.[66] The "I love you" virus destroyed data on millions

57. MICHAEL ERBSCHLOE, TROJANS, WORMS, AND SPYWARE: A COMPUTER SECURITY PROFESSIONAL'S GUIDE TO MALICIOUS CODE 35, 36 (2005).

58. Thomas M. Chen & Jean-Marc Robert, *Worm Epidemics in High-Speed Networks*, 37 COMPUTER 48, 49 (2004).

59. *Id.*

60. Posting of Jacques Distler, to Musings, Software Monoculture, http://golem.ph.utexas.edu/~distler/blog/archives/2003_10.shtml#s000236 (Oct. 15, 2003).

61. For current statistics on Apache's market share, see Netcraft, Web Server Survey Archives, http://news.netcraft.com/archives/web_server_survey.html.

62. LWN.net, The Apache Vulnerability, Full Disclosure, and Monocultures, http://lwn.net/Articles/2756/ (June 18, 2002).

63. John Quarterman, *Managing Internet Risk in a Scale-Free World*, *in* SCIENCE AND SECURITY: INFORMING NEW ZEALAND 79, 81 (2005).

64. Amit Singh, A Taste of Computer Security, Unix vs. Microsoft Windows, http://www.kernelthread.com/publications/security/uw.html.

65. *See, e.g,* MARK F. GRADY & FRANCESCO PAIRISI, THE LAW AND ECONOMICS OF CYBERSECURITY 119 (2006) (a discussion of monoculture immediately raising the issue of Windows dominance).

66. An earlier program called "Melissa" was also very fast-spreading, but was relatively benign compared with "I love you." John Markoff, *April 30-May 6; An "I Love You" Virus Becomes Anything But*, N.Y. TIMES, May 7, 2000, § 4, at 2; John Markoff, *A Disruptive Virus Invades Computers Around the World*, N.Y. TIMES, May 5, 2000, § A, at 1.

of computers, and was the first volley in what was to be several years of fast-spreading and damaging computer viruses and worms. Another incident leading to greater consciousness of the problem of a security vulnerability being discovered and exploited on a large number of computers simultaneously was the "Code Red" worm (named in part because the researchers who identified it drank Code Red Mountain Dew to "fuel[] their efforts").[67] The Code Red worm demonstrated "the speed at which a malicious exploit of a ubiquitous software bug can incapacitate host machines."[68] Drawing a biological analogy, the authors also noted that "[a]s is the case with biologically active pathogens, vulnerable hosts can and do put everyone at risk."[69] High-profile computer worms demonstrated the risks and costs of a software monoculture could be very high. Insecure software was widely deployed to end users who may not have had much computer expertise, but "machines operated by home users or small businesses (hosts less likely to be maintained by a [sic] professional systems administrators) [were] integral to the robustness of the global Internet."[70] This widespread deployment of insecure software operated by nonexpert users led to several years of high-profile security exploits, as names like "Nimba," "Blaster," and "Slammer" joined the rogue's gallery with "I love you" and "Code Red."[71]

A more exotic phenomenon facilitated by the Microsoft software monoculture is the "zombie botnet." A "botnet" is a network of computers that have been infected by some computer worm that then connects them, unbeknownst to their owners, to a network of other computers that have also been so infected.[72] Each infected computer is known as a "zombie," and the botnet is then indirectly controlled by its "owner" to perform some nefarious act.[73] Common uses of botnets are Distributed Denial of Service attacks (DDoS), whereby many thousands of computers simultaneously try to access some resource on a target computer, overloading and perhaps damaging the target.[74] Another use is

---

67. David Moore, Colleen Shannon, & K. Claffy, *Code-Red: A Case Study on the Spread and Victims of an Internet Worm*, PROCEEDINGS OF THE 2ND ACM SIGCOMM WORKSHOP ON INTERNET MEASUREMENT 2002, at 273-74 (2002).

68. *Id.* at 282.

69. *Id.*

70. *Id.*

71. *See generally* Evan Cooke, Z. Morley Mao, & Farnam Jahanian, *Hotspots: The Root Causes of Non-Uniformity*, *in* SELF-PROPAGATING MALWARE, INT'L CONFERENCE ON DEPENDABLE SYS. & NETWORKS 179-80 (2006) (listing some of the "most significant worms to strike the Internet").

72. Lilian Edwards, *Dawn of the Death of Distributed Denial of Service: How to Kill Zombies*, 24 CARDOZO ARTS & ENT. L.J. 23, 26-27. (2006).

73. *Id.*

74. *Id.*

to send spam–it is difficult to block unsolicited email based on its source, when its source is thousands of seemingly unrelated computers located around the world.[75] The magnitude of botnets is hard to overstate. In 2005, Dutch police discovered and managed to shut down a botnet comprised of 1.5 million infected zombies,[76] and Vint Cerf has estimated that one in four computers connected to the Internet are part of one or more botnets.[77] Peter Gutmann, a computer scientist at the University of Auckland, recently stated that the "Storm" botnet could be viewed as the most powerful supercomputer in the world.[78]

II. RESPONSES TO MONOCULTURE

Concerns about the negative consequences of monocultures and dependence on bottleneck technologies or commodities are very domain-specific, although certain common features can be noted. The responses either seek to do away with the monoculture by increasing diversity in some way, or give the government a regulatory role in limiting the harms caused by the monoculture. In crops, different planting and crop rotation techniques can limit the bad effects of monocultures.[79] To counter the heavy reliance on oil, energy independence has become a watchword not only among environmentalists, but among those concerned with national security.[80] For most of the twentieth century, telecommunications regulation was premised on the assumption that telecommunications networks are natural monopolies.[81] More recently, the net neutrality

75. Jacqui Cheng, *Botnets Cause Significant Surge in Spam*, ARS TECHNICA, Oct. 30, 2006, http://arstechnica.com/news.ars/post/20061030-8111.html.

76. Gregg Keizer, *Dutch Botnet Bigger Than Expected*, INFORMATION WEEK, Oct. 21, 2005, http://www.informationweek.com/news/security/government/showArticle.jhtml?articleID=172303265.

77. Tim Weber, *Criminals "May Overwhelm the Web*," BBC NEWS, Jan. 25, 2007, http://news.bbc.co.uk/1/hi/business/6298641.stm.

78. *See* Insecure.org, World's Most Powerful Supercomputer Goes Online, http://seclists.org/fulldisclosure/2007/Aug/0520.html (archiving a message of computer science professor Peter Gutmann); Sharon Gaudin, *Storm Worm Botnet More Powerful than Top Supercomputers*, INFORMATIONWEEK, Sept. 6, 2007, http://www.informationweek.com/news/internet/showArticle.jhtml?articleID=201804528.

79. *See* James F. Power, *Legumes and Crop Rotations*, *in* SUSTAINABLE AGRICULTURE IN TEMPERATE ZONES 178, 198 (Charles A. Francis et al. eds., 1990).

80. *See* Stephen D. Solomon, *For National Security, Get Off Oil*, SCIENTIFIC AM., Oct. 2008, http://www.sciam.com/article.cfm?id=is-oil-a-threat (former CIA director R. James Woolsey sees oil dependence as a national security threat); Set America Free Coalition, An Open Letter to the American People, http://www.setamericafree.org/openletter.htm ("our present dependency creates unacceptable vulnerabilities. In Iraq and Saudi Arabia, America's enemies have demonstrated that they can advance their strategic objective of inflicting damage on the United States, its interests and economy simply by attacking critical overseas oil infrastructures and personnel.").

81. *See* NUECHTERLEIN & WEISER, *supra* note 4, at 55.

movement has advocated the regulation of ISPs in order to prevent them from becoming Internet gatekeepers.[82] The responses to perceived excess homogeneity in software, and its attendant negative consequences, have been similarly varied, ranging from regulation of the dominant software provider designed to increase interoperability, to a modification of tort and contract law principles, to more pragmatic approaches designed to better protect important computer systems from security vulnerabilities.

A. Geer

In 2003, a report titled "CyberInsecurity: The Cost of Monopoly–How the Dominance of Microsoft's Products Poses a Risk to Security" was published.[83] The report's principal author, Daniel Geer, was shortly thereafter fired from his position at @Stake, a computer security firm with ties to Microsoft.[84] This widely-publicized firing helped make the report (whose thesis was attention-grabbing in itself) famous. More than any other document, Geer's report kicked off the monoculture debate.[85] Although primarily focused on perceived problems with Microsoft's engineering practices, the report explored the risks of software monocultures generally. For instance, it notes that "[a] monoculture of networked computers is a convenient and susceptible reservoir of platforms from which to launch attacks . . . [t]his susceptibility cannot be mitigated without addressing the issue of that monoculture."[86] It further notes that "[t]he NIMDA and Slammer worms that attacked millions of Windows-based computers . . . spread from one to another computer at high rates. Why? Because these worms did not have to guess much about the target computers because nearly all computers have the same vulnerabilities."[87]

However, the bulk of Geer's argument is focused on problems specific to Microsoft.[88] He argues that certain engineering practices of Microsoft exacerbate network effects and create a level of consumer lock-in that would not otherwise exist.[89] He explains that Microsoft tightly

---

82. SaveTheInternet.com, Frequently Asked Questions, http://savetheinternet.com/=faq ("The nation's largest telephone and cable companies—including AT&T, Verizon, Comcast and Time Warner—want to be Internet gatekeepers, deciding which Web sites go fast or slow and which won't load at all.") (last visited May 5, 2009).

83. DAN GEER ET. AL., CYBER *IN*SECURITY: THE COST OF MONOPOLY (2003), http://www.ccianet.org/papers/cyberinsecurity.pdf (emphasis in original).

84. *See* Ellen Messmer, *Oh Dan Geer, Where Art Thou?*, NETWORKWORLD, Dec. 22, 2003, http://www.networkworld.com/weblogs/security/003879.html.

85. *See Warning: Microsoft "Monoculture"*, WIRED, Feb. 15, 2004, http://www.wired.com/politics/security/news/2004/02/62307.

86. GEER ET AL., *supra* note 83, at 7.

87. *Id.* at 10.

88. *See id.*

89. *Id.* at 13.

integrates its operating system and its application software in ways that give its own applications significant advantages, arguing that it uses "inter-module interfaces so complex, undocumented and inaccessible"[90] that no one outside Microsoft can effectively exploit them. He also argues that Microsoft integrates certain components, such as its Internet Explorer software, more deeply into the operating system than necessary from an engineering perspective, thereby making it difficult to replace or replicate Microsoft software or any of its components.[91] Geer argues that "[t]ight integration of applications and operating system achieves user lock-in by way of application lock-in. It works."[92] In other words, from a business perspective, Microsoft's engineering strategy has been a very successful means of holding on to customers.

An unintended side effect of Microsoft's engineering strategy, however, has been to increase the complexity of its products. As Geer points out, "[t]he central enemy of reliability is complexity."[93] By achieving user lock-in through creating a high level of dependence between different pieces of software, Microsoft has created a software ecosystem that is both dominant (because difficult to switch away from or replace) and highly unreliable (because overly complex). He argues that "[a]bove some threshold level of code complexity, fixing a known flaw is likely to introduce a new, unknown flaw"[94] and that the Microsoft's code base is unlikely to ever become secure.[95]

Finally, while Geer notes an increased awareness of security issues in Microsoft at the time of his article's publication, he worries that certain solutions then pushed for by Microsoft, such as those known under the rubric of "Trusted Computing," could serve to increase Microsoft's dominance.[96] His solution to these problems is quite broad. For instance, he proposes requiring that Microsoft release comparable versions of its application software such as Office for Mac OS X and Linux before it is allowed to release Windows versions. He would also support requiring thorough and open documentation of Microsoft APIs, to allow better competition with its products.[97]

In the "Coda" section of the report, Geer writes that "[t]hese comments are specific to Microsoft, but would apply to any entity with similar dominance under current circumstances. Indeed, similar moments of truth have occurred, though for different reasons, with IBM

90. *Id.* at 13.
91. *See id.*
92. GEER ET AL., *supra* note 83, at 13.
93. *Id.* at 14.
94. *Id.* at 15.
95. *Id.*
96. *See id.* at 16-17.
97. *See id.* at 18-19.

or AT&T."[98] It is true that dominant firms are often accused of tying their products together, and attempting to unfairly leverage a powerful position in one market into a powerful position in another market. Dominance by particular firms has long been a phenomenon of high technology and telecommunications markets. However, no company besides Microsoft has been accused using bad engineering practices to accomplish a kind of tying that has such severe negative security consequences. There are no other firms with "similar dominance under current circumstances,"[99] and new ones are unlikely to arise. In its specifics, Geer's monoculture argument is applicable to Microsoft alone.

B. Picker

Randal Picker agrees that insecure software can be a real problem, but he argues for what he sees as a more cost-effective response.[100] In the first instance, he does not disagree that the rise of the networked economy has been accompanied by regrettable side effects. He argues that just as networking computers together has given rise to positive externalities in the form of what Yochai Benkler has described as "shareable goods," it has also given rise to negative externalities.[101] He cites spam and phishing as prominent examples, and goes on to a broader discussion of problems of computer security.[102]

Picker does not address the arguments made by Geer that network effects and negative security consequences are exacerbated by specific engineering choices made by Microsoft. Neither does he deny that homogenous networks can have negative consequences. He writes that "there is a real downside to all of this connectivity: problems percolate quickly throughout an interconnected system, and problems that might have been just local disturbances end up everywhere."[103] He later continues that

> [t]he monoculture is another name for a homogenous, connected system. In the monoculture framework, heterogeneity is used as a barrier to the spread of a virus throughout a connected computer system. The anti-monoculture idea also taps into our sense of

---

98. GEER ET AL., *supra* note 83, at 20.
99. *Id.*
100. Randal C. Picker, *Cyber Security: Of Heterogeneity and Autarky*, (Univ. of Chicago Law Sch. John M. Olin Law & Economics Working Paper No. 22, 2004), *available at* http://ssrn.com/abstract=590927).
101. *Id.* at 1-5.
102. *Id.*
103. *Id.* at 12.

necessary biodiversity.[104]

Although he has many quibbles with the specifics of the monoculture argument as laid out by Geer and others, the thrust of his argument is quite simple: Even if Geer is right about the security consequences of software monoculture, a simpler and more cost-effective solution than forced heterogeneity is what he calls "autarky"—simply isolating important computer systems from the Internet, so that they are immune to the negative externalities associated with networking computers together.[105] Picker's argument does little, however, to address the concerns that remain for those machines that, for various reasons, must stay connected to the Internet.

C. *Government Support of Open Source*

Some governments have sought to counteract problems in the software market by adjusting their policies in ways that benefit alternatives—primarily open source software.[106] For instance, in 2001, several Brazilian municipalities began giving open source software preference.[107] There are many similar initiatives throughout the world.[108] These actions are undertaken for a variety of reasons, not all of which are specifically aimed at reducing software monoculture. But some actions, such as the Japanese government's recently-announced policy to promote open source software, are expressly designed to lessen their dependence on Microsoft software.[109] Government policies favoring software diversity, even if they are limited to shaping the government's own purchasing decisions, have the potential to reduce monoculture by sustaining alternative products that otherwise would not thrive in the marketplace.

These policies, however, are not without their critics. David S. Evans and Bernard J. Reddy argue that government preferences for open source software will likely cause more problems than they solve.[110] They

---

104. *Id.*

105. *Id.* at 6.

106. *See* ROBERT WILLIAM HAHN, GOVERNMENT POLICY TOWARD OPEN SOURCE SOFTWARE 4-5 (2002) ("Open source" software has many interesting characteristics and has license terms that prevent vendor lock-in. From the perspective of governments, however, sometimes its chief selling point is that it is not from Microsoft.)

107. *Id.* at 5.

108. For a frequently updated list of resources concerning the use of open source software in government, see California Environmental Protection Agency Air Resources Board, Government-Related Open Standards/Open Source Software Articles, http://www.arb.ca.gov/oss/articles/articles.htm.

109. Phil Hochmuth, *The Japanese Government Looks to Go Open Source*, LINUXWORLD, May 9, 2007, http://www.linuxworld.com/newsletters/linux/2007/0507linux2.html.

110. David S. Evans & Bernard J. Reddy, *Government Preferences for Promoting Open-*

note that they are

> aware of no compelling evidence that governments have special expertise in analyzing the software industry to effect solutions . . . . Whether 'open code' in any given situation is actually 'as powerful' as 'closed code' is an everyday business judgment that should be made by businesses, governments, and private users; it does not strike us as a policy issue that should be decided by bureaucrats or legislators, or even by lawyers and economists.[111]

If open source software (or any software alternative) has advantages, then government IT professionals and purchasers would be well-advised to carefully consider those products when making their purchasing decisions, as part of their business judgment.[112] Even skeptics of the monoculture argument acknowledge that it can be rational to take into account the effects of buying into or supporting a software monoculture when making a technology choice.[113]

But, since government decisions can have negative unintended consequences and can distort markets, governments should hesitate before fixing software preferences as law or policy.

D. Extension of Law

Several commentators argue that legal reform increasing liability for software vendors who ship insecure products may alleviate negative consequences of technology monoculture. As Robert W. Hahn and Anne Layne-Farrar note, "the liability rules governing the distribution and use of software remain unclear, even after some thirty years of debate."[114] A few scholars have introduced proposals to clarify those rules. They have generally noted that the current legal climate does not assign liability to those parties best able to bear it, and certain behaviors in the marketplace create negative externalities for third parties. Through various means, they propose to realign the economics of software security by internalizing negative externalities. By causing costs to be borne by those who create them, they attempt to define a legal and economic

---

Source Software: A Solution in Search of a Problem, 9 MICH. TELECOMM. & TECH. L. REV. 313, 394 (2003) ("The net effect is likely to be a reduction in the total 'externality' benefits of software.").

111. *Id.* at 393-94.

112. *See id.* at 394.

113. Greg Goth, *Addressing the Monoculture*, IEEE SECURITY AND PRIVACY, Nov/Dec 2003, at 8-10 (quoting John Carroll as saying "[i]t is good that consumers factor monoculture costs into their calculations when choosing a particular platform. It is not good to treat those costs as more important than any others.").

114. Robert W. Hahn & Anne Layne-Farrar, *The Law & Economics of Software Security*, 30 HARV. J.L. & PUB. POL'Y 283, 327 (2006).

environment more likely to result in secure software. After briefly touching on ideas of Pamela Samuelson and Jennifer A. Chandler, this section will focus in detail on Emily Kuwahara's approach.

1. Chandler, Samuelson

Jennifer A. Chandler has argued that given the unique nature of Distributed Denial of Service attacks,[115] only holding software vendors liable would properly internalize risks to those most able to prevent them. Unlike many other security problems, DDOS attacks can cause harm to computers that do not themselves have any security vulnerabilities—and the users whose computers *are* compromised may not suffer any economic loss. Increasing the liability for the creators of insecure software is therefore the only way to create incentives to prevent the harm. Chandler therefore proposes to create a new tort of "negligently creating an unreasonable risk of harm from third parties."[116]

In an article from the early days of the online revolution, Pamela Samuelson notes that the policy reasons explaining why information vendors are generally not held liable in the same way that products vendors are for defects or errors do not necessarily apply to software or electronic information.[117] Concerns about free expression have led courts to limit liability for defective or erroneous information to defamatory statements and situations where a person claims to have specialized knowledge (for instance, through malpractice actions against doctors or lawyers).[118] But some kinds of "information" seem more like products than like communications, and Samuelson observes that, in a case involving aeronautical charts, there is precedent for treating "information" as a product governed by standard liability rules.[119] She notes that in cases where an information product "behaves like a machine," courts are likely to apply products liability principles,

---

115. Distributed Denial of Service ("DDOS") attacks occur when a large number of computers simultaneously attempt to access resources on a remote computer. In one common scenario, a large number of computers are compromised by software that allows them to be remotely controlled by a malicious hacker. Those compromised computers then simultaneously send common network requests to a target computer system, overtaxing its ability to deal with them. A DDOS attack therefore allows a computer that is not itself subject to any particular security vulnerabilities to be brought down by a large network of computers that are. *See* Jennifer A. Chandler, *Security in Cyberspace: Combatting Distributed Denial of Service Attacks*, 1 U. OTTAWA L. & TECH. J. 231, 236 (2004).

116. *Id.* at 261. Much of her discussion concerns Canadian cases, although the principles discussed are applicable in American law.

117. Pamela Samuelson, *Liability for Defective Electronic Information*, COMMC'N OF THE ACM, Jan. 1993, at 21.

118. *Id.* at 21-22.

119. *Id.* at 23-24.

including, in some cases, strict liability.[120]

2. Kuwahara

Emily Kuwahara argues that product liability law could be extended to hold software vendors liable for defective products, provided the current liability disclaimers are invalidated and an exception is created to the economic loss rule.[121] Kuwahara observes that the "prevalence of viruses and worms on the Internet is astounding",[122] noting that an unprotected computer on the Internet has a 94% chance of being infected within an hour.[123] She writes, though, that the current state of case law suggest that recovery is not available against a software vendor such as Microsoft either in cases of extensive damage caused by a widespread security incident, such as the Slammer worm, or in situations where an individual brings an action after her personal computer has been hacked.[124] She offers a survey of the thinking about increased tort liability for software vendors, from Howard Schmidt, "who oppose[d] liability for software companies because it will raise costs and prices, stifle innovation, and lead to job cuts,"[125] to Bruce Schneier, "who strongly believes that the cost of insecure software is an externality that should not be borne by users, but by software companies."[126] She also discusses more exotic proposals, such as the creation of a new tort of "negligent enablement of cybercrime,"[127] or the creation of a code of professional practice for software engineers, which would open the door to malpractice actions.[128] She also notes the argument that Microsoft's dominant market places special burdens on it that wouldn't necessarily be shared by other software vendors. For instance, the Computer and Communications Industry Association issued a report that placed a "special burden . . . upon Microsoft because of [the] ubiquity of its product."[129]

Kuwahara goes on to detail a number of policy rationales for allocating risk to Microsoft particularly, including: compensation of victims; a lack of competition that reduces its incentives to increase its software's security; its superior ability to bear financial risk; the beneficial

---

120. *Id.* at 21.
121. *See* Emily Kuwahara, *Torts v. Contracts: Can Microsoft Be Held Liable to Home Consumers For Its Security Flaws?*, 80 S. CAL. L. REV. 997, 1030 (2007).
122. *Id*. at 1000.
123. *See id*.
124. *Id*. at 998-99.
125. *Id*. at 1001-02.
126. *Id*. at 1002.
127. Kuwahara, *supra* note 121, at 1003.
128. *Id*. at 1002-03.
129. *Id*. at 1007.

establishment of a standard of care in software design; the reluctance to allow Microsoft to use contract law to evade liability when consumers often have little choice but to use its products; and little actual bargaining ever occurs; the fact that liability insurance would likely remain available and affordable for all software companies; the fact that increased liability hasn't deterred innovation in other fields; and the fact that mere disclosure of software flaws does not offer consumers a sufficient remedy.[130]

At its core, her argument is that, because products liability has been successful in other areas of commerce, it is likely to be successful in software, as well. It is rooted in an assumption that software is best understood as a "product" or "good"[131] more similar to an automobile than a service.[132] Her alternative approach of a non-disclaimable statutory warranty offers a reasoned compromise to tort liability and addresses the imbalance in bargaining power between large software vendors and consumers. Her argument, however, depends on a number of historically-bound circumstances. It may make sense to be skeptical of adhesion contracts in the context of Microsoft, given that most consumers see no choice but to run its software, and must accept the terms of its licenses. It is also true, however, that in recent years Microsoft competitors, such as Apple, have met with increasing success,[133] and the web is increasingly becoming an important platform for software development. At the margins, at least, these developments may have an effect on how Microsoft does business. Because Kuwahara's argument depends heavily on facts that are specific to Microsoft, and because the market may already be acting to curb some of Microsoft's perceived defects, it is probably premature to adopt her proposed reforms. Additionally, if the software security problem remains primarily a Microsoft problem, as opposed to a problem that is endemic to an industry, it may be more prudent to enact regulations that target Microsoft particularly. Introducing principles of general application based on the behavior of a single company may have unintended consequences on non-culpable parties.

E. Policy Should Not Be Based on Contingent Circumstances

An analysis that proposes to introduce changes to the legal

---

130. *Id*. at 1012-15.
131. *Id*. at 1019-20; *see* Samuelson, *supra* note 117.
132. Kuwahara, *supra* note 121, at 1014; Samuelson, *supra* note 117.
133. 2007 saw Macintosh hardware sales jump by as much as 40% over the previous year, which is a growth rate between two and three times higher than the computer industry average. Charles Jade, *Apple 2007: Best Year Ever*, ARS TECHNICA, Dec. 24, 2007, http://arstechnica.com/journals/apple.ars/2007/12/24/apple-2007-best-year-ever.

environment is best made without too much reliance on the historically specific (and likely transitory) circumstance of a dominant software firm also having extremely vulnerable products. This circumstance was brought about by a number of specific businesses, technological, and historical reasons and is unlikely to be reproduced again. For example, Windows was initially developed for computers that had rare and transient network connectivity. Today's always-on broadband environment changes that, and increases the exposure of the computer to the outside world for attacks.[134] Computer systems designed for the older world have shown themselves to be not very well suited for the new world, and security incidents proliferated. However, as the risks of always-on network connections have become internalized by software developers, it is likely that the number of vulnerabilities will decrease. For example, in its first year of deployment, Vista had fewer security vulnerabilities than either Windows XP or Mac OS X.[135]

The risk of unintended consequences is too great to justify a change to the law unless there is a real, concrete problem to be addressed. A poorly calibrated liability regime could result, for instance, in more money being spent to prevent security vulnerabilities than the vulnerabilities themselves are likely to cause, resulting in a net social cost. As Steven Pinker suggests, it may sometimes be better to look for practical, engineering solutions to social problems, than to immediately think of redesigning the legal environment. He writes,

> [t]here are many other issues for which we are too quick to hit the moralization button and look for villains rather than bug fixes. What should we do when a hospital patient is killed by a nurse who administers the wrong drug in a patient's intravenous line? Should we make it easier to sue the hospital for damages? Or should we redesign the IV fittings so that it's physically impossible to connect the wrong bottle to the line?[136]

While economic incentives may cause software providers to develop new technologies and improve their products' security in ways they would not have done absent those incentives, without a *technological* solution to the underlying problems that cause software insecurity,

134. Pratyusa K. Manadhata & Jeannette M. Wing, Attack Surface Measurement, http://www.cs.cmu.edu/~pratyus/as.html ("Intuitively, a system's attack surface is the set of ways in which an adversary can enter the system and potentially cause damage. Hence the larger the attack surface, the more insecure the system.").

135. Michael Calore, *Microsoft: Vista Has Fewer Security Flaws in First Year Than XP, Mac OS*, WIRED, Jan. 24, 2008, http://blog.wired.com/monkeybites/2008/01/microsoft-vista.html.

136. Steven Pinker, *The Moral Instinct*, N.Y. TIMES, Jan. 13, 2008, § 6 (Magazine), at 632.

modifications to the liability regime of software markets will amount to little more than a series of transfer payments. Such modifications may be justified as matters of equity, or to harmonize software liability with other areas of products liability. But market and social incentives over the past several years have already increased the focus of the software industry on security issues. Given that the current incentives to create secure software may be adequate, and given that software security has measurably increased in the past few years, changes to the liability environment for software may be premature, and the risk of unintended consequences may be too great, to justify any drastic changes solely on the basis of improving security and counteracting the negative security consequences of a software monoculture.

III. TECHNOLOGY HAS PROVEN SUFFICIENT TO DEAL WITH MOST COMPUTER SECURITY ISSUES

As noted above, discussions of the negative security consequences of software monocultures are generally focused on the problems of a Microsoft monoculture particularly. While any software monoculture can be threatened by the rapid exploitation of a software vulnerability (and, as demonstrated by the Internet Worm, non-Microsoft monocultures have been), in the case of Microsoft, the monoculture effect is seen as a "force multiplier"[137] that greatly increased the effects that are ultimately caused by flawed software in the first place. Therefore, my analysis of the proper policy response to a software monoculture will be based primarily on an analysis of the factors that have led to Microsoft's products being widely viewed as insecure, and on the responses that Microsoft has deployed in order to deal with this problem. It is also informed by an understanding that government interventions in markets often have unintended consequences. As Hahn and Layne-Farrar write,

> From an economist's perspective, before the government decides to intervene to impose software security, it must be reasonably certain that private parties are unable to do so on their own. In other words, it must be clear that the market failed in some way. Otherwise, interventions run the risk of interfering with properly functioning

137. "Force multiplication" is a military concept whereby some factor increases a unit's combat potential. A force multiplier can be favorable weather, decoys, or even sunscreen. About.com: US Military, "Force Multiplier", http://usmilitary.about.com/od/glossarytermsf/g/f2536.htm. Network effects have been analogized to the concept of a force multiplier. *See* LTC Roland Ng Kian Huat, *Force Multiplication Through Network And Networking: A Frame For Discourse*, POINTER: J. OF THE SINGAPORE ARMED FORCES Vol. 30 No. 2 (2004), *available at* http://www.mindef.gov.sg/imindef/publications/pointer/journals/2004/v30n2/features/feature4.html.

> markets and, therefore, of introducing inefficiencies where none existed before–what could be termed a "government failure" as opposed to a market failure.[138]

After a comprehensive review of the marketplace for computer security, those authors remain skeptical that government intervention is needed. They even point out that seemingly benign reforms, such as a "lemon law" for software, could have negative consequences.

Because, as discussed below, technological solutions to many fundamental computer security issues (including the problem of monoculture itself) appear to be making progress, in order to avoid potential negative consequences, the government should not regulate to increase software diversity.

Geer's analysis of the problematic nature of Microsoft's software engineering principles is sound.[139] However, it bears keeping in mind that Microsoft is a software company that became successful in a time before ubiquitous, always-on computer networking. Indeed, broadband adoption is not yet complete: in 2007, 23% of Internet users still used dial-up connections.[140] Microsoft's design strategies may have always been bad from a software engineering standpoint. But most computer worms, virus and trojans today spread over the Internet.[141] In the days where the primary vector of computer malware transmission was the floppy disk or BBS downloads,[142] many computer vulnerabilities would simply not be exploited. The penalty throughout the 1980s and 1990s for insecure software design was not as severe as it is today. It is reasonable to assume that even without any policy action, Microsoft's software engineering strategies will change to reflect the new, networked reality.

In fact, Microsoft's approach to software engineering *has* changed in the past several years. The year before Geer's paper, Microsoft issued its "Trustworthy Computing" whitepaper. This paper called for a fundamental reengineering of computers, down to the level of the

138. Hahn & Layne-Farrar, *supra* note 114, at 299.

139. Indeed, despite the progress Microsoft has made in increasing the security of Windows Vista, Windows is still widely seen as overly complicated and slowed in its development cycle by Microsoft's commitment to retain backwards compatibility with older software, and hardware compatibility with as much of the PC ecosystem as possible. *See* Steve Lohr & John Markoff, *Windows Is So Slow, But Why?; Sheer Size Is Causing Delays for Microsoft*, N.Y. TIMES, Mar. 27, 2006, § C, at 1.

140. JOHN B. HORRIGAN & AARON SMITH, HOME BROADBAND ADOPTION 2007 1 (2007), http://www.pewinternet.org/pdfs/PIP_Broadband%202007.pdf.

141. *But see* Gregg Keizer, *Best Buy Sold Infected Digital Picture Frames*, N.Y. TIMES, Jan. 23, 2008, http://www.nytimes.com/idg/IDG_002570DE00740E18002573D9007CF01E.html.

142. *See* DAVID J. STANG, NETWORK SECURITY 237 (1992) (a contemporary source describing PC malware of the early 1990s).

microprocessors, with the aim of increasing security and preventing unauthorized code from running. Many, including Geer himself, have criticized that paper's proposals, arguing that the proposal for a Next Generation Secure Computing Base, commonly referred to as "Palladium," threatened to put too much control of what software can run on a computer into too few hands and to exacerbate the risk of vendor lock-in.[143] Microsoft has since abandoned the most ambitious of its "trusted computing" plans.[144] Although overly ambitious and perhaps misguided, the Trusted Computing whitepaper did at least demonstrate an increased awareness of security issues.

Several other initiatives have had more of a practical impact. In 2002, Microsoft undertook a two-month hiatus in the development of its software in order to focus on security concerns.[145] It has shown itself to be more nimble in its response to problems as they are uncovered.[146] Its research arm has begun to look for long-term security solutions that, unlike secure computing, do not rely on changes to hardware.[147] However, Microsoft's improved dedication to security issues can most clearly be seen on a practical level by looking at a few of the security-related improvements found in the most recent version of the Windows operating system, Vista.[148]

One longstanding weakness in Windows had been that it possessed a "file permissions system" that did not adequately prevent untrained users or rogue programs from making damaging changes to the operating system. Vista addresses this by introducing a more robust, Unix-style permissions system whereby even computer administrators need to supply a password before certain settings or files can be changed. Under Vista, Internet Explorer now runs in a "sandbox" that makes it so neither it, nor any programs it spawns (such as malware from a web site) can do much damage to the underlying system.[149] Vista also contains security features designed to prevent a user's computer from becoming part of a botnet,[150] and the most notorious current worm, Storm, which makes

---

143. GEER ET AL., *supra* note 83, at 16.

144. Paula Rooney, *Microsoft Shelves NGCSB Project as NX Moves to Center Stage*, CHANNELWEB, May 5, 2004, http://www.crn.com/security/18841713.

145. Peter Judge, *Microsoft Security Push Cost $100m for .Net Server Alone*, ZDNet UK, Jul.CO. UK, July 2, 2002, http://news.zdnet.co.uk/internet/0,1000000097,2118314,00.htm.

146. Matt Mondok, *Microsoft Sets Company Record with WMF Patch*, ARS TECHNICA, Jan. 7, 2006, http://arstechnica.com/journals/microsoft.ars/2006/01/07/2394.

147. Jeremy Reimer, *Microsoft Hefts A Heavy Mithril BrowserShield*, ARS TECHNICA, Sept. 5, 2006, http://arstechnica.com/news.ars/post/20060905-7668.html.

148. Grant Gross, *Microsoft Talks Up Vista Security in DC*, INFOWORLD, Jan. 30, 2007, http://www.infoworld.com/article/07/01/30/HNdcvistalaunch_1.html.

149. MARK JUSTICE HINTON, PC MAGAZINE WINDOWS VISTA SOLUTIONS 70 (2007).

150. Microsoft, Bots, Botnets, and Zombies,

computers it infects part of the Storm botnet, currently does not infect Windows Vista.[151]

These examples show that there are often technological solutions to problems created by technology–solutions that make a policy response unnecessary. One technological change in particular, however, has the potential to alleviate many of the negative externalities caused by software monocultures. This technology, Address Space Layout Randomization (ASLR), uses software techniques to produce a kind of virtual diversity, limiting the vectors by which malware can spread.[152] Elements of software traditionally load into a particular part of a computer's memory. Malware can take advantage of this fact to more easily spread from one computer to another. ASLR reduces the ability of malware to spread from one computer to another by randomly changing the memory location software loads into.[153] As Ollie Whitehouse writes,

> ASLR is a prophylactic security technology that strengthens system security by increasing the diversity of attack targets. Rather than increasing security by removing vulnerabilities from the system, ASLR makes it more difficult to exploit existing vulnerabilities. . . . By randomizing the memory layout of an executing program, ASLR decreases the predictability of that layout and reduces the probability that an individual exploit attempt will succeed.[154]

Although ASLR is not a new technology, its inclusion in Windows Vista shows technological methods taken by Microsoft can lessen the effects of software monoculture. It is the flexible nature of software that gives it the ability to create virtual diversity of this sort–it is difficult to imagine an analogous solution to the problem of, for example, agricultural monoculture. The impressive number of technological solutions Microsoft has brought to bear in Vista in order to address software security should at least argue in favor of giving technology, rather than law and policy, the chance to solve problems in computer security.

Only time will tell whether Vista's improved security model will indeed lead to a more secure system in the long term. But the early signs

---

http://www.microsoft.com/mscorp/safety/technologies/bots.mspx (Last accessed Mar. 8, 2009).

151. Posting of Jim Thompson to Chron.com TechBlog, This Worm is One Quiet Storm, Houston Chronicle Tech Blog, http://blogs.chron.com/techblog/archives/2007/10/is_this_worm_the_perfect_storm.html (Oct. 14, 2007).

152.. OLLIE WHITEHOUSE, AN ANALYSIS OF ADDRESS SPACE LAYOUT RANDOMIZATION ON WINDOWS VISTA 4 (2007).

153. *Id.*

154. *Id.*

are encouraging. For instance, Peter Bright noted in January 2008 that "[a]fter a year on the market, Vista has had fewer security vulnerabilities discovered than XP did in its first year. According to a post on the Windows Vista Security blog, Vista has had 36 fixed and 30 unfixed security vulnerabilities, compared to 68 fixed and 54 unfixed vulnerabilities in XP. Patches have been issued on 9 occasions so far with Vista, compared to 26 for XP."[155]

CONCLUSION

The fact that Microsoft has improved the security of its flagship product in the absence of the kinds of reforms argued for by Geer, Kuwahara, and others, argues against the need for government action or legal reform as a means to improve computer software security. Nevertheless, some of the reform proposals may have other reasons that would justify their adoption. It may be that increased tort liability against Microsoft and other software vendors for shipping vulnerable products is justified from principles of equity—software companies may be superior risk-bearers, even if the added financial incentives were not necessary to get them to improve their products' security. Geer's proposals for forcing Microsoft to be more "open" may be justified as a means of increasing competition in the software market, or as a means to reduce the risk of monopoly. Certain measured responses may be justified even in the absence of evidence that they are necessary to counteract the problems of a software monoculture. For instance, Picker's autarky proposal is probably a sound prophylactic engineering practice under any circumstance. Governments and companies desirous of avoiding vendor lock-in should consider using open data formats, communications protocols, and software. Education of IT buyers could lead to an increase in the awareness of alternative software, which may have its own merits. Finally, governments should ensure that their actions do not *promote* the creation of a software monoculture unnecessarily.[156]

Extraordinary efforts by governments are not needed to address what appears to be a transient, technology-driven problem. In recent years, Microsoft has undertaken a number of security initiatives and adopted a number of new security technologies, including those like Address Space Layout Randomization that partially undermine the

155. Peter Bright, *Microsoft: Vista's Not as Insecure as XP. Please Buy It!*, ARS TECHNICA, Jan. 26, 2008, http://arstechnica.com/news.ars/post/20080126-microsoft-vistas-not-as-insecure-as-xp-please-buy-it.html.

156. Similarly, it has been argued that governments at the very least ought to end subsidies that increase agricultural monocultures to levels perhaps beyond what the market itself would produce. *See* Michael Pollan, *You Are What You Grow*, N.Y. TIMES, Apr. 22, 2007, § 6 (Magazine), at 615.

monoculture argument. The arguably flawed nature of Microsoft products should be a concern for IT managers and technologists, not policy-makers. If Microsoft's continued dominance is to be challenged by regulators, it should be because of established, economics-based antitrust reasons, and not under the guise of an attempt to improve computer security.

WEAVING THE NAVAJO.NET

DAVID B. WILSON*

* The Author practices intellectual property law at the Denver office of Townsend and Townsend and Crew, where he focuses on patent prosecution within the firm's Electronics and Software Group. He received his J.D. from the University of Colorado in 2008 and was Editor-in-Chief of the Journal on Telecommunications & High Technology Law for volume 6. He would like to thank Professor Sarah Krakoff along with Michelle Albert, Laura Makar, Melissa Pingley, Ethan Plaut, Richard Williams, and Lisa Yellow Eagle for reviewing and commenting on the Article, along with JTHTL article editor Brian Wolf for his incredible help in editing the final version of this article. Professors Brad Bernthal, Richard Collins, Dale Hatfield, Phil Weiser, and Charles Wilkinson also deserve endless thanks for helping inspire the Author's interest in both telecommunications and federal Indian law. Thanks also to Faegre & Benson for making the Silicon Flatirons writing competition possible. Lastly, the Author wishes to thank his family members for their support and encouragement.

INTRODUCTION

> Today indigenous peoples must use the arrows of 0's and 1's and satellites. Information is a way to overcome today's monsters.[1]

Speaking before the World Summit on the Information Society, Navajo Nation President Joe Shirley, Jr.'s statement captures a vision of the telecommunications revolution the Navajos find themselves leading today. A geosynchronous satellite orbiting thousands of miles above the rugged remote Four Corner landscape is breaking down the "Digital Divide"[2] and creating a communication revolution for North America's largest native tribe. Over the past ten years, the tribe has begun a process of technological integration and innovation that will play an important role in shaping, developing, and preserving Navajo culture.

Ten years ago, the U.S. Commerce Department's National Telecommunications and Information Administration ("NTIA") began using the term "Digital Divide" to capture the growing separation between the Internet "haves" and "have-nots."[3] The NTIA report showed that Native Americans living on rural reservations lagged considerably behind the U.S. national average in access to the Internet and to computers.[4]

Historically, extending even basic telephone service onto tribal lands has proven difficult. The Navajo Nation's remote and rugged landscape, low population densities, and low incomes have discouraged the deployment of telephone infrastructure.[5] Even with the backing of

---

1. Tara Tidwell Cullen, *Sovereignty Unplugged: Wireless Technology and Self-Governance in the Navajo Nation*, CULTURAL SURVIVAL Q., June 15, 2004, at 2 (internal brackets omitted) (quoting Navajo Nation President Joe Shirley, Jr. at preparation meeting for the World Summit on the Information Society), http://www.culturalsurvival.org/ourpublications/csq/article/sovereignty-unplugged-wireless-technology-and-self-governance-navajo-nat.

2. *See* National Telecommunications and Information Administration, Falling Through the Net: Toward Digital Inclusion 13 (Oct. 2000) (discussing computer and internet access for Native Americans), *available at* http://search.ntia.doc.gov/pdf/fttn00.pdf.

3. *Id.* at 2-3.

4. *Id.* at 12-13; *see also* Press Release, The Navajo Nation, Navajo President Joe Shirley, Jr., receives tribal leadership technology award at 20th RES 2006 (Feb. 12, 2006) (noting that in 2002, 15% of Navajo families had computers and only 10% had Internet access), http://www.navajo.org/images/pdf%20releases/George%20Hardeen/feb06/Navajo%20President%20receives%20ICT%20award%20at%20RES%202006%20%20for%20Feb%2013.pdf [hereinafter Shirley Press Release 1].

5. *See* USDA RURAL DEVELOPMENT TELECOMMUNICATIONS PROGRAM SUCCESS STORY: COMMUNITY CONNECT PROGRAM; TRADITIONAL TELEPHONE PROGRAM HUERFANO, NM (Oct. 2007), http://www.usda.gov/rus/telecom/highlights/huerfano-comm-connect-success.pdf; *see also* FEDERAL COMMUNICATIONS COMMISSION, TELEPHONE SUBSCRIBERSHIP ON AMERICAN INDIAN RESERVATIONS AND OFF-RESERVATION TRUST LANDS 1 (May 2003), http://www.fcc.gov/Bureaus/Common_Carrier/Reports/FCC-

Universal Service Fund ("USF") money, which provides federal dollars to back basic phone service to rural America, including tribal lands, the Navajo Nation has remained far behind the national average, and even behind the average of other tribes, in obtaining phone service.[6] Data from the 2000 Census estimated that 67.9% of Native American households on reservations had telephone service.[7] In stark contrast, Navajo telephone penetration rates fell far below this average at 37.4%.[8] At the same time, 94% of non-native rural households had telephones.[9]

Since 2000, the Navajo Nation has begun to overcome this communication divide with the help of earth-orbiting satellites.[10] This development reflects the shift towards more spectrum-based technology[11] that has occurred across the United States over the past decade. This technological revolution through wireless technology helps to tackle a long-standing problem facing the Navajo Nation—how to spread telecommunications and information technology across the Navajo Nation.[12]

Spectrum-based technologies remove the need to lay wires to reach individual Navajo households on the Nation. Some have described this as "technological leapfrogging."[13] As Navajos incorporate wireless technologies into their lives, an opportunity arises for the Nation to consider the broader question of how to shape the use of these telecommunications and information services in a way that helps the Nation develop economically and still preserve its culture. One of the first comprehensive studies of contemporary communications technology in Indian Country made the following important point:

State_Link/IAD/subsai03.pdf (providing statistics regarding telephone subscribership on the Navajo Nation).

6. *Id.* at tbl.3.

7. *Id.* at tbl.1.

8. *Id.* at tbl.3.

9. John C. Miller & Christopher P. Guzelian, *A Spectrum Revolution: Deploying Ultrawideband Technology on Native American Lands*, 11 COMMLAW CONSPECTUS 277, 278 (2003).

10. *See* Cullen, *supra* note 1, at 1.

11. Spectrum-based technologies range from radio to cell phones; I will use the term to cover any communications technology that uses the electromagnetic spectrum, rather than wires, to transmit signals.

12. The terms "telecommunications" and "information services" help differentiate how the FCC regulates these different services under the 1996 Telecommunications Act. Generally, "telecommunications" refers to basic telephone services, while "information services" are more advanced services like broadband Internet services. *See* NUECHTERLEIN & WEISER, *infra* note 21, at 152. The significance of the difference between these two types of service will be discussed below and will help shape several of the arguments made within this paper. *See infra* Part II.

13. *See* Kade L. Twist, *Native Networking Trends: Wireless Broadband Networks*, BENTON FOUNDATION, Sept. 21, 2001, http://www.digitaldivide.net/articles/view.php?ArticleID=280 (discussing the ideas of technological leapfrogging).

> Community and cultural development is perhaps the development area most commonly considered for tribal communications. Many projects within Indian Country fit under this label. Noncommercial radio and television stations have been part of the tribal landscape for years, and have as their main purpose the provision of culturally appropriate services to Indian communities. Other examples include Internet education and access projects, satellite services, and public safety radio networks.[14]

This paper examines the cultural and legal issues surrounding the Navajo Nation's efforts to provide and regulate advanced telecommunications services ("ATSs"), specifically high-speed wireless Internet access, on tribal lands. Part I connects the Navajo's history of technological incorporation, such as weaving and silversmithing, with the tribe's growing use of Internet technologies. The discussion focuses on the Navajo's most recent plans to develop its own high-speed wireless Internet network called "Internet to the Hogan." Part II provides an overview of federal telecommunications law and explores how the Federal Communications Commission historically regulated the telecommunications and electronic information industry. The role that the states have taken in this field will be reviewed. The discussion ends with an examination of the specific statutory language that covers ATSs. Part III then provides an analysis of how tribal sovereignty, federal regulation, and state jurisdiction issues should be viewed to help provide the tribe with the strongest arguments for asserting control over ATSs on tribal lands. Part IV concludes this paper with a look to the future of the Navajo Internet network, making several recommendations for policy and statutory changes to further empower Navajo cultural and legal control over its Internet infrastructure.

I. THE INTERNET AND NAVAJO CULTURAL ADAPTABILITY

> We used silver that came from outside and learned to silversmith and became world-renowned for that. We took the technology of weaving and made it our own and we've become world-renowned. Internet technology is a tool that has come from outside. Now we have to take that and become world-renowned.[15]

Navajo culture is well known for its ability to adapt and to

14. JAMES CASEY ET AL., NATIVE NETWORKING: TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY IN INDIAN COUNTRY 15 (Jean Smith ed., 1999), http://www.benton.org/publibrary/native/bentonne.pdf.

15. Cullen, *supra* note 1, at 4 (quoting Ernest Franklin of the Navajo Nation Department of Capital Improvement).

incorporate new technologies that it encounters.[16] "The Navajos have always brought in new people, new ideas, and new elements and, over time, made them Navajo."[17] The Internet is no different. Just as the Navajos have developed their own style of weaving and silversmithing, "the Navajos w[ill] continue to demonstrate their capacity for learning initially from others and ultimately develop[] forms of cultural expression that, regardless of their derivation, w[ill] emerge as fundamentally, centrally Navajo."[18]

Over the past ten years, the Navajo Nation has begun the process of weaving the Internet into Navajo life. Starting in 1998, grants from the Bill and Melinda Gates Foundation's Native American Access to Technology Program have helped to launch a program that provides funding to install computers in all 110 Navajo chapter houses and to connect each chapter house to the Internet.[19] Because of the rugged landscape and remoteness of many of the chapter houses, the primary ways that most Americans obtain access to the Internet, either through their phone line or through cable, are not practical for most Navajos. Rather, the Navajos turn to the skies above and receive their connection to the Internet using a geosynchronous satellite orbiting thousands of miles above the Navajo Nation.[20]

The use of satellite technology to connect to the Internet shows one way that the Navajos are tapping into the rapidly evolving world of telecommunications technology. Commentators call it technological convergence and digital migration.[21] Where once communications were sent via analog signals, whether over wires or airwaves, today communications are usually sent via digital signals. While the significance of this analog to digital switch may not be obvious, the key is that digital technology reduces the world of communications to a binary system, where messages are translated into a coded series of 0s and 1s.[22] As a result, it does not matter if one is trying to connect to the Internet using a phone line, fiber optic cable, radio tower, or satellite. All this information, in the end, is reduced to the same basic building blocks of

16. *See* PETER IVERSON, DINÉ: A HISTORY OF THE NAVAJOS 3 (2002).

17. *Id.*

18. *Id.* at 32.

19. Andrew C. Gordon et al., *Native American technology access: the Gates Foundation in Four Corners*, 21 ELEC. LIBR. 428, 431-32 (2003); Navajos – Overview, History, Modern Era, The first Navajos in America, Settlement, Acculturation, and Assimilations, http://www.everyculture.com/multi/Le-Pa/Navajos.html ("The basic unit of local government in the Navajo Nation is the Chapter, each with its own Chapter House. The Chapter system was created in 1922 as a means of addressing agricultural problems at a local level").

20. Gordon et al., *supra* note 19, at 431-32

21. *See* JONATHAN E. NUECHTERLEIN & PHILIP J. WEISER, DIGITAL CROSSROADS: AMERICAN TELECOMMUNICATIONS POLICY IN THE INTERNET AGE 23-27 (2005).

22. *See id.* at 116.

0s and 1s.

The significance of this digital revolution is that many forms of technology can more easily switch from one means of reaching individuals to another, and one device can handle many different types of applications. This is known as technological convergence.[23] In the past, different technologies required different devices, whether it be televisions, phones, or computers. Now, technological convergence and digital technology allows all these technologies to come together as one. Therefore, the satellite Internet connection to each Navajo chapter house can allow for a wide variety of technologies through one device.

The relevance of this technological convergence as it pertains to tribal and cultural sovereignty is that the evolving state of telecommunications technology provides different options for tribes to consider when developing telecommunications infrastructure and how to regulate it. Although the Navajo Nation may have lagged behind the rest of the nation with respect to basic telecommunications services through the year 2000, technological convergence now allows the Nation to make a technological leapfrog forward.[24] The Navajo Nation faces critical decisions about where it wants to land in this new technological world. As will be discussed in subsequent sections of this paper, the state of telecommunications regulation in America has established several avenues by which tribes can seek to regulate these services on tribal lands with an eye towards establishing greater cultural sovereignty and self-governance.[25]

The Navajo Nation has already started making these choices by the ways that it utilizes the Internet and makes the Internet its own. One of the original motivations for connecting each chapter house to the Internet was so that the Nation could help satisfy "the Local Governance Act (LGA), which had been passed in 1998 to encourage the Navajo Nation's 110 chapters to become more self-sufficient."[26] Each chapter house has its own website through which community planning and development information is provided.[27] As a result, over fifty local Navajo communities have developed land-use plans based on information gathered by locals through the Internet that have

---

23. *Id.* at 23.

24. *See* Electa Draper, *Tech Tidal Wave Heads Toward Tribe*, DENVERPOST.COM, Mar. 1, 2007, http://www.denverpost.com/news/ci_5326782 (discussing Navajo technological leapfrogging).

25. *See infra*.

26. *See* Cullen, *supra* note 1, at 1-2; *see generally*, NAVAJO NATION CODE tit. 26, § 1B (1998).

27. Marty Logan, *Vast Navajo Nation Connects Communities via Web*, INTER PRESS SERVICE, Dec. 6, 2005, *available at* http://www.finalcall.com/artman/publish/article_2314.shtml.

"empower[ed] individual communities to take the lead in rebuilding their lives."[28] These decentralized efforts at community empowerment go to the heart of Navajo culture: "the Diné are far from being a unified nation but rather constitute a series of autonomous groups with highly localized leadership patterns."[29] At the same time, as Navajo Nation President Joe Shirley, Jr. said more broadly, "The most important thing is that our people are communicating with their government, and their government is communicating with them."[30]

Navajo chapter houses are using their satellite connection to the Internet as a cultural preservation and adaptation tool on several other levels. Along with providing a means for better government services, the Internet also helps bridge generations.[31] Young Navajos drive much of the interest in integrating the Internet into Navajo society.[32] "Some chapters have created programs in which youth teach elders to use computers that have been installed in senior centers. In return, the elders teach the youth traditional Navajo stories."[33] Furthermore, the Navajo Nation is providing Internet access at all the Navajo's Head Start schools.[34] As a tool, the Internet allows teachers to "use online resources to plan their classroom curricula and find out about changing federal [education] requirements."[35] Navajo President Shirley commented, "Education, as we see it, as an indigenous nation, is one of the ways to get back to standing on our own two feet, sustaining ourselves as individuals, as families, as communities, and, ultimately, as a nation."[36]

More broadly, the Internet helps protect Navajo culture through the Internet's other educational and economic aspects. Older Navajos are taking online distance education classes and earning both undergraduate and graduate degrees.[37] Navajos are also participating in eCommerce; over 600 Navajo artists presently sell their works, including jewelry, rugs, and pottery, online.[38] Navajo President Shirley is now talking with other

28. Marty Logan, *Navajos Go Global: An Indigenous Web Builds Up*, INTER PRESS SERVICE, Nov. 17, 2005, http://www.ipsterraviva.net/TV/Tunis/viewstory.asp?idnews=383.

29. IVERSON, *supra* note 16, at 25.

30. Kathy Helms, *Navajo president attends World Summit in Africa*, GALLUPINDEPENDENT.COM, Nov. 16, 2005, http://www.gallupindependent.com/2005/nov/111605wsmt.html.

31. Cullen, *supra* note 1, at 3.

32. *Id.*

33. *Id.*

34. *Id.*

35. *Id.*

36. Press Release, Navajo Nation, Navajo President Joe Shirley, Jr., addresses international UN conference on information, communication, technology (Apr. 24, 2007), *available at* http://opvp.org/cms/kunde/rts/opvporg/docs/828919834-04-26-2007-15-20-49.pdf.

37. *See* Shirley Press Release 1, *supra* note 4, at 2.

38. *Id.*

less developed nations about the "Navajo Model" for utilizing Internet technology: "[t]he Navajo Nation has demonstrated to the world that a people who value culture, language, and tradition can use satellite and wireless technology to help maintain their way of life."[39]

The next step on this path of technological transformation involves spreading the "Internet to the Hogan."[40] Leaders of the transformation plan "to leapfrog the Navajo Nation ahead of what's available in the finest homes and communities in New York City or Denver."[41] The Internet to the Hogan Plan involves establishing a state of the art wireless broadband network that is owned and operated by the Navajos rather than relying on a non-Navajo Internet Service Provider ("ISP") as the Nation has done to date.[42] Furthermore, instead of utilizing satellite communications, the Internet to the Hogan network will rely on terrestrial wireless technology.[43] This will allow the Navajo network to become true high-speed broadband, as the satellite technology used to date is more comparable to slow narrow-band dial-up access to the network.[44] The new network will allow Navajos to achieve Internet speeds 10 to 100 times faster than what they have experienced so far.[45] Furthermore, Navajo chapter houses and Navajo colleges will have portable mini-supercomputers onsite to help build the backbone of this network.[46]

As the Navajo Nation establishes its Internet to the Hogan network, the Nation will use wireless technology to allow individual Navajo homes to go online through wireless connections to communication towers, similar to the technology used for cell phones. According to the Internet to the Hogan Plan, "[t]he Navajo Nation culture and values will help shape the strategic plan[] and provide guidance in developing [and] implementing modern information

---

39. *Id.*

40. *See* DEP'T OF INFORMATION TECH., NAVAJO NATION INFORMATION TECH. PLAN 2007 – 2008: "INTERNET TO THE HOGAN – ENTERING THE INFORMATION AGE" (Jan. 2, 2007), http://www.dit.navajo.org/pdf/NAVAJO%20NATION%20Strategic%20Information%20Technology%20Plan%205.0.pdf.pdf [hereinafter INTERNET TO THE HOGAN TECH. PLAN].

41. Draper, *supra* note 24, at 1 (quoting Tom Davis, Dean of Instruction at Navajo Technical College).

42. OnSat Network Communications provides satellite Internet to the Navajo chapter houses. *See* Gordon et al., *supra* note 19, at 431. A recent audit, however, suggests that OnSat inappropriately billed the Navajos over $650,000. *See* John Christian Hopkins, *Bates: 'Gross negligence' in OnSat dealings*, GALLUPINDEPENDENT.COM, June 21, 2007, http://www.gallupindependent.com/2007/june/062107jch_onsatdealings.html.

43. *See* Draper, *supra* note 24, at 1.

44. *See* Digital Equity Network, http://www.nnden.org/ (providing information about OnSat, the satellite Internet Service Provider to the Navajos).

45. *See id.*

46. *See* Draper, *supra* note 24, at 2.

technology within the Navajo Nation['s] boundaries and beyond."[47] "From Chapter Houses to the central government Navajos still integrate traditional prayers in hopes of acquiring the best for [the Navajo] people. Values generated by [Navajo] way of life will be the driving mechanism in this plan; integrat[ion] with current technology will also drive the implementation of [information technology]."[48]

At its root, the Internet to the Hogan Plan will allow the Navajos to control their connection to the Internet. Navajo President Shirley has talked about the importance of such indigenous Internet portals that connect peoples like the Navajo to the rest of the world: "Our portal will allow us to share, with our own voices, our traditions, values, history and language as well as our aspirations for the future."[49] The creation of indigenous Internet portals like the Navajo Network will help "stave off complete cultural assimilation . . . [as] indigenous peoples are embracing the technology of the digital age to ensure their continued survival."[50] Nonetheless, there are several important jurisdictional issues that may impact the Navajo Nation's power to control their own wireless network and its connection to the global Internet. The following sections of this paper examine how the different jurisdictional arguments that the federal, state, and tribal governments may assert could affect the Internet to the Hogan Plan.

II. ADVANCED TELECOMMUNICATIONS SERVICES: UNDERSTANDING FEDERAL TELECOMMUNICATIONS AND INFORMATION LAW IN THE NAVAJO CONTEXT

> The Navajo way of life is a process of moving from a domain of perfect beauty into history, the threat of chaos motivates a return to ritual in order to achieve re-creation and renewal. Navajo life can be portrayed as a pathway out of the domain of the perfect beauty of fresh creation and into history, into the profane world.[51]

Understanding the legal issues that may arise when the Navajo Nation provides and regulates Internet services requires delving into the complex world of federal communications law. Several broad conceptual

---

47. INTERNET TO THE HOGAN TECH. PLAN, *supra* note 40, at 9-10.

48. *Id.* at 5.

49. Joe Shirley, Jr., President, Navajo Nation, Address to the World Summit on the Information Society: Towards an Indigenous Portal (Nov. 23, 2005), http://www.itu.int/wsis/tunis/statements/docs/pe-indigenous/1.doc.

50. Angela R. Riley, *"Straight Stealing": Towards an Indigenous System of Cultural Property Protection*, 80 WASH. L. REV. 69, 113 (2005).

51. SAM D. GILL, NATIVE AMERICAN RELIGIONS: AN INTRODUCTION 24-25 (2nd ed. 2005).

ideas will help clarify the legal and jurisdictional issues pertinent to this paper. First, the Communications Act of 1934 ("the 1934 Act"), provides the basis for almost all regulation of communications over wire or air.[52] Federal and state regulation which arose under the 1934 Act compartmentalized communications under different statutory titles.[53] Technological convergence and digital migration are now causing problems with this compartmentalization. Fortunately, the Telecommunications Act of 1996 ("the 1996 Act"),[54] which amended the 1934 Act, provides some regulatory flexibility that is important for the analysis of the provision and regulation of wireless broadband services by the Navajo through the Internet to the Hogan Plan.

A. The Communications Act of 1934

Federal legislation regarding telecommunications started during the early part of the twentieth century, in part as a result of the tragic sinking of the Titanic and also the need to create a functional national telephone system. During the sinking of the Titanic, distress signals were sent out over radios, but the lack of a coordinated, unified system for using radios led to calls for help going unheeded.[55] Around the same time, local telephone companies were starting up around the country. However, because again no coordination amongst telephone companies existed, connecting users of different phone services often proved impossible.[56] By 1934, Congress passed the 1934 Act, providing broad statutory authority for a new federal agency, the Federal Communications Commission ("FCC"), to oversee telecommunications development across the country.[57] While the FCC's power to regulate telecommunications preempted the states from regulating interstate communications, state public utility commissions could still regulate intrastate communications.[58]

52. Communications Act of 1934, Pub. L. No. 73-416, 48 Stat. 1064 (codified as amended in scattered sections of 47 U.S.C.).

53. *Id*.

54. Telecommunications Act of 1996, Pub. L. No. 104-104, 110 Stat. 56 (codified as amended in scattered sections of 47 U.S.C.) [hereinafter The 1996 Act].

55. *See* NUECHTERLEIN & WEISER, *supra* note 21, at 232.

56. *See id.* at 5-6.

57. *See* STUART MINOR BENJAMIN ET AL., TELECOMMUNICATIONS LAW AND POLICY Ch. 2 (Carolina Academic Press 2006) (2001) (providing an overview of the FCC and its regulatory powers).

58. *See* NUECHTERLEIN & WEISER, *supra* note 21, at 47-49 (discussing dual jurisdiction between federal and state governments with respect to traditional telephone service).

i. The Multi-Silo Approach to Telecommunications Regulation

Depending on the nature of the service, whether over the air (like radio broadcasts) or over wires (like traditional telephone service), both the FCC and state public utility commissions apply regulations under several statutory titles of the 1934 Act, as amended. For the purposes of this paper, Titles I, II, and III provide the key provisions shaping the discussion. Title I ancillary jurisdiction is the most significant to this paper as it is a catch-all title allowing for regulation of communication services and technologies that do not fit neatly into any of the other categories yet may impact the services provided under the other titles.[59] Title II provides for regulation of traditional common carrier services over wires, such as basic telephone service. Common carrier regulation typically involves industries that have natural monopolies, and thus the regulation imposes upon phone companies the obligation to ensure equality in how they provide services.[60] Title II's importance to this discussion relates to the division between federal and state regulation of common carriers, creating dual jurisdiction issues; the FCC regulates interstate telecommunications, while the states regulate intrastate telecommunications.[61] Under Title III, the FCC regulates the use of the airwaves, or electromagnetic spectrum. Historically, this involved regulating radio and television broadcasters. Today, however, with the rise of more wireless technology, whether wi-fi hubs or cell phones, Title III covers a broader category of services than it originally did.[62] Typically, Title III federal regulation completely preempts state regulation. Several other titles fill out the FCC's regulatory authority: Cable Enforcement (Title IV), Penalties (Title V), and Miscellaneous Provisions (Title VII).[63]

Because FCC regulation is divided into several different titles, the rise of digital technology and technological convergence has caused numerous regulatory headaches. A new telecommunications technology often fits under several different titles. The following sections of this paper will present specific regulatory aspects affecting different communication services and technologies.[64] These sections will provide a

---

59. *See id.* at 218 (providing an overview of Title I ancillary jurisdiction). Section 4(i) gives the FCC the power to "perform any and all acts, make such rules and regulations, and issue such orders, not inconsistent with this [Act] as may be necessary in the execution of its functions." 47 U.S.C. § 154(i) (2000).

60. *See* NUECHTERLEIN & WEISER, *supra* note 21, at 23.

61. *Id.* at 47-48.

62. *See id.* at 24.

63. *See* BENJAMIN ET AL., *supra* note 57, at 54-55.

64. *See infra*.

broader discussion regarding how tribal sovereignty over advanced telecommunications services, such as the broadband wireless network of the Internet to the Hogan Plan, will allow the Navajo Nation to determine how it wants to provide and regulate Internet infrastructure on tribal lands. Each of the following sections examines the key elements of federal regulation but also provides a closer look at where states play a role in regulating some of these telecommunications services. This provides a basic framework for further analysis when considering how tribes like the Navajo might take on more roles that bear resemblance to telecommunications regulation by federal or state regulators.

ii. Common Carrier Regulation under Title II

Under Title II, the FCC regulates interstate common carrier services that are provided over wires.[65] Long distance telephone service is the best known interstate common carrier service.[66] However, states regulate intrastate telephone service, typically through state public utility commissions.[67] On both the state and federal levels, regulators play an important role in approving the rates that are charged for phone services.[68] In addition, regulators also ensure that telephone service is provided in a non-discriminatory way.[69] As discussed in later sections, both the federal and state governments have established Universal Service Funds, which are additional charges tacked onto phone bills.[70] These help to subsidize underserved communities, either because of income disparity or difficulty in providing service (i.e., remote communities).[71] The important take-away is that both the federal government and state governments play a role in regulating phone service because of its separate, yet interconnected, interstate and intrastate aspects.

Although the Navajo's Internet to the Hogan Plan involves using wireless technology to create a communications network for the Navajo Nation, it does have aspects that are reminiscent of common-carriers. The network will provide access to a potentially essential service, which is similar to many common carriers like traditional phone. In addition, the Internet to the Hogan Plan has similarities to intrastate phone services as it provides service within the reservation boundary. If the Internet to the Hogan Plan becomes the only real means for Navajos to connect to the

65. *See* The 1996 Act, *supra* note 54.
66. *See* NUECHTERLEIN & WEISER, *supra* note 21, at 48.
67. *Id.*
68. *See id.* at 214.
69. *Id.*
70. *Id.* at 52.
71. *Id.*

broader communications world, the resemblances to common carriers may suggest that the Internet to the Hogan Plan be regulated as a common carrier under Title II.

iii. Spectrum Regulation under Title III

Because electromagnetic radio signals, such as AM and FM, can easily cross state boundaries, the federal government has taken exclusive jurisdiction over spectrum regulation since 1927 when Congress passed the precursor to the 1934 Communications Act.[72] Two essential principles provide guidance to spectrum regulation. First, spectrum regulation at its root is about avoidance of interference between different users.[73] As a result, much of the past eight decades of spectrum regulation placed the FCC in a command-and-control position, where the FCC dictated who gets to use which slices of spectrum for what purposes and where.[74] Nonetheless, since the passage of the 1996 Telecommunications Act, a slow shift has begun towards deregulation that chips away at the FCC's historic command-and-control of spectrum.[75]

The other essential principle is that spectrum is owned by the public.[76] Because the "airwaves" are seen as being owned and shared by the public, Congress wrote language into the original Communications Act that imposed public trust obligations on those who used spectrum to broadcast to the public; Congress called on the FCC to regulate spectrum "from time to time, as public convenience, interest, or necessity requires."[77] When a radio or television station needs to use spectrum to transmit its signal, it has to go to the FCC and get a license.[78] Historically, these licenses were, in effect, given away. More recently, spectrum has been auctioned off, most notably to cell phone companies.[79]

Because the Internet to the Hogan Plan will rely on wireless technology and spectrum, Title III regulation has the most significance to the future of the Navajo network. On some level, the Navajo Nation will have to work with the FCC to coordinate spectrum usage. In Part III, this paper discusses in more detail the jurisdictional issues that could impact this issue.

72. Radio Act of 1927, 44 Stat. 1162 (1927); NUECHTERLEIN & WEISER, *supra* note 21, at 232.

73. *See* NUECHTERLEIN & WEISER, *supra* note 21, at 229.

74. *See* BENJAMIN ET AL., *supra* note 57, at 67.

75. *See* NUECHTERLEIN & WEISER, *supra* note 21, at 239.

76. *See id.* at 229.

77. 47 U.S.C § 303 (2000).

78. *See* NUECHTERLEIN & WEISER, *supra* note 21, at 236.

79. *Id.* at 237.

iv. Internet Regulation under Title I

Generally, Congress and the FCC have avoided regulating the Internet. In the 1996 Telecommunications Act, Congress added a statement suggesting its Internet policy: "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."[80] However, services that rely on the Internet have faced scrutiny to determine if regulation is necessary. Two active areas caught both Congressional and FCC Commissioners' attention over the past few years (1) Internet Service Provider ("ISP") discrimination towards content providers and (2) Internet telephone services, like Voice-over-Internet-Protocol ("VoIP").[81] Under the FCC's Title I ancillary authority, the FCC can take steps to regulate telecommunications industries that may impact the other spheres that the FCC more formally regulates.[82] VoIP is of significance to this paper as this Internet-based telephone service has drawn the attention of state regulators.[83] Furthermore, the Navajo Nation has taken active steps to begin using VoIP to provide phone service for its government agencies.[84] For the purposes of this paper, what is most relevant is that Section 706,[85] discussed in the following section, specifically contemplates both the FCC and state utility commissions using different regulatory approaches to help ensure that all Americans have access to high-speed broadband Internet connections, regardless of the specific applications that might be used.

80. 47 U.S.C. § 230(b)(2) (2000); *see also* NUECHTERLEIN & WEISER, *supra* note 21, at 197.

81. *See* NUECHTERLEIN & WEISER, *supra* note 21, at 155, 197 (discussing network neutrality and VoIP regulation, respectively).

82. *See id.* at 218; *see also* United States v. Sw. Cable, 392 U.S. 157, 178 (1968) (upholding FCC ancillary authority to regulate industries like cable that may impact broadcasters).

83. However, the FCC has determined that VoIP is "indivisibly interstate," thus preempting state regulation. NUECHTERLEIN & WEISER, *supra* note 21, at 205.

84. *See* INTERNET TO THE HOGAN TECH. PLAN, *supra* note 40, at 70.

85. 47 U.S.C.A. § 1302 (2008) (codifying Pub. L. No. 110-385, 122 Stat. 4096 (2008)) [hereinafter Section 706].

B. *The 1996 Act and Section 706: Advanced Telecommunications Services*

In 1996, Congress amended the 1934 Act to account for the many changes that had occurred in the ways we communicate.[86] The Telecommunications Act of 1996 focuses on deregulating the telecommunications industry by having the FCC promote competition within the markets.[87] In addition, the 1996 Act contains new provisions to account for emerging technologies like the Internet. Section 706 specifically provides the FCC and state public utilities commissions with broad discretionary powers as to how these commissions can either regulate or deregulate Internet infrastructure.[88] The language of Section 706 focuses on what are called "advanced telecommunications capabilities."[89] Basically, it refers to high-speed broadband service, which is the form of telecommunications and information service within the Navajo Nation's Internet to the Hogan technology plan.[90] The goal of Section 706 is to see the timely deployment of broadband Internet service to all Americans, and the section includes a requirement that the FCC make regular inquiries into the deployment of this advanced telecommunications service.[91] The specific language of Section 706 is informative:

> The Commission and each State commission with regulatory jurisdiction over telecommunications services shall encourage the deployment on a reasonable and timely basis of advanced telecommunications capability to all Americans (including, in particular, elementary and secondary schools and classrooms) by utilizing, in a manner consistent with the public interest, convenience, and necessity, price-cap regulation, regulatory forbearance, measures that promote competition in the local telecommunications market, or other regulating methods that remove

---

86. *See* The 1996 Act, *supra* note 54.

87. *See* NUECHTERLEIN & WEISER, *supra* note 21, at ch.3.

88. *See* BENJAMIN ET AL., *supra* note 57, at 955.

89. Section 706, *supra* note **Error! Bookmark not defined.**, § 1302(d) ("The term 'advanced telecommunications capability' is defined, without regard to any transmission media or technology, as high-speed, switched, broadband telecommunications capability that enables users to originate and receive high-quality voice, data, graphics, and video telecommunications using any technology.").

90. *See* INTERNET TO THE HOGAN TECH. PLAN, *supra* note 40, at 13-14.

91. *Id.* For the remainder of this paper, I will use the terms "advanced telecommunications capabilities," "advanced telecommunications services," and "ATSs" interchangeably.

barriers to infrastructure investment.[92]

In effect, Section 706 opens up the regulatory and deregulatory toolboxes that the FCC and the state commissions have used in the past and anticipated needing in the future in order to ensure that broadband Internet access (along with other advanced technologies) reach Americans in a reasonable and timely fashion as called for in the 1996 Act.

The FCC has applied many of these different approaches over the past ten years. Most notably, the FCC has moved towards minimizing regulation of advanced telecommunications services ("ATSs"). For example, the FCC categorized both cable modems and telephone digital subscriber lines ("DSL"), the two prominent ways that many Americans achieve access to high-speed broadband Internet service, as "information services," rather than as offerings of "telecommunications services."[93] This distinction brought these ATSs outside of the sphere of Title II common carrier regulation and put them under the more deregulatory approach of Title I ancillary jurisdiction. While wireless broadband ATSs—like those the Navajo's Internet to the Hogan Plan anticipate—have yet to be categorized, the general trend of FCC regulation would suggest the FCC will not treat wireless broadband providers as common carriers.

However, the specific provisions of Section 706 do give the FCC and state commissions the power to impose regulation in order to help ensure that all Americans have access to broadband Internet services.[94] Thus, if regulation is needed to help ensure the viability of the Internet to the Hogan Plan, Section 706 provides a means to achieve these regulations. As discussed in the next two Parts of this paper, the Navajo Nation should play a role in determining what type of regulation will be helpful for the Internet to the Hogan Plan.

The states also play a role, though more limited by the jurisdictional restriction to intrastate matters, in regulating ATSs. Most important for this discussion, some state legislatures have passed laws that prohibit local municipalities from providing wireless broadband services.[95] This

---

92. Section 706, *supra* note **Error! Bookmark not defined.**, § 1302(a).

93. *See* Nat'l Cable & Telecomm. Ass'n v. Brand X Internet Serv., 545 U.S. 967 (2005) (giving *Chevron* deference to the FCC's categorization of cable modem service as an information service rather than an offering of telecommunications services); Appropriate Framework for Broadband Access to the Internet over Wireline Facilities, *Notice of Proposed Rule Making*, 17 FCC Rcd. 3019 (2002) (developing rules to treat digital subscriber line broadband as an information service).

94. *See* Section 706, *supra* note **Error! Bookmark not defined.**, § 1302(a); *see also* BENJAMIN ET AL., *supra* note 57, at 955.

95. *See* Nixon v. Mo. Mun. League, 541 U.S. 125 (2004) (affirming a state's power to

could potentially impact the Navajo Internet to the Hogan Plan if a state, such as Utah, were to try to assert this authority over the Navajo wireless broadband network; this issue is further analyzed in the following section.[96]

III. INTERNET TO THE HOGAN: TRIBAL SOVEREIGNTY, FEDERAL TRUST, AND STATE JURISDICTION

> [Jurisdiction] gives [tribes] the governmental power required to operate cultural and economic programs on which to build a future for the people living on the reservation. Indians see this ability to make and enforce laws in particular territory as an essential force necessary to preserve a geographic and cultural core, and to perpetuate their survival as tribal peoples.[97]

> The extent to which federal and state regulatory authority over telecommunications services in Indian Country exists has never been universally defined. In most cases, for example, because the tribes have not exercised their authority to regulate telecommunications services within reservation boundaries, the state regulatory agencies have exercised jurisdiction over telecommunications services within Indian Country by default.[98]

As the tension between these two quotes suggests, tribes have an important role in asserting jurisdiction over activities that impact a tribe's culture. Yet within the arena of common carrier telecommunications, tribes have generally not asserted such jurisdiction and, in some cases, states have stepped in to fill the void.[99] At the same time, however, some states have gone back and forth as to whether they have jurisdiction over telecommunications on tribal lands.[100] In the area of spectrum regulation, the FCC has generally regulated Title III telecommunications services

---

prohibit local municipalities from providing broadband Internet services); *see also* FTC, MUNICIPAL PROVISION OF WIRELESS INTERNET (FTC Staff Report, Sept. 2006) (providing an analysis of the pros and cons of municipal wireless), http://www.ftc.gov/os/2006/10/V060021municipalprovwirelessinternet.pdf [hereinafter MUNICIPAL PROVISION REPORT].

96. *See* Cybertelecom, Municipal Broadband, http://cybertelecom.org/broadband/muni.htm (providing a list of states that prohibit or restrict municipal broadband, including the state of Utah) [hereinafter Municipal Broadband].

97. DAVID H. GETCHES ET AL., CASES AND MATERIALS ON FEDERAL INDIAN LAW 456-57 (5th ed. 2005).

98. CASEY ET AL., *supra* note 14, at 15.

99. *Id.*

100. *Id.* (pointing out that both California and Arizona have at different times stated they do not have jurisdiction over telecommunications services on tribal lands).

on tribal lands since the passage of the Communications Act of 1934.[101] However, the FCC did little to promote and support tribal use of spectrum until recently.[102] As the remainder of this paper argues, an important shift occurred over the past decade in which the Navajo began to assert more control and jurisdiction over telecommunications services on tribal lands.[103]

This section examines more closely the principles of tribal sovereignty and the federal trust relationship as they impact tribal jurisdiction and regulation of ATSs on tribal lands. In addition, this section looks at issues that may arise if states assert jurisdiction in the telecommunications arena. This section provides the foundation for the conclusion where this paper argues that the FCC should promote the Navajo Nation's desire to require that the FCC and the states "[r]ecognize the rights of tribal governments and communities to exercise regulatory jurisdiction over telecommunication activities within the boundaries of reservations."[104]

Despite the apparent de facto FCC and state jurisdiction over telecommunications service to native nations, some commentators argue that the questions of jurisdiction still remain unresolved therefore leaving ambiguity over the authority the FCC has to regulate telecommunications on tribal lands.[105] This section addresses three specific questions: (A) how does tribal sovereignty affect the power of the FCC to regulate on tribal lands; (B) what role does the federal trust relationship play in defining the FCC's relationship with tribes; and (C) what role do states play in regulating telecommunications services on tribal lands. Through examining these three questions, this paper will develop a vision of tribal telecommunications regulation.

---

101. Miller & Guzelian, *supra* note 9, at 280.

102. *See, e.g.*, Extending Wireless Telecommunications Services To Tribal Lands, *Third Report & Order*, 19 FCC Rcd. 17,652 (2004); Extending Wireless Telecommunications Services To Tribal Lands, *Second Report & Order & Second Further Notice of Proposed Rulemaking*, 18 FCC Rcd. 4775 (2003); Extending Wireless Telecommunications Services To Tribal Lands, *Report & Order & Further Notice of Proposed Rule Making*, 15 FCC Rcd. 11,794 (2000).

103. *See* Navajo Nation Telecommunications Regulatory Commission, History, http://www.nntrc.org/content.asp?CustComKey=21364&CategoryKey=21413&pn=Page&DomName=nntrc.org.

104. PowerPoint presentation: Ron Lee, Navajo National Telecommunication Regulatory Commission, Ariz. Telecomm. & Information Council Pre-Summit, *Telecommunications Issues in Indian Country* 13 (Apr. 19, 2007) (on file with author).

105. *See* CASEY ET AL., *supra* note 14, at 15; *see also* Daniel J. Adam, Note, *Tribal Telecom: Telecommunications Regulation in Indian Country*, 27 J. LEGIS. 153 (2001) (providing a basic overview of telecommunications regulation in Indian country).

A. Tribal Sovereignty and Federal Telecommunications Jurisdiction

Tribal sovereignty goes to the heart of the Navajo Nation's power to self-govern. Within the tribe's inherent sovereignty exists cultural sovereignty, "the effort of Indian nations and Indian people to exercise their own norms and values in structuring their collective futures."[106] This section examines the different arguments the Navajo Nation could make in asserting exclusive jurisdiction over advanced telecommunications services on tribal lands. While exclusive jurisdiction might do the most to empower Navajo cultural sovereignty, it is a difficult case to argue. Instead, this paper suggests that the Navajo Nation should work with the federal government to establish a telecommunications regulation for a tribal status somewhere between that of a state and a foreign nation. This is the best position to take to avoid individual state control while also taking advantage of the FCC's telecommunications expertise and financial support programs.

The Communications Act of 1934, as amended, is best described as a federal statute of general applicability. While a Supreme Court decision held that "general Acts of Congress apply to Indians as well as to all others in the absence of a clear expression to the contrary,"[107] the 9th Circuit outlined the generally applied exceptions to this rule in *Donovan v. Coeur d'Alene Tribal Farm*:

> A federal statute of general applicability that is silent on the issue of applicability to Indians tribes will not apply to them if: (1) the law touches "exclusive rights of self-governance in purely intramural matters"; (2) the application of the law to the tribe would "abrogate rights guaranteed by Indian treaties"; or (3) there is proof "by legislative history or some other means that Congress intended [the law] not to apply to Indians on their reservations. . . ."[108]

Because the 1934 Act is a federal statute of general applicability that is silent with regards to Indians, it is appropriate to apply the *Donovan* exceptions to this analysis. One may first argue that the 1934 Act should not apply to the Navajo Nation because it impacts tribal self-governance in purely intramural matters. The Nation is developing its telecommunications infrastructure under the Internet to the Hogan Technology Plan in large part to implement the Nation's Local

---

106. Wallace Coffey & Rebecca Tsosie, *Rethinking the Tribal Sovereignty Doctrine: Cultural Sovereignty and the Collective Future of Indian Nations*, 12 STAN. L. & POL'Y REV. 191, 196 (2001).

107. Fed. Power Comm'n v. Tuscarora Indian Nation, 362 U.S. 99, 120 (1960).

108. 751 F.2d 1113, 1116 (9th Cir. 1985) (quoting United States v. Farris, 624 F.2d 890, 893-94 (9th Cir. 1980)).

Governance Act.[109] The use of telecommunications infrastructure further aids chapter house self-sufficiency.[110] Thus, telecommunications is being used as part of the Nation's self-governance in purely intra-mural matters. The Nation should therefore have exclusive jurisdiction over the telecommunications issues related to the Internet to the Hogan Technology plan.

Closely reading different sections of the 1934 Act, several commentators have further argued that tribal governments should have exclusive jurisdiction over telecommunications on tribal lands based in part upon factors similar to the *Donovan* exceptions.[111] For example, Section 152 of the 1934 Act reads broadly, applying to "all persons engaged within the United States in such communication or such transmission of energy by radio."[112] Section 301 goes on to refer to radio transmissions taking place in "any State, Territory, or possession of the United States."[113] Some commentators have used this language to argue that because cases "have held that reservations are not states, territories, or possessions of the United States pursuant to full faith and credit statutes, the National Labor Relations Act, and the Occupational Safety and Health Act," tribes should have exclusive jurisdiction over telecommunications on tribal lands.[114]

Because the Communications Act does not explicitly apply to tribal lands, arguments are made that any implicit intent to regulate telecommunications services on tribal lands is ambiguous. In *NLRB v. Pueblo of San Juan*, the 10th Circuit held that tribal economic regulation can preempt a federal statute of general applicability when federal intent to regulate the tribes is uncertain.[115] The Navajo's establishment of a regulatory commission to oversee telecommunications services,[116] along with the provisions and plans the Navajo have created to oversee the development of advanced telecommunications services,[117] contains enough similarities to tribal economic regulation that the reasoning of

109. *See supra* Part I.

110. *Id.*

111. *See* Miller & Guzelian, *supra* note 9, at 286-88.

112. 47 U.S.C. § 152 (2000); *see also* Miller & Guzelian, *supra* note 9, at 286-88 (providing a more detailed analysis of this argument).

113. 47 U.S.C. § 301 (2000).

114. Miller & Guzelian, *supra* note 9, at 286-87. However, the National Labor Relations Board in 2004 held that federal labor law does apply to tribes as employers, "overruling 30 years of its own precedent." GETCHES ET AL., *supra* note 97, at 339 (citing San Manuel Indian Bingo & Casino, 341 NLRB No. 138 (2004)).

115. Miller & Guzelian, *supra* note 9, at 293 (discussing NLRB v. Pueblo of San Juan, 276 F.3d 1186 (10th Cir. 2002)).

116. *See* Navajo Nation Telecommunications Regulatory Commission, http://www.nntrc.org.

117. *See generally* INTERNET TO THE HOGAN TECH. PLAN, *supra* note 40.

the Pueblo of San Juan decision may be applied.

Upon further examination of the Navajo Nation's Internet to the Hogan Plan, several points of the plan favor an exclusive tribal jurisdiction position. First, the initial drive to develop Internet infrastructure on Navajo land was the desire to help Navajo connect to their government, while maintaining their decentralized ways of living.[118] The Navajo Nation thus use the Internet for intra-cultural purposes. Furthermore, the Internet to the Hogan Plan creates an intra-network for the entire reservation.[119] While the network will allow access to Internet resources outside the Navajo Nation, the purpose of developing the broadband wireless network is essentially an internal one. One could reasonably argue that the way that the Navajo Nation has put the Internet to the Hogan Plan together fits *Donovan*'s first exception.

However, making these arguments may prove risky, especially because the wireless broadband plans for Navajo involve the use of spectrum rather than wireline connections. Using spectrum inevitably implicates the key reasons for federal exclusive jurisdiction. The airwaves know no boundaries; thus the Navajo's use of spectrum could potentially cause interference problems for other spectrum users outside of the reservation. In comparison, if the Navajo were using wire or cable infrastructure, like DSL over phone lines or cable modems, the Navajo would be able to argue that they were creating an exclusive intra-tribal infrastructure and could then regulate as do states with regard to intrastate phone services. However, this argument also runs into problems since the Internet itself seems "indivisibly interstate," a feature that has been used to preempt state regulation for Voice over Internet Phone service.[120]

Rather than seeking exclusive jurisdiction over advanced telecommunications services on their tribal lands, the Navajo Nation might find strategic advantages in working with the FCC to help establish and regulate the Internet to the Hogan broadband wireless network. Examining the potential trust relationship between the FCC and the Navajo Nation represents the next step in advancing this discussion.

118. *See Cullen, supra* note 1, at 1; *see also* INTERNET TO THE HOGAN TECH. PLAN, *supra* note 40, at 7.

119. *See* INTERNET TO THE HOGAN TECH. PLAN, *supra* note 40, at 41.

120. *See* NUECHTERLEIN & WEISER, *supra* note 21, at 205-06; *see also* Vonage Holding Corporation Petition for Declaratory Rule Concerning an Order of the Minnesota Public Utilities Commission, *Memorandum Opinion & Order*, 19 FCC Rcd. 22,404 (2004) (ruling that VOIP is indivisibly interstate), *aff'd in part* Minn. Pub. Utils. Comm'n v. FCC, 483 F.3d 570 (8th Cir. 2007).

B. *The Trust Relationship*

Although the Navajo Nation could attempt to assert complete jurisdiction over advanced telecommunications services on its tribal lands, a more fruitful approach may involve an examination of the federal trust relationship. Invoking the trust relationship does not necessarily go against the Nation's efforts to assert more cultural sovereignty and self-governance. Supreme Court jurisprudence with respect to the trust relationship focuses on the specific obligations of the federal government with respect to tribes based on the language of federal statutes. The *Mitchell* cases show the extremes, from a bare trust to a fiduciary trust.[121] The key consideration is often the degree of management a federal agency takes in overseeing and regulating a tribal resource or asset.[122]

Within the context of telecommunications law, there are several important points in this trust relationship analysis. First, when focusing on providing telephone services to underserved low-income and rural Americans under the USF provisions of the 1996 Act, the FCC interpreted the USF to include efforts to reach tribal communities who do not have adequate phone service.[123] Furthermore, the FCC effectively asserted control of spectrum available to tribal lands because of its exclusive jurisdiction over spectrum.[124] Thus, while the 1996 Act does not explicitly invoke the word "trust," the actions of the FCC show its intent to establish this relationship.

This last argument is bolstered by the fact that the FCC in 2000 established a specific government-to-government policy statement regarding its relationship to native nations like the Navajo.[125] To

121. United States v. Mitchell, 463 U.S. 206, 225 (1983) (finding a fiduciary relationship is established when a statute authorizes an agency to manage a tribal resource) [hereinafter *Mitchell II*]; United States v. Mitchell, 445 U.S. 535, 542 (1980) (holding that the General Allotment Act "created only a limited trust relationship").

122. *See Mitchell II*, 463 U.S. at 225; *see also* Seminole Nation v. United States, 316 U.S. 286 (1942) (discussing the fiduciary duty of the federal government regarding tribal trust funds).

123. *See* Federal-State Joint Board on Universal Service, *Twelfth Report & Order*, *Memorandum Opinion & Order, & Further Notice of Proposed Rulemaking*, 15 FCC Rcd. 12,208 ¶ 124 (Jun. 8, 2000) [hereinafter *Tribal ETC*].

124. *See supra* Part II, A, iii.

125. FCC, Statement of Policy on Establishing a Government-to-Government Relationship with Indian Tribes 4-5, FCC-00-207, June 23, 2000 (

> 1.The Commission will endeavor to work with Indian Tribes on a government-to-government basis consistent with the principles of Tribal self-governance to ensure, through its regulations and policy initiatives, and consistent with section 1 of the Communications Act of 1934, that Indian Tribes have adequate access to communications services.
>
> 2.The Commission, in accordance with the federal government's trust responsibility, and to the extent practicable, will consult with Tribal governments prior to implementing any regulatory action or policy that will significantly or

establish this policy, the FCC looked first to the 1996 Act, which mandates that "consumers in all regions of the Nation, including low-income consumers and those in rural, insular, and high[-]cost areas, should have access to telecommunications and information services."[126] The FCC realized that tribal lands fall distinctly within this category and affirmed its "commitment to promote a government-to-government relationship between the FCC and federally-recognized Indian Tribes."[127] This government-to-government relationship suggests several things. First, it provides a basis for treating tribal governments, such as the Navajo, who have telecommunications regulatory structures as if the tribal governments have inherent power to regulate in this area. This gives support for the idea that with respect to telecommunications regulation on tribal lands, the Navajo Nation should be treated on a level at least comparable to how the states are treated with respect to intrastate telecommunications regulation. At a minimum, the policy requires that tribes like the Navajo be brought into all proposed rule making that impacts their interests. This is discussed more in the last section of this

---

uniquely affect Tribal governments, their land and resources.

3.The Commission will strive to develop working relationships with Tribal governments, and will endeavor to identify innovative mechanism to facilitate Tribal consultation in agency regulatory processes that uniquely affect telecommunications compliance activities, radio spectrum policies, and other telecommunications service-related issues on Tribal lands.

4.The Commission will endeavor to streamline its administrative process and procedures to remove undue burdens that its decisions and actions place on Indian Tribes. As administrative and organizational impediments that limit the FCC's ability to work with Indian Tribes, consistent with this Policy Statement, are identified, the Commission will seek to remove those impediments to the extent authorized by law.

5.The Commission will assist Indian Tribes in complying with Federal communications statutes and regulations.

6.The Commission will seek to identify and establish procedures and mechanism to educate Commission staff about Tribal governments and Tribal culture, sovereignty rights, Indian law, and Tribal communications needs.

7.The Commission will work cooperatively with other Federal departments and agencies, Tribal, state and local governments to further the goals of this policy and to address communications problems, such as low penetration rates and poor quality of services on reservations, and other problems of mutual concern.

8.The Commission will welcome submissions from Tribal governments and other concerned parties as to other actions the Commission might take to further goals and principles presented herein.

9.The Commission will incorporate these Indian policy goals into its ongoing and long-term planning and management activities, including its policy proposals, management accountability system and ongoing policy development processes.

), *available at* http://www.fcc.gov/Bureaus/OGC/Orders/2000/fcc00207.doc [hereinafter FCC Tribal Policy].

126. *Id.* at n.2 (quoting 47 U.S.C. § 254(b)(3) (2006)).

127. *Id.* at 2-3.

paper.

C. *Dealing with State Jurisdiction Assertions*

State legislatures and their public utility commissions present potentially the greatest problem for tribes like the Navajo that try to assert control over advanced telecommunications services. Because federal communication laws like the 1934 and 1996 Acts allow state jurisdiction over telecommunications that pertain to intrastate matters, some states have taken this power to impose their state regulation on tribes.[128] As the quotes at the beginning of this Part suggests, these state public utility commissions often act to fill the regulatory void where tribes have not regulated telecommunications.[129]

At the same time, state public utility commissions have not taken the initiative to help tribes develop telecommunications infrastructure even if commissions do assert jurisdiction over tribal lands in this arena.[130] Part of the problem is that states have often gone back and forth as to their jurisdictional power over tribal telecommunications. Consider Arizona, for example, within which the largest percentage of Navajo lands is found.[131] For many years, Arizona's state public utility commission did not assert telecommunications jurisdiction over Indian-owned and operated telecommunications companies.[132] More recently, it started asserting jurisdiction when the telecommunications provider is non-tribally owned.[133] However, with the growing governmental structures of many tribes, most notably the Navajo,[134] the time has come to reconsider the whole question as to whether states should have any regulatory power over telecommunications services on tribal lands, regardless of who provides the telecommunications services.

Asserting tribal jurisdiction over advanced telecommunications services like the Navajo's Internet to the Hogan broadband plan is especially important today because of recent regulatory efforts on the part of several states. The Utah state legislature and numerous other states have passed laws banning local municipalities from providing wireless

128. CASEY ET AL., *supra* note 14, at 15.
129. *Id.*
130. *Id.*
131. Encarta, Navajo (people), http://encarta.msn.com/encyclopedia_761576887/Navajo_(people).html.
132. *See* CASEY ET AL., *supra* note 14, at 15; *see also* Federal State Joint Board on Universal Service, Order, 15 FCC Rcd. 8544, ¶ 9 (1999) (finding that Arizona state public utility commission does not have jurisdiction over tribally owned and operated telecommunications companies on tribal lands).
133. *See Tribal ETC*, *supra* note 123, ¶ 142.
134. *See, e.g.*, Navajo Nation Telecommunications Regulatory Commission, *supra* note 116.

Internet or other forms of broadband to their community members.[135] The Supreme Court held that such state laws are valid.[136] If a state like Utah, Arizona, or New Mexico were to try to assert this type of law over the Navajo Nation, it could have a devastating effect on the Internet to the Hogan project.

However, the Navajo Nation has several strong arguments it can make to counter any argument a state might use if the state were to try to assert jurisdiction over Navajo-provided telecommunications services. In *Nixon v. Missouri*, a local government which provided broadband internet to its community relied upon 47 U.S.C. § 253,[137] which states that: "No state or local statute or regulation, or other State or local legal requirement, may prohibit or have the effect of prohibiting the ability of any entity to provide any interstate or intrastate telecommunications service."[138] When the state of Missouri tried to shut down the local government broadband service, the U.S. Supreme Court upheld the right of states to restrict local municipalities from delivering telecommunications services explaining that:

> [W]hen a government regulates itself (or the subdivision through which it acts) there is no clear distinction between the regulator and the entity regulated. Legal limits on what may be done by the government itself (including its subdivisions) will often be indistinguishable from choices that express what the government wishes to do with the authority and resources it can command. That is why preempting state or local government self-regulation (or regulation of political inferiors) would work so differently from preempting regulation of private players that it is highly unlikely that Congress intended to set off on such uncertain adventures.[139]

In effect, the Court pointed out that because local governments receive their power from the state government, local governments cannot try to bypass that state authority using federal preemption, even if the federal government may have appeared to have preempted such power. Because tribal governments do not receive their power from the states, it would be a simple argument to make that states can thus not limit a tribal government from providing telecommunications services.

However, in Nixon the Court acknowledged that a local government is an "entity," yet argued, "that the strange and indeterminate results of using federal preemption to free public entities

135. *See* Municipal Broadband, *supra* note 96.
136. *See Nixon*, 541 U.S. at 131-40.
137. *Id.* at 130.
138. 47 U.S.C. § 253(a) (2000).
139. *Nixon*, 541 U.S. at 134.

from state or local limitations is the key to understanding that Congress used 'any entity' with a limited reference to any private entity when it cast the preemption net."[140] So, while the Court on one hand emphasized that local governments cannot bypass state law through a preemption analysis, in this case, the Court may have inadvertently limited Section 253 to private entities.

Because Section 253 may only apply to private entities, any analysis regarding tribal jurisdiction over telecommunications services in Indian Country should also examine the core Supreme Court rulings pertaining to state civil jurisdiction over Indian Country. The following analysis shows that none of the three states within which the Navajo Nation resides should have regulatory powers over the tribe's advanced telecommunications services in the Internet to the Hogan Plan.

In general, the Court applies a fact-specific, balancing-of-interest test when determining whether federal law preempts state jurisdiction in Indian Country.[141] "State jurisdiction is preempted by the operation of federal law if it interferes or is incompatible with federal and tribal interests reflected in federal law, unless the state interests at stake are sufficient to justify the assertion of state authority."[142]

Further Supreme Court decisions help fill out this understanding of how state jurisdiction may be precluded in Indian Country, including the "infringement" analysis from *Williams v. Lee*[143] and the "preemption" analysis from *McClanahan v. Arizona Tax Comm'n*.[144] Under all these tests, state regulation of Navajo owned and operated advanced telecommunications services should not be allowed. Furthermore, state regulation of non-tribally owned and operated advanced telecommunications should also be preempted based on the analysis in *Warren Trading Post v. Arizona Tax Comm'n*.[145] In addition, *California v. Cabazon Band of Mission Indians*[146] shows that states cannot prohibit the Navajo Nation from providing advanced telecommunications services if the state allows its municipalities to provide similar services.

---

140. *Id.* at 133.

141. New Mexico v. Mescalero Apache Tribe, 462 U.S. 324, 333-34 (1983).

142. *Id.* at 334.

143. 358 U.S. 217, 223 (1959) (providing for exclusive tribal jurisdiction when state jurisdiction would infringe on tribes' right to self-government).

144. 411 U.S. 164, 172 (1973) (precluding state regulation on reservation when federal law preempts state authority, based on a balancing of state, federal, and tribal interests).

145. Warren Trading Post v. Ariz. Tax Comm'n, 380 U.S. 685, 690 (1965) (ruling that states cannot regulate non-Indian traders in Indian Country because of federal preemption based on extensive federal regulation and tribal interests test).

146. California v. Cabazon Band of Mission Indians, 480 U.S. 202, 209 (1987) (holding that if a state permits certain forms of conduct, and doesn't absolutely ban such conduct, it does not have authority to enforce its regulation within an Indian reservation if the state regulation is incident upon the tribe or its members).

The Navajo Nation has its strongest arguments against state regulation when the activity regulated is performed by its own tribal members or the tribe itself.[147] Since the Internet to the Hogan Plan directly involves advanced telecommunications services that the Navajo Nation will provide, allowing state regulation would directly infringe upon the rights of the Navajo to make their own laws and be ruled by them, thus violating the principles in *Williams* and *McClanahan*.[148] The incidence of any such state regulation would fall directly on individual Navajos or the tribe itself.

The strongest argument that a state could make for regulatory authority in this case would be if the state was directly funding the Internet to the Hogan Plan through a funding mechanism like the state's Universal Service Fund. However, even in this case, the Navajo Nation would have a strong argument against state jurisdiction because funding alone does not create a regulatory nexus.[149] The Navajo Nation should nonetheless be cautious before accepting such funds from the state, and should turn to federal USF for support instead.

A stronger case for state regulatory authority arises when non-tribal telecommunications service providers are involved. Presently, the Navajo Nation does rely on a non-tribal telecommunications company to provide its satellite-based Internet service to chapter houses. In such cases, the preemption test in *Warren Trading Post*, *White Mountain Apache Tribe v. Bracker*, and *Cent. Mach. Co. v. Arizona Tax Comm'n* is appropriate.[150] These cases again involve balancing-of-interest tests like that used in *McClanahan*. Here, the Navajo Nation has a strong argument because tribal members and the tribe itself are directly benefiting from the services on the reservation. The only real reason for the state to regulate is to provide possible revenues for its Universal Service Fund. Again, as discussed above, this is a critical issue. If states are funding telecommunications infrastructure development or subsidizing tribal telecommunications in some way, then states have a stronger argument for regulation. However, as discussed further in the following section, private grants and some federal dollars have primarily helped fund the

147. *See Williams*, 358 U.S. at 223; *see also McClanahan*, 411 U.S. at 172.

148. *See generally Williams*, 358 U.S. 217; *McClanahan*, 411 U.S. 164.

149. *Cf. Montana v. Blackfeet Tribe of Indians,* 471 U.S. 759, 105 S.Ct. 2399 (1985) (invalidating a state tax on tribal royalty interests).

150. These three cases all involve non-tribal companies that have commercial relationships for services on the reservations with tribal members or the tribe itself. Preemption precludes state regulation in *White Mountain* and *Central Machinery* because the tribes are the buyers and the value added is on the reservation. *See* White Mountain Apache Tribe v. Bracker, 448 U.S. 136 (1980); Cent. Mach. Co. v. Ariz. Tax Comm'n, 448 U.S. 160 (1980); *see also Warren Trading Post*, 380 U.S. at 690.

advanced telecommunications infrastructure on Navajo land to date.[151] Furthermore, the fact that the FCC has worked to establish a government-to-government relationship with the Navajo Nation through policy statements goes to the heart of preemption analysis.[152] If a state were to assert authority over the Internet to the Hogan, it would need to show that its interests in regulating outweighed the interferences it would place on the government-to-government relationship between the FCC and the Navajo Nation.

Lastly, *Cabazon* provides an argument for final preclusion of any state effort to stop the Navajo Nation from providing advanced telecommunications services itself, even if a state—like Utah—prohibits municipal wireless broadband services.[153] In *Cabazon*, the Supreme Court held that if a state permits certain forms of conduct, and doesn't completely ban such conduct, it does not have authority to enforce its regulation within an Indian reservation if the state regulation is incident upon the tribe or its members.[154] Although a state might try to stop local governmental entities from providing ATSs, they rarely if ever ban private entities from providing advanced telecommunications services.[155] Thus, states would not have authority to enforce their telecommunications regulation in Indian Country, such as the Navajo Nation. Furthermore, ATS regulation falls under a civil/regulatory category, rather than a criminal/prohibitory category, the critical distinction found in *Cabazon*.[156] Accordingly, states should not be allowed to prohibit tribal governments from providing ATSs such as the Navajo's Internet to the Hogan Plan.[157]

151. *See infra* Part IV.

152. *See* FCC Tribal Policy, *supra* note 125.

153. *See Cabazon*, 480 U.S. at 209.

154. *Id.*

155. *See generally Nixon*, 541 U.S. at 125 (discussing situation in Missouri where private companies could provide broadband service but state wanted, and succeeded, in stopping local municipalities from providing municipal broadband services).

156. *Id.*

157. Furthermore, a state's ability to stop a municipality from providing ATSs is based on the fact that municipalities get their power delegated to them from the state. Tribal governments have inherent sovereignty that is not delegated to them from either the states or the federal government. *See* Talton v. Mayes, 163 U.S. 376, 384 (1896) (holding that tribal powers of self-government are not delegated but rather are inherent). Congress could step in with its plenary power and try to change this, effectively giving the state the power to stop tribally-provided ATSs. *See* Lone Wolf v. Hitchcock, 187 U.S. 553, 564-65 (1903) (holding that Congress has plenary power over tribes). However, the efforts in Congress recently have tended to go the other direction, geared at stopping the states that have passed prohibitions on municipal ATSs. *See* MUNICIPAL PROVISION REPORT, *supra* note 97, 35-38 (providing examples of proposed federal legislation preempting states from limiting municipal provision of wireless internet).

IV. WEAVING THE FUTURE NAVAJO.NET

> Navajo people consider the opening [of the hogan] as a pathway leading out of the enclosed space. It is always seen as the road out, and this road is the road of life. . . .The pathway through the opening in the circle shows that the space inside and the space outside, the time of story and the time of humans in history, the world of beauty and order and the world of ugliness and disorder, are intimately interdependent.[158]

Looking to the future of the Navajo Nation's Internet to the Hogan Plan, this paper concludes with several policy, regulatory, and statutory suggestions that could help foster more cultural empowerment to the Navajos as they work with their advanced telecommunications services. These recommendations focus on actions both the federal government and the Navajo Nation can take to enhance the provision and regulation of ATSs. At their root, these suggestions explore the idea of developing advanced telecommunications regulation in the tribal interest.

The idea of tribal interest regulation is a merging of two key principles within present telecommunications law and policy. The first foundation of tribal interest regulation comes from Title III spectrum regulation. Within Title III, the FCC is mandated to regulate the use of the airwaves in the name of public interest, necessity, and convenience.[159] This public interest mandate has resulted in numerous regulations that encourage diversity of voices over the public airwaves and localism for broadcasters.[160] While the public interest standard generally applied to TV and radio, it is still applicable to all uses of the electromagnetic spectrum. Emphasizing the importance of localism, the second foundation of tribal interest, the FCC's government-to-government relationship with the tribes,[161] would allow the Navajo Nation and the FCC together to establish principles for regulating in the tribal interest. The root of the tribal interest standard would be based on notions of cultural sovereignty and tribal self-government. The following sections address specific federal and Navajo programs—beyond the Internet to the Hogan Plan—and how they might incorporate this notion of regulating in the tribal interest.

158. GILL, *supra* note 51, at 24-25.
159. *See supra* Part II, A, iii.
160. *See* NUECHTERLEIN & WEISER, *supra* note 21, at 358.
161. *See supra* Part III, B.

A. Federal Programs

After releasing its government-to-government policy statement,[162] the FCC took several steps to try to improve access to telecommunications services on tribal lands like the Navajo Nation. At the same time, they also took steps that may have inadvertently thwarted these efforts. The FCC needs to reconsider how these programs and regulatory decisions may impact the interest of tribes like the Navajo in light of tribal programs like the Internet to the Hogan Plan.

i. FCC's Indian Telecommunications Initiative

In 2002, the FCC established the Indian Telecommunications Initiative ("ITI") to improve telecommunications services on tribal lands.[163] The ITI has 3 primary goals (1) increase the number of Native Americans who have phone service; (2) increase telecommunications infrastructure on tribal lands; and (3) "inform consumers in Indian Country about financial support available through federal government programs."[164] To date, the ITI has focused most of its work on educating tribes about the Enhanced Lifeline and Link-Up programs (discussed in the following section), which provide federal dollars to subsidize the cost of telephone service for individuals.[165] Senior FCC officials have also met with tribal leaders to discuss telecommunications issues on tribal lands.[166]

While educating tribes and their members about the availability of federal dollars to help pay for phone service may be useful, it does not address some of the more fundamental problems that plague tribal lands, where the basic technological infrastructure has not even been established to provide services like basic telephone. However, the Navajo's Internet to the Hogan Plan provides one means to leapfrog this problem—use wireless technology instead of traditional wirelines. The Internet to the Hogan Plan will allow for phone service to reach individual hogans, using VoIP technology.[167] The FCC's ITI should take into consideration such innovative ways of addressing long-standing infrastructure problems. Furthermore, instead of just educating tribes about the programs the FCC has to offer, the FCC should also use the ITI to educate itself about the specific tribal cultural issues that impact

---

162. *See* FCC Tribal Policy, *supra* note 125.

163. *See* FCC, Federal Tribal Initiatives, http://www.fcc.gov/indians/iti.html.

164. *Id.*

165. *See* UNITED STATES GOV'T ACCOUNTABILITY OFFICE, TELECOMMS.: CHALLENGES TO ASSESSING AND IMPROVING TELECOMMS. FOR NATIVE AMS. ON TRIBAL LANDS 27 (GAO-06-189, Jan. 2006) [hereinafter GAO TRIBAL TELECOM REPORT].

166. *Id.*

167. *See* INTERNET TO THE HOGAN TECH. PLAN, *supra* note 40, at 35.

telecommunications services on tribal lands. For example, the FCC could educate itself about how Navajos are utilizing Internet access via chapter houses to provide the FCC with more insight as to how Navajos perceive, incorporate, and maybe even reject the Internet within the context of Navajo culture. Considering cultural issues will thus allow the FCC to tailor its ITI program to help fit the notion of regulating in the tribal interest.

ii. The Universal Service Fund

As noted above, since 2000, the FCC has directed its ITI work towards educating tribes about federal dollars that were historically targeted at subsidizing low-income or rural individuals to help with paying for telecommunications services. Under the federal Universal Service Fund, the FCC implements two national programs, the Enhanced Link-Up and Lifeline programs.[168] The Enhanced Link-up program provides a one-time discount to help with installation fees for setting up the cost of basic phone service.[169] The Lifeline program provides ongoing discounts on basic telephone service, allowing low-income Indians on tribal lands to pay as low as $1 a month for such service.[170]

Unfortunately, USF data is not broken down into enough detail to allow for analysis of how many individual Indians or individual tribes have been able to take advantage of this program.[171] One important step that the FCC's ITI could take is to start collecting data on individual Indian and tribal utilization of USF funds. These subsidies, however, are generally targeted at phone service, either through wireline phones or wireless cell phones. The FCC should adjust this program so that Indians can utilize both programs to help connect individual Indian homes, like Navajo hogans, through alternative communications technology, like the broadband wireless network being established through the Internet to the Hogan Plan.

iii. Eligible Telecommunications Carriers

Another aspect of the USF is its High Cost program. This program provides funds to telephone service providers rather than subsidizing individuals. The funds are intended to help companies provide service to customers in remote or rural areas, where the cost of providing these

168. *Id.* at 23.
169. *Id.*
170. *Id.*
171. *Id.* at 24.

services is substantially higher than in urban areas.[172] In 2004, the High Cost Fund dwarfed the Lifeline and Link-up programs by nearly 360% ($3.5 billion compared to $758.8 million).[173] It is through programs like High Cost that communications companies can actually find the federal dollars needed to help deploy infrastructure into native communities. In addition, High Cost Funds are also used to support the deployment of broadband Internet infrastructure.[174]

In order for a telecommunications company to get High Cost Fund dollars, it must be designated as an Eligible Telecommunications Carrier ("ETC") by either the FCC or a state public utility commission.[175] Receiving ETC status commits these companies to certain obligations to provide service to rural and remote areas, and the ETC receives in exchange money to subsidize the work. A critical problem that has arisen with regard to ETCs is uncertainty over whether the surrounding state or the FCC needs to make the ETC designation for a telecommunications provider wanting to serve tribal lands. As the FCC has stated, this issue is "a legally complex and fact specific inquiry, informed by the principles of tribal sovereignty, federal Indian law, treaties, as well as state law."[176] The FCC decided these issues based on principles from *Montana v. United States*.[177]

In *Montana*, the Court held that Indian tribes, in general, cannot regulate the activities of non-Indians on land held in fee simple within reservation boundaries by non-tribal members.[178] There are several exceptions, however. First, a tribe may regulate such activities when the non-Indians have entered into contractual or business agreement with the tribe.[179] Second, a tribe may regulate such activities if the activities "so threaten the Tribe's political or economic security as to justify tribal regulation."[180] As argued in Part III.C above, the preemption analysis under *McClanahan* or *Warren Trading Post* is more appropriate in this context. First, the telecommunications services provided either by or for the Navajo Nation involve activities that will occur predominantly on

172. GAO TRIBAL TELECOM REPORT, *supra* note 165, at 22.

173. *See id.* at 21.

174. *See* Federal-State Joint Board on Universal Service, *Recommended Decision*, 22 FCC Rcd. 20,477, ¶ 10 (2007) [hereinafter Federal-State High Cost Decision].

175. *See Tribal ETC*, *supra* note 123, ¶¶ 112-127 (discussing ETC designation process on both non-tribal and tribal lands).

176. GAO TRIBAL TELECOM REPORT, *supra* note 165, at 29 (quoting *Tribal ETC*, *supra* note 123, ¶ 8).

177. Western Wireless Corporation Petition for Designation as an Eligible Telecommunications Carrier for the Pine Ridge Reservation, *Memorandum Opinion & Order*, 16 FCC Rcd. 18,145, ¶ 14 (2001).

178. 450 U.S. 544, 564-65 (1981).

179. *Id.* at 566.

180. *Id.*

tribal lands as there is very little land held by non-tribal members within the borders of the Navajo Nation; this suggests that the *Montana* rules do not even apply in this case. Second, in the case of non-tribal related telecommunications providers, the providers in almost all cases would have some contractual or business relationship with the Navajo Nation; this would give the Nation the authority to regulate. So, while the Navajo Nation may regulate based on this analysis, the surrounding states such as Arizona should have no authority to regulate based on the preemption analysis of Part III.C above.

One should note that one applicant for ETC status, Smith Bagley Inc., has been waiting over 3 years to hear if it would receive ETC status.[181] The FCC should reconsider the whole basis for how it is making these decisions. In order to regulate in the tribal interest, the FCC should allow individual tribes to determine ETCs that may provide subsidized telecommunications services on a tribe's land, taking on a role comparable to a state public utility commission.

Unfortunately, the FCC recently took steps that could curtail native telecommunications programs from tapping the High Cost Fund.[182] Because of widespread criticism of the USF's growing size, the FCC adopted an interim cap on competitive eligible telecommunications carriers ("CETC").[183] Thus, High Cost Fund dollars are only available to incumbent ETCs. Tribal lands often do not have an established or incumbent telephone provider, so this policy could effectively stop High Cost Fund dollars from reaching these communities. Furthermore, new companies are often the entities that are actively seeking opportunities to help deploy facilities and services to Native American households.[184] The FCC should reconsider its cap on CETCs, at least for new CETCs on tribal lands.

The issues surrounding ETCs suggest that the FCC needs to hear tribal voices more clearly when addressing these critical USF issues. In the continuing discussion over regulation in the tribal interest, placing tribal representatives on the Federal-State Joint Board on Universal Service would help deal with the problems tribes face in dealing with access to High Cost USF.

---

181. GAO TRIBAL TELECOM REPORT, *supra* note 165, at 29.

182. *See* Federal-State High Cost Decision, *supra* note 174, ¶35.

183. *Id.*

184. Comments of General Communications to *High Cost Universal Service Support Federal-State Joint Board on Universal Service*, WC Dkt. No. 05-337, ii (June 6, 2007), *available at* http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6519518397.

B. *Empowering the Navajo Nation Telecommunications Regulatory Commission*

Over twenty years ago, the Navajo Nation took its first steps to actively try to regulate telecommunications services on its tribal lands. In 1984, the Navajo Nation Council established the Navajo Nation Telecommunications Regulatory Commission ("NNTRC").[185] Through the NNTRC, the Nation "develop[ed] and adopt[ed] a regulatory code subject to Navajo Nation Council enactment, which will govern any and all Navajo Nation telecommunications activities, not inconsistent with Federal Communications Commission (FCC) regulation."[186] The NNTRC works to establish policies regarding regulation of telecommunications activities on the Navajo Nation including telephone, cellular, satellite, internet, and two-way radio.[187] The NNTRC's policies show indirectly how the Nation sees its role in regulating this industry as an exercise of tribal sovereignty over telecommunications on tribal lands. Besides the Internet to the Hogan Plan, the NNTRC has focused on other wireless technology deployment like cell phone towers.[188]

i. NNTRC Right of Way Regulation

To date, the NNTRC has directed most of its efforts at providing policies that facilitate the deployment of telecommunications infrastructure on tribal lands, focusing primarily on "enabl[ing] telecommunications providers to furnish comprehensive and efficient wireless communication services to the community, while minimizing the adverse impacts their facilities may have on neighboring properties."[189] The NNTRC presently works on ways to streamline the process to obtain tribal consent for rights-of-way ("ROW") on tribal land for telecommunications equipment and infrastructure.[190]

The ROW issue is one of the factors that has contributed to the

---

185. Navajo Nation Telecommunications Regulatory Commission, History, http://www.nntrc.org/content.asp?CustComKey=21364&CategoryKey=21413&pn=Page&DomName=nntrc.org.

186. *Id.*

187. *See* Navajo Nation Telecommunications Regulatory Commission, FAQs, http://www.nntrc.org/content.asp?CustComKey=24433&CategoryKey=24434&pn=FAQ&DomName=nntrc.org#1.

188. *See* Navajo Nation Telecommunications Regulatory Commission, Policy Documents, http://www.nntrc.org/content.asp?CustComKey=106193&CategoryKey=120237&pn=Page&DomName=nntrc.org.

189. Navajo Nation Telecommunications Regulatory Commission, Policy on ROW, http://www.nntrc.org/content.asp?CustComKey=106193&CategoryKey=120226&pn=Page&DomName=nntrc.org.

190. *See* GAO TRIBAL TELECOM REPORT, *supra* note 165, at 51.

slow deployment of telecommunications infrastructure onto tribal lands. The ROW process on tribal lands tends to be more complicated than on non-tribal land because the Bureau of Indian Affairs ("BIA") must give its approval for all ROW requests on Indian land.[191] The complexity of getting BIA approval has caused confusion and delays for some telecommunications service providers.[192] Part of the problem is that ROW regulations are outdated, focusing on "the size of the right-of-way needed and voltage levels of electrical equipment that may be installed for commercial purposes, but similar descriptions and guidance are not available for advanced telecommunications rights-of-way."[193] The NNTRC's effort to streamline the process for a wireless company to gain tribal approval for a ROW is one way to address this problem.

ii. A Proposal for Future NNTRC Regulation

While this paper has focused on the Navajo Nation's development of advanced telecommunications service infrastructure as a tool to develop their culture, another critical issue not explored in any great detail is the actual regulation of the content that is accessed and distributed through the Navajo's intranet network. As Angela Riley argues:

> [T]he proliferation of the Internet presents a new threat to indigenous cultural property. Sacred materials can be taken, distorted, and sent around the world almost instantaneously. Until recently, if a sacred ceremony was viewed without authorization from the tribe, the viewer had limited means of communicating the ceremony's contents. Today, if that same ceremony is recorded on a digital camera, for example, it can be placed on the Internet and sold or disseminated globally in a matter of moments. [194]

While traditional intellectual property law may not have proved effective in guarding against this threat to indigenous cultures,[195] the Navajo Nation may have another regulatory tool to consider—Intranet content regulation. As a first step down this path, the NNTRC, working with the tribal council, could work to fashion tribal codes that would impose penalties on anyone who uses the Navajo's Internet to the Hogan network to distribute secret sacred information.[196] While the

191. *Id.* at 37.
192. *Id.* at 37-39.
193. *Id.* at 38.
194. Riley, *supra* note 50, at 116.
195. *See id.*
196. For an interesting discussion of Navajo secret sacred information, see KATHY M'CLOSKEY, TOWARDS AN UNDERSTANDING OF NAVAJO AESTHETIC 14 (2004),

jurisdictional boundaries of the Navajo Nation may limit the effectiveness of this proposed regulation, it has potential for dealing in part with the problem of cultural misappropriation that has historically plagued tribes like the Navajo.

This proposed regulation suggests several questions that will have to await a future paper. Can the Navajo Nation legally regulate Internet content distributed through the Navajo network? Several issues immediately come to mind. If the Internet itself is "indivisibly interstate," would the FCC's or Congress's power in this area preempt any such regulation? What about the free speech implications of such regulation? The U.S. Congress has faced strict scrutiny from the Supreme Court every time that it has tried to regulate Internet content.[197] Yet, tribal governments are not limited by the federal Constitution, but rather the Indian Civil Rights Act ("ICRA"). What are the implications of *Santa Clara Pueblo v. Martinez*[198] in assessing the remedies someone could seek if the Navajo Nation imposed penalties on them for violating this regulation? Does the compelling interest standard apply to free speech claims brought before tribal courts with respect to tribal Internet content regulation? All these questions suggest there is still much to explore within the topic of the Internet and Indian tribes.

CONCLUSION

As Navajo President Shirley has recognized, the Internet and information technology will play a critical role in the future of the Navajo Nation.[199] This paper examined the cultural and legal issues surrounding the Navajo Nation's efforts to provide and regulate ATSs, specifically high-speed wireless Internet access, on tribal lands. Navajo culture has a long history of technological incorporation, such as weaving and silversmithing. New broadband infrastructures and the Internet used as a cultural preservation and advancement tool will be no different. The Navajo's most recent plans to develop its own high-speed wireless Internet network called "Internet to the Hogan" reflect this continued cultural adaptation. It is time for the Navajo Nation to take a more active

---

http://www.library.utoronto.ca/see/SEED/Vol4-1/M%27Closkey.pdf.

197. *See* Reno v. ACLU, 521 U.S. 844, 879 (1997) (applying strict scrutiny to Communications Decency Act of 1996, finding the Act not narrowly tailored).

198. *See* 436 U.S. 49, 58-59 (1978) (holding that a federal court does not have jurisdiction to resolve actions brought against an Indian tribe or one of its officers under ICRA, except for a habeas corpus petition); *see also* Tribal Restrictions on Sharing of Indigenous Knowledge on Uses of Biological Resources, Memorandum for the Assistant Attorney General for the Environment and Natural Resources Division, Oct. 12, 1999 (suggesting federal precedents under analogous constitutional provisions should be applied to tribal government action under ICRA), *available at* http://www.usdoj.gov/olc/biodiv14.htm.

199. *See* Cullen, *supra* note 1, at 1.

role in overseeing the regulation of their advanced telecommunication services. While the Federal Government has a part to play in this regulation, the Navajo Nation has several strong arguments it can make if the states try to interfere with the Navajo's Internet to the Hogan Plan. In weaving the Navajo.Net, the Nation can develop regulation for its advanced telecommunications in the tribal interest and look to other ways to help preserve Navajo culture through Internet regulation.