

BERKELEY TECHNOLOGY LAW JOURNAL

VOLUME 25

NUMBER 4

FALL 2010

TABLE OF CONTENTS

ARTICLES

DATA SHARING, LATENCY VARIABLES, AND SCIENCE COMMONS	1601
<i>Jorge L. Contreras</i>	
OPERATING EFFICIENTLY POST-BILSKI BY ORDERING PATENT DOCTRINE DECISION-MAKING.....	1673
<i>Dennis Crouch & Robert P. Merges</i>	
PATENT LAW'S PARSIMONY PRINCIPLE	1693
<i>Alan Devlin</i>	
THE ILLUSION OF THE COMMONS	1751
<i>Jonathan M. Barnett</i>	
THE ROLE OF THE RULE OF LAW IN VIRTUAL COMMUNITIES	1817
<i>Nicolas Suzor</i>	
WHY LICENSE AGREEMENTS DO NOT CONTROL COPY OWNERSHIP: FIRST SALES AND ESSENTIAL COPIES.....	1887
<i>Brian W. Carver</i>	

NOTES & COMMENTS COMPETITION WINNER

IQBAL-ING SEAGATE: PLAUSIBILITY PLEADING OF WILLFUL PATENT INFRINGEMENT	1955
<i>Damon C. Andrews</i>	

SUBSCRIBER INFORMATION

The *Berkeley Technology Law Journal* (ISSN1086-3818), a continuation of the *High Technology Law Journal* effective Volume 11, is edited by the students of the University of California, Berkeley School of Law (Boalt Hall), and published four times each year (May, August, November, February) by the Regents of the University of California, Berkeley. Journal Publications, School of Law. Periodicals Postage Rate Paid at Berkeley, CA 94704-9998, and at additional mailing offices. POSTMASTER: Send address changes to Journal Publications, 2850 Telegraph Avenue, Suite 561 #7220, Berkeley, CA 94705-7220.

Correspondence. Address all correspondence regarding subscriptions, address changes, claims for non-receipt, single copies, advertising, and permission to reprint to Journal Publications, 2850 Telegraph Avenue, Suite 561 #7220 Berkeley, CA 94705-7220; (510) 643-6600; JournalPublications@law.berkeley.edu. Authors: see section entitled Information for Authors.

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

Form. The text and citations in the *Journal* conform generally to the THE CHICAGO Manual OF STYLE (15th ed. 2003) and to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (Columbia Law Review Ass'n et al. eds., 19th ed. 2010). Please cite this issue of the *Berkeley Technology Law Journal* as 25 BERKELEY TECH. L.J. ____ (2010).

BTLJ ONLINE

The full text and abstracts of many previously published *Berkeley Technology Law Journal* articles can be found at <http://www.btlj.org>. Our site also contains a cumulative index, general information about the *Journal*, selected materials related to technology law, and links to other related pages.

DATA SHARING, LATENCY VARIABLES, AND SCIENCE COMMONS

Jorge L. Contreras[†]

TABLE OF CONTENTS

I.	INTRODUCTION	1602
II.	COMMONS AND LATENCY	1609
A.	COMMONS THEORY BACKGROUND	1610
1.	<i>Physical Resource Commons</i>	1610
2.	<i>Commons of Information</i>	1611
B.	THE TEMPORAL DIMENSION OF INFORMATION COMMONS.....	1613
1.	<i>Dynamic and Static Commons</i>	1613
2.	<i>Knowledge, Rights, and Timing in Information Commons</i>	1614
a)	The Dynamic Character of Information Commons: An Example.....	1614
b)	The Changing Bases of Knowledge and Rights in an Information Commons.....	1614
3.	<i>A Proposed Latency Analysis for Information Commons</i>	1616
a)	Knowledge Latency.....	1616
b)	Rights Latency.....	1618
c)	Latency Analysis	1618
d)	Implementing Latency Analysis in Commons Design	1620
III.	COMMONS IN THE SCIENCES	1621
A.	CONSTRUCTING SCIENCE COMMONS.....	1621
1.	<i>Incentives to Share</i>	1621
2.	<i>Modes of Data Sharing</i>	1625
3.	<i>Stakeholders and Policy Considerations</i>	1628

© 2011 Jorge L. Contreras.

[†] Acting Director of Intellectual Property Programs and Senior Lecturer in Law, Washington University in St. Louis. The author wishes to thank the following for their thoughtful comments: Scott Baker, Sara Bronin, Dan Burk, Rebecca Eisenberg, Terry Fisher, Charlotte Hess, Kimberly Kaphingst, Michael Madison, Charles McManis, Ruth Okediji, Nirej Sekhon, the participants in the 2009 SALT Junior Faculty Development Workshop, the 2010 Junior Scholars in Intellectual Property (JSIP) workshop and the 2010 Intellectual Property Scholars Conference (IPSC), and research assistant Jeff Johnson.

B.	EXOGENOUS INFLUENCES ON LATENCIES WITHIN A SCIENCE COMMONS	1630
1.	<i>Publication Delay</i>	1631
2.	<i>Copyright and Other Restraints on the Dissemination of Data</i>	1633
3.	<i>Patents and Restraints on the Use of Data</i>	1635
a)	Effect of Patents on Scientific Data.....	1635
b)	Patents and Latency	1636
c)	Patents and Anticommons.....	1637
IV.	TIMING AND DATA SHARING: TWO CASE STUDIES	1641
A.	THE GENOME COMMONS.....	1641
1.	<i>Development of Genomics Data Release Policies</i>	1641
a)	The Human Genome Project (HGP)	1641
b)	The Bermuda Principles	1644
c)	Bermuda, Data Release, and Patents	1645
d)	Growing Tension Between Data Generators and Users	1647
e)	GWAS and Embargo Policies	1648
f)	Private Sector Policies.....	1650
B.	SCIENTIFIC PUBLISHING.....	1652
1.	<i>Scientific Publishing and the Open-Access Debate</i>	1652
2.	<i>Open Access and Government-Funded Research</i>	1654
3.	<i>Alternative Open-Access Models</i>	1655
V.	APPLYING THE LATENCY FRAMEWORK	1657
A.	LATENCY IN THE GENOME COMMONS.....	1657
1.	<i>Stakeholders and Policy Design Considerations</i>	1657
2.	<i>Latency as a Policy Design Tool—Retention versus Embargo</i>	1660
B.	LATENCY AND SCIENTIFIC PUBLISHING.....	1664
1.	<i>Stakeholders and Policy Design Considerations</i>	1664
2.	<i>Compromises Effected Through Latency Choices</i>	1665
C.	SYNTHESIS AND BROADER APPLICATIONS.....	1667
VI.	CONCLUSION	1671

I. INTRODUCTION

The fossilized remains of *ardipithecus ramidus* (“*Ardi*”), an early human ancestor and important evolutionary link, were unearthed in 1992 by archeologists from the University of California, Berkeley, and the University of Tokyo. The scientists published their initial findings in 1994 to great

acclaim,¹ but did not release detailed data about the fossils until 2009,² seventeen years after the initial discovery. This long delay, justified by the researchers as necessary due to the fragile condition of the remains and the difficulty of extracting and reassembling them, has nevertheless been widely criticized as contrary to the practice of “open science.”³

Data is the currency of science, and the sharing of data is fundamental to the scientific enterprise.⁴ Since the emergence of modern scientific practice, scientists have published their observations and experimental data to earn professional recognition, permit verification of their theories, and further the overall progress of science.⁵ Today, powerful database and networking technologies have enabled the dissemination of data on a scale not

1. Tim D. White et al., *Australopithecus Ramidus, a New Species of Early Hominid from Aramis, Ethiopia*, 371 NATURE 306 (1994); see also John Noble Wilford, *Fossil Find May Show if Prehuman Walked*, N.Y. TIMES, Feb. 21, 1995, at C1 (calling *a. ramidus* “potentially the most significant discovery in early human studies since the Lucy fossils in 1974”).

2. Tim D. White et al., *Ardipithecus Ramidus and the Paleobiology of Early Hominids*, 326 SCI. 65 (2009).

3. See, e.g., Joel Achenbach, *Ancient Skeleton Could Rewrite the Book on Human Origins*, WASH. POST, Oct. 2, 2009 (noting “impatience” in the scientific community while waiting for the *Ardi* data); Editorial, *Fossils for All: Science Suffers by Hoarding*, SCI. AM., Aug. 2009, at 26 (“[F]ossil hunters often block other scientists from studying their treasures, fearing assessments that could scoop or disagree with their own. In so doing, they are taking the science out of paleoanthropology.”).

4. See JOHN M. ZIMAN, PROMETHEUS BOUND: SCIENCE IN A DYNAMIC STEADY STATE 40 (1994); Yochai Benkler, *Commons-Based Strategies and the Problems of Patents*, 305 SCI. 1110, 1110 (2004). In this context, I (and most other commentators) use the term “science” to refer to “basic” research, or the investigation of fundamental natural laws and properties, rather than applied research or technology development that is directed to commercial application. See, e.g., ZIMAN, *supra*, at 24–26; Richard R. Nelson, *The Simple Economics of Scientific Research*, 67 J. POL. ECON. 297, 300–01 (1959) (discussing the imprecise line between basic and applied research). The norms of openness that apply to basic research do not necessarily apply to these later categories of investigation, and many of these are, quite naturally, conducted by corporate and industrial concerns in secret. See, e.g., Arti Kaur Rai, *Regulating Scientific Research: Intellectual Property Rights and the Norms of Science*, 94 NW. U. L. REV. 77, 93 (1999).

5. See Robert K. Merton, *The Normative Structure of Science (1942)*, in THE SOCIOLOGY OF SCIENCE 267, 267–78 (Norman W. Storer ed., 1973); discussion *infra* Section III.A. Merton famously identifies four “norms” of science: universalism (scientific claims should be evaluated using objective criteria consistent with observation), communism (now generally referred to as “communalism”) (scientific findings should belong to the scientific community as a whole), disinterestedness (scientists should have no emotional or financial attachment to their work), and organized skepticism (scientists should act dispassionately, without regard to personal beliefs). See also ZIMAN, *supra* note 4, at 77; Benkler, *supra* note 4, at 1110 (“[O]pen distribution of the inputs and outputs of the scientific process are its organizational norms.”); Rai, *supra* note 4, at 90–91 (discussing in detail the norms of scientific research, including the Mertonian norm of communalism).

imaginable just a few decades ago. Developed primarily by government-funded research projects, vast collections of publicly accessible⁶ data now exist in fields such as chemistry,⁷ meteorology,⁸ geophysics,⁹ astronomy,¹⁰ paleontology,¹¹ and, as will be discussed *infra*, molecular biology and genomics.¹² Other data collections have been proposed in a wide range of additional disciplines.¹³ These aggregations of public scientific data, sometimes referred to as “science commons,”¹⁴ can serve as useful resources

6. My use of the term “publicly accessible” connotes that data may be accessed by any researcher, though not necessarily without charge. Some of the databases discussed in this Article charge for access, just as most scientific journals charge for subscriptions. For a discussion of this issue, see *infra* Section IV.B.1. Also, even to the extent that data is “publicly” available within a commons, various practical, technical, and logistical issues may impact individuals’ ability to access and use that data. See Jorge L. Contreras, *Prepublication Data Release, Latency and Genome Commons*, 329 SCI. 393, 394 n.16 (2010); see also *infra* note 63 and accompanying text.

7. See Robert Potenzzone, *Opportunities for Commercial Exploitation of Networked Science and Technology Public-Domain Information Resources*, in NAT’L RESEARCH COUNCIL, THE ROLE OF SCIENTIFIC AND TECHNICAL DATA AND INFORMATION IN THE PUBLIC DOMAIN: PROCEEDINGS OF A SYMPOSIUM 52, 53 (2003) [hereinafter NRC, PUBLIC DOMAIN].

8. See U.S. NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION’S (“NOAA”) NATIONAL DATA CENTERS, <http://nesdis.noaa.gov>; BRITISH ATMOSPHERIC DATA CENTER (BADc), <http://badc.nerc.ac.uk> (discussed in DIGITAL ARCHIVING CONSULTANCY ET AL., LARGE-SCALE DATA SHARING IN THE LIFE SCIENCES: DATA STANDARDS, INCENTIVES, BARRIERS AND FUNDING MODELS app. A-2 (2005) [hereinafter JOINT DATA STANDARDS STUDY]).

9. See U.S. GEOLOGICAL SURVEY’S EARTH RESOURCES OBSERVATION SYSTEMS (“EROS”) DATA CENTER, <http://eros.usgs.gov> (last visited Nov. 22, 2010).

10. See, e.g., M. Jordan Raddick & Alexander S. Szalay, *The Universe Online*, 329 SCI. 1028 (2010) (discussing the massive Sloan Digital Sky Survey (www.sdss.org)); NATIONAL AERONAUTICS AND SPACE ADMINISTRATION’S (NASA) SPACE SCIENCE DATA CENTER, <http://nssdc.gsfc.nasa.gov>; *About the PDS*, PLANETARY DATA SYSTEM (PDS), <http://pds.nasa.gov/about/about.shtml> (last visited Sept. 17, 2010).

11. See THE OPEN DINOSAUR PROJECT, <http://opendino.wordpress.com> (last visited Sept. 17, 2010).

12. See *infra* Section IV.A.

13. See, e.g., JAMES BOYLE, THE PUBLIC DOMAIN: ENCLOSING THE COMMONS OF THE MIND 171–78 (2008) (describing the MIT Registry of Standard Biological Parts); Geoff Brumfiel, *Chemists Spin a New Web of Data*, 453 NATURE 139 (2008) (describing a new open-access source of data on molecular chemistry); David Einhorn & Rita Heimes, Letter to the Editor, *Creating a Mouse Academic Research Commons*, 27 NATURE BIOTECHNOLOGY 890 (2009) (describing the impetus for creating a public resource for genetically modified mouse strains); Peter A. Stott & Peter W. Thorne, *How Best to Log Local Temperatures?*, 465 NATURE 158 (2010) (describing efforts to develop international data sharing of local temperature information).

14. This term finds its root in shared physical resources that have been broadly categorized as “commons.” See *infra* notes 19–23 and accompanying text; *infra* Section II.A. A “science commons” is a species of commons that is devoted to scientific data and

for the global scientific community and enable collaboration and information sharing at unprecedented levels.

But just as these public data resources proliferate, there are signs that all is not well in the science commons. As the example of *Ardi* strikingly illustrates, significant delays between the generation of scientific data and its disclosure to the public arise for a host of reasons, including publication lag time, intellectual property restrictions, and institutional and interpersonal inefficiencies.¹⁵ The recent outcry over BP's attempt to maintain the secrecy of data collected from the catastrophic *Deepwater Horizon* oil spill¹⁶ reflects growing public and academic unease with the concealment and delayed release of scientific data.¹⁷ This unease is shared by scholars who debate the benefits and effects of science-based commons, whether intellectual property protection threatens the existence of such commons, and whether intellectual property systems hinder or promote innovation and discovery.¹⁸

information. My use of the term "science commons" is not intended to reference the non-profit organization known as Science Commons, which undertakes various projects, including tool development, to make scientific data more accessible to the public. *See About Science Commons*, SCIENCE COMMONS, <http://sciencecommons.org/about/> (last visited Sept. 14, 2010).

15. *See infra* Section III.B.

16. *See* Amanda Mascarelli, *Freedom of Spill Research Threatened*, 466 NATURE 538 (2010); Lauren Schenkman, *After Outcry, Oil Data Inches into the Open*, 329 SCI. 888 (2010) (describing BP's retreat from its initial position following public criticism of its data concealment practices).

17. *See, e.g.*, ZIMAN, *supra* note 4, at 40–41 (“[Increasing secrecy] not only slows the advance of knowledge: it also puts a damper on public assessments of research claims, which are the ultimate arbiter of scientific validity.”); *see also* David Blumenthal et al., *Data Withholding in Genetics and the Other Life Sciences: Prevalences and Predictors*, 81 ACAD. MED. 137, 137 (2006) [hereinafter Blumenthal et al., *Prevalences and Predictors*] (“[D]ata withholding is common in biomedical science.”); David Blumenthal et al., *Withholding Research Results in Academic Life Science: Evidence from a National Survey of Faculty*, 277 J. AM. MED. ASSN. 1224 (1997) [hereinafter Blumenthal et al., *National Survey*] (reporting that 8.9% of academic life scientists have refused to share research results with other scientists within the past three years); Ryan P. O'Donnell et al., *Hindrance of Conservation Biology by Delays in the Submission of Manuscripts*, 24 CONSERVATION BIOLOGY 615 (2010) (examining delays in submission of data for publication in the field of conservation biology); Elizabeth Pisania et al., *Time for Fair Trade in Research Data*, 375 LANCET 703 (2010) (describing a lack of data sharing in the field of public health research); Josh Sommer, *The Delay in Sharing Research Data is Costing Lives*, 16 NATURE MED. 744 (2010); *Fight Over Access to 'Hobbit' Bones*, NEWSIDENTIST, Dec. 11, 2004, available at <http://www.newscientist.com/article/mg18424772.900-fight-over-access-to-hobbit-bones.html> (reporting that a researcher was isolating valuable samples from other researchers).

18. *See* BOYLE, *supra* note 13; Elinor Ostrom & Charlotte Hess, *A Framework for Analyzing the Knowledge Commons*, in UNDERSTANDING KNOWLEDGE AS A COMMONS: FROM

Commons theory offers an attractive and timely framework for the analysis of scientific data sharing. Over the past two decades, legal scholarship, economic analysis, and organizational theory have become deeply invested in the theoretical construct of the commons.¹⁹ Taking as its foundational metaphor the non-proprietary physical resources shared by traditional societies such as pastures, fisheries, and forests, the systematic study of commons systems was pioneered by Elinor Ostrom in the 1980s and early 1990s.²⁰ Her work in this area earned her the 2009 Nobel Prize in Economics. Among Ostrom's many insights was the application of the well-known Institutional Analysis and Development (IAD) framework applied by organizational theorists to privately ordered commons structures.²¹ Ostrom, together with Charlotte Hess, has recently extended this methodology to the analysis of so-called "information commons," shared intangible resources as varied as audiovisual content, open-source software, digital book archives, medical records, and traditional knowledge.²² More recently, Michael Madison, Brett Frischmann, and Katherine Strandburg have undertaken a thorough re-examination of the IAD framework in relation to commons in the "cultural environment."²³

Despite the abundance of debate and scholarship on this topic, until now, little attention has been paid to the *temporal* dimension of information commons—that is, the rate at which information is added to, or subtracted from, the commons.²⁴ These temporal characteristics are crucial, as they

THEORY TO PRACTICE (Charlotte Hess & Elinor Ostrom eds., 2006) [hereinafter KNOWLEDGE AS A COMMONS]; *infra* Section III.B.

19. A sense of the volume of this scholarship can be obtained from the online Digital Library of the Commons hosted at Indiana University, <http://dlc.dlib.indiana.edu/dlc/> (last visited July 24, 2010), which, at last count, included nearly 6,000 scholarly articles relating to the commons.

20. *See* ELINOR OSTROM, GOVERNING THE COMMONS — THE EVOLUTION OF INSTITUTIONS FOR COLLECTIVE ACTION (1990).

21. *See infra* Section II.A.2.

22. *See* Ostrom & Hess, *supra* note 18, at 42–43.

23. Michael J. Madison, Brett M. Frischmann & Katherine J. Strandburg, *Constructing Commons in the Cultural Environment*, 95 CORNELL L. REV. 657, 659 (2010). Madison, Frischmann, and Strandburg refer to aggregations of shared information as "cultural commons" and include within their far-ranging analysis shared resource structures as varied as patent pools, open source software, Wikipedia, the Associated Press, and jamband fan communities. *Id.* at 660–63. For purposes of this Article, I adopt Ostrom's terminology of "information commons," as this term more intuitively characterizes the scientific data that is the central focus of this Article.

24. It is important to distinguish between the dynamic nature of an information resource, which is the focus of this Article, and the dynamic nature of the community that interacts with and creates an information commons, which has been analyzed by others. For

define the very nature of the resource being shared and are essential to the overall functioning and value of the commons. In this Article, I describe a novel analytical tool, which I term “latency analysis,” for analyzing and designing the temporal features of information commons.²⁵ I place this analytical tool within the organizational frameworks proposed by Ostrom and Hess, and Madison, Frischmann, and Strandburg, though its applicability is not necessarily limited to these frameworks.

Latency analysis utilizes two key variables that characterize all information commons: the rate at which information enters the commons, its *knowledge latency*, and the rate at which the knowledge in the commons becomes freely utilizable, its *rights latency*.²⁶ With these two variables in mind, latency analysis provides a three-step analytical methodology that consists of (1) determining the stakeholder groups relevant to the information commons,²⁷ (2) determining the policy objectives that are relevant to each stakeholder group,²⁸ and (3) mediating among the differing positions of the stakeholder groups through adjustments in the latency variables of the commons.

The latency analysis that I develop in this Article is both descriptive and prescriptive. Not only is it a useful tool for analyzing existing information

example, while Ostrom and Hess acknowledge the dynamic nature of information commons, Ostrom & Hess, *supra* note 18, at 41–42 (“[T]he framework can also be used to analyze dynamic situations where individuals develop new norms, new rules, new physical technologies.”), the focus of their analysis is on the dynamics of the institutional and organizational *structures* that utilize the commons, rather than the dynamic characteristics of the commons themselves. Likewise, Madison, Frischmann, and Strandburg, *supra* note 23, at 682, acknowledge the “iterative” nature of rules and resources within a cultural commons and state that “[d]ynamic effects are central to [their] analysis.” *Id.* at 673. However, like Ostrom and Hess, they focus on the dynamic characteristics of the communities (musicians, open source developers, etc.) who participate in commons construction. While they refer briefly to the expiration of patents and copyrights when explaining that “[t]he durability of shared resources must be considered,” *id.* at 689, by and large they do not focus on the dynamic characteristics of the common resource itself.

25. I first described this approach in Contreras, *supra* note 6, at 393.

26. *Id.* at 393.

27. See Madison, Frischmann & Strandburg, *supra* note 23, at 690 (prescribing the identification of constituencies and their part in commons formation); Ostrom & Hess, *supra* note 18, at 48–50 (noting the importance of defining the “attributes of the community”).

28. See Madison, Frischmann & Strandburg, *supra* note 23, at 691–93 (identifying community goals and objectives); Ostrom & Hess, *supra* note 18, at 48–50 (considering goals and objectives among the attributes of the community).

commons;²⁹ it also offers policy designers an objective set of variables with which to mediate otherwise value-based negotiations among stakeholders.³⁰ The introduction of latency variables into the commons formation discourse thus enables parties to avoid often fruitless debate over entrenched values-based preferences (for example, intellectual property protection over the public domain) and to achieve compromise based on simple numerical calculations.³¹ In this sense, the use of latency variables in policy design helps to reduce the transaction costs of negotiating commons policy, to achieve efficient and equitable results for all stakeholders, and thereby to facilitate the formation of socially valuable commons of information.

In Part II, I review the history and current state of commons theory, then describe latency analysis and its place within existing analytical frameworks for the commons. In Part III, I describe the formation of commons in the sciences, relevant stakeholder groups, and modes of data sharing. I then address the principal exogenous factors that have been implicated in limiting the growth of science commons, particularly delays induced by the scientific publication process and intellectual property rights that either limit the entry of knowledge into a commons (i.e., copyright and trade secret) or limit the usability of knowledge once it enters (i.e., patent). In this Part, I also discuss the well-known debate over the “anticommons” and its implications for the formation and utility of science commons.³²

In Part IV, I describe two well-known narratives of commons formation in the sciences: the “genome commons”³³ and open-access scientific publishing.³⁴ In each of these cases, I analyze the cultural, legal, political, and historical influences that shaped the development of the relevant scientific commons, as well as the competing policy objectives of the relevant

29. In this sense, answering the call of Madison, Frischmann, and Strandburg for greater study of existing commons structures and institutions. Madison, Frischmann & Strandburg, *supra* note 23, at 708–09.

30. *See infra* Section V.C.

31. This technique has long been recognized as an effective negotiation practice. *See, e.g.*, ROGER FISHER, WILLIAM URY & BRUCE PATTON, *GETTING TO YES: NEGOTIATING AGREEMENT WITHOUT GIVING IN* 4–12, 81–92 (2d ed. 1991) (illustrating the differences between “positional” and “principled” negotiation and the value of negotiating from objective criteria).

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

33. I adopt this term in Contreras, *supra* note 6.

34. Madison, Frischmann, and Strandburg advocate blending the traditional functionalist account of cultural commons with a “metaphorical or narrative approach.” Madison, Frischmann & Strandburg, *supra* note 23, at 671.

stakeholder groups. Section IV.A begins with an overview of the groundbreaking set of privately ordered principles³⁵ adopted in 1996 requiring that all genetic sequence data generated by the Human Genome Project (HGP) be released in publicly accessible databases a mere twenty-four hours after generation. The Bermuda Principles continue to shape data release practices in genomics and other fields,³⁶ but not without some important caveats. Most significant among these are timing-based policies that were implemented to mediate among the competing interests of data-generating scientists, data-using scientists, and public funding institutions, and which have resulted in two distinctly different approaches to commons formation. In Section IV.B, I discuss the recent debate over “open access” scientific publishing and the commons of scientific knowledge more broadly. In particular, I draw attention to private bilateral compromises among research institutions and publishers as well as new federal policies that are intended to balance the competing interests of publishers and the public by imposing specified time periods before the public release of scientific papers is required.

In Part V, I apply the methods of latency analysis to each of these case studies, analyzing in greater detail the various timing-based policy features that enabled compromise among competing stakeholder positions. I then draw a number of conclusions regarding the success of latency-based mediation of commons policy negotiation positions and suggest areas in which further theoretical and empirical work may be of value.

II. COMMONS AND LATENCY

35. HUGO, *Summary of Principles Agreed at the First International Strategy Meeting on Human Genome Sequencing*, HUMAN GENOME PROJECT INFORMATION (1996), http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml/ [hereinafter Bermuda Principles]. A more detailed discussion of the Bermuda Principles can be found in Jorge L. Contreras, *Bermuda's Legacy: Policy, Patents and the Design of the Genome Commons*, 12 MINN. J. L. SCI. & TECH. 61 (2011).

36. See, e.g., Francis Collins, *Opinion: Has the Revolution Arrived?*, 464 NATURE 674, 675 (2010) (referring to the “radical ethic of immediate data deposit” adopted by the HGP as the current “norm for other community research projects”); Jane Kaye et al., *Data Sharing in Genomics — Re-shaping Scientific Practice*, 10 NATURE REV. GENETICS 331, 332 (2009) (“[T]hese policies have created a climate in which data sharing has become the default, and [grant] applicants must demonstrate why their data should be exempt from the requirement that it should be deposited for use by other scientists.”); Nikos Kyrpides, *Fifteen Years of Microbial Genomics: Meeting the Challenges and Fulfilling the Dream*, 27 NATURE BIOTECHNOLOGY 627, 627 (2009) (“Over time, as the substantial benefits of prepublication release of genome data have been recognized, many funding agencies and most of the large sequencing centers now adhere to the rapid data release policy set forth as the Bermuda Principles.”).

A. COMMONS THEORY BACKGROUND

1. *Physical Resource Commons*

Since the Middle Ages, the term “commons” has denoted shared physical spaces such as fields, pastures, and forests that were open and free for exploitation by farmers, herdsmen, and other local peoples.³⁷ In 1968, biologist Garrett Hardin described an effect he termed the “tragedy of the commons,” rooted in Malthusian notions of population growth and the deleterious effects of unbounded utilization of depletable common resources.³⁸ Hardin employed the metaphor of a shared pasture to illustrate the social harm that may arise when individual herdsmen seek to maximize their own gain by allowing ever more animals to graze on the common land. The result, of course, is the critical depletion or destruction of the shared resource. “Therein is the tragedy,” Hardin wrote, “[e]ach man is locked into a system that compels him to increase his herd without limit—in a world that is limited.”³⁹ For Hardin, a “commons” connoted a resource that was accessible to many without constraint, such as a pasture, a forest, or the ocean. Later scholars have defined this attribute of commons as “non-excludability”: the inability to limit the use of a resource due to its inherent features, such as size or accessibility.⁴⁰

37. See Ostrom & Hess, *Introduction to KNOWLEDGE AS A COMMONS*, *supra* note 18, at 12; NANCY KRANICH, *THE INFORMATION COMMONS — A PUBLIC POLICY REPORT 10* (2004), available at <http://www.fepproject.org/policyreports/InformationCommons.pdf>. In the U.S., “commons” have also been associated historically with New England’s open town squares that served as popular venues for speechifying and pamphleteering. Ostrom & Hess, *supra*, at 13. In both cases, “commons” terminology has a strong traditional association with freedom and openness.

38. Garrett Hardin, *The Tragedy of the Commons*, 162 *SCI.* 1243, 1244 (1968). The first modern economic analyses of commons structures are usually credited to H. Scott Gordon and Anthony Scott, who based their models on studies of fisheries in the mid 1950s. See Charlotte Hess & Elinor Ostrom, *Ideas, Artifacts, and Facilities: Information as a Common-Pool Resource*, 66 *LAW & CONTEMP. PROBS.* 111, 115–16 (2003).

39. Hardin, *supra* note 38, at 1244. Hardin’s concern is not specifically with pasture land, but with all scenarios in which natural resources may be depleted due to a “tragedy of the commons,” including ocean fish stocks, national parks, the environment, and the earth’s ability to support a rapidly growing population. Elinor Ostrom cites numerous additional problems that have been analyzed using commons theory, including famines, acid rain, and urban crime. OSTROM, *supra* note 20, at xv.

40. See OSTROM, *supra* note 20, at 30. Other definitions and variations on this theme have been addressed in the literature. See, e.g., Hess & Ostrom, *supra* note 38, at 114–18 (summarizing several commons approaches and theories).

As noted above, Elinor Ostrom conducted the seminal analysis of social and organizational structures governing physical commons in the 1980s.⁴¹ Among Ostrom's many insights was the applicability of the well-known Institutional Analysis and Development (IAD) framework, employed since the 1970s to evaluate organizational characteristics and institutional decision-making, to common-pool resources.⁴² Under the IAD framework, commons structures may be examined with respect to three broad sets of characteristics: (1) those of the common resource itself, (2) the "action arena" in which the common resource is utilized, and (3) the desired or actual outcomes of the commons structure.⁴³ Each of these broad areas may be subdivided into further analytical components, so that the common resource, for example, is assessed with respect to its bio-physical characteristics, the attributes of the relevant community, and the applicable rules set, whether legal or norms-based.⁴⁴ The application of the IAD framework analysis results in a deeper understanding of the factors that should be considered when structuring or evaluating a commons structure, and Ostrom and others have persuasively applied the IAD framework to common resource arrangements ranging from fisheries to irrigation systems to environmental governance.⁴⁵

2. *Commons of Information*

In the mid-1990s, scholars began to apply commons theory to intangible shared resources and information.⁴⁶ Since then, much has been written about so-called "information commons" in areas including computer software, network capacity, artistic content, scholarly learning, and scientific data.⁴⁷ In

41. OSTROM, *supra* note 20.

42. See Charlotte Hess & Elinor Ostrom, *A Framework for Analysing the Microbiological Commons*, 58 INT'L SOC. SCI. J. 335, 339 (2006); Ostrom & Hess, *supra* note 18, at 42–43.

43. See Ostrom & Hess, *supra* note 18, at 44–45.

44. *Id.* at 45–53.

45. See Hess & Ostrom, *supra* note 42, at 339.

46. Hess & Ostrom, *supra* note 37, at 4 (noting the "explosion" of information commons scholarship beginning around 1995).

47. See, e.g., HAL ABELSON, KEN LEDEEN & HARRY LEWIS, BLOWN TO BITS — YOUR LIFE, LIBERTY, AND HAPPINESS AFTER THE DIGITAL EXPLOSION 277 (2008) (discussing the application of commons theory to broadcast spectrum); LAWRENCE LESSIG, THE FUTURE OF IDEAS 85–86 (2001) (arguing that commons systems have encouraged innovation, specifically with respect to software, telecommunications, and the Internet); JONATHAN ZITTRAIN, THE FUTURE OF THE INTERNET AND HOW TO STOP IT 78–79 (2008) (discussing commons approaches both to Internet content and hardware); Yochai Benkler, *Coase's Penguin, or Linux and the Nature of the Firm*, 112 YALE L.J. 369 (2002) (arguing that "commons-based peer production" of software has proven to be both viable and efficient, as

these discussions, information and technology are analogized to the grasses in Hardin's pasture, as public knowledge is, by its nature, non-excludable. That is, once an item of information becomes generally known, it is difficult (absent the contractual or statutory structures described below) to prevent others from sharing it.⁴⁸

Information commons are, of course, different than aggregations of finite physical resources inasmuch as information is generally viewed as "non-rivalrous," meaning that any number of individuals may enjoy its benefits without depleting it: a fact, whether known by one person or by one million, remains undiminished.⁴⁹ As such, Hardin's "tragedy of the commons," which arises from self-interested over-exploitation of a common resource, is unlikely to occur within the context of information commons. However, as will be discussed below, other potential "tragedies" that are more specific to information commons have been postulated.⁵⁰

Building upon their earlier work on physical commons, Ostrom and Hess have applied the IAD framework to the analysis of knowledge-based commons structures, reasoning that both physical resource commons and information commons share numerous attributes.⁵¹ They caution, however, that further evolution of the IAD model may be required to understand the peculiar attributes of information commons more thoroughly.⁵² Michael Madison, Brett Frischmann, and Katherine Strandburg recently accepted this challenge and have undertaken a thorough re-examination of the IAD

demonstrated by the model of the Linux operating system); James Boyle, *The Second Enclosure Movement and the Construction of the Public Domain*, 66 LAW & CONTEMP. PROBS. 33, 44–49 (2003) (discussing open source software).

48. See, e.g., Hess & Ostrom, *supra* note 37, at 8–9. In much of the political and economics literature on commons, a distinction is drawn between "public goods" such as sunsets, which are owned by no one but which may be enjoyed by all and "common pool resources," such as highways and public libraries, which may be owned by a government or collective, but which are open to enjoyment by all. See Hess & Ostrom, *supra* note 38, at 119–21. In these terms, this Article addresses common-pool resources as opposed to naturally occurring public goods.

49. See, e.g., Dana Dalrymple, *Scientific Knowledge as a Global Public Good: Contributions to Innovation and the Economy*, in NRC, PUBLIC DOMAIN, *supra* note 7, at 35, 48 (noting that scientific knowledge, in its pure form, is "the epitome of a global public good," as it is "freely available to all and is not diminished by use—indeed it may grow with use"); Hess & Ostrom, *supra* note 37, at 8–9; Boyle, *supra* note 47, at 41–42. But see David W. Opperbeck, *The Penguin's Genome, or Coase and Open Source Biotechnology*, 18 HARV. J.L. & TECH. 167, 209–10 (2004) (seeking to refute the characterization of information as non-rivalrous).

50. See *infra* Section III.B and accompanying notes.

51. Ostrom & Hess, *supra* note 18, at 43.

52. *Id.* at 68.

framework in relation to commons in the “cultural environment.”⁵³ In doing so, they recognized that, unlike the farmers and fishermen who exploit physical commons of natural resources, users of information commons not only *use* the common resource, but *produce* it as well.⁵⁴ This insight led them to propose a modified framework that more closely links the features of the common resource to its user–producers, as mediated through constructed “rules in use,” and which also seeks to combine the functionalist approach of the IAD framework with metaphorical and narrative accounts of commons formation.⁵⁵

B. THE TEMPORAL DIMENSION OF INFORMATION COMMONS

1. *Dynamic and Static Commons*

The classical physical common resource posited by Hardin and others is both shared and finite, meaning that, absent husbandry and control, the resource will be over-consumed and therefore decline and possibly disappear over time: if the pasture continues to be over-grazed, it will become depleted within a predictable timeframe.⁵⁶ Curiously, the discussion of information commons typically lacks this dynamic element. Whereas physical commons are often characterized specifically in terms of their alteration (usually diminution) over time, information commons are more frequently described in one state or another without much reference to dynamic alteration of the resource, and *quantitative* changes to the commons have not traditionally been viewed as analytically significant, except in the most general terms.⁵⁷ Certainly, close analysis of information commons in their static state is a necessary and important exercise, as the fundamental arguments regarding, for example, the effects of intellectual property protection on common resources can be considered time-independent. Moreover, information commons are different from physical resource commons in that the contents of information commons generally do not diminish with increased use and

53. Madison, Frischmann & Strandburg, *supra* note 23, at 659.

54. *Id.* at 681. In this respect, they echo the well-known principle that users of intellectual property are also its producers. See WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW* 13–14 (2003).

55. Madison, Frischmann & Strandburg, *supra* note 23, at 671, 681–83.

56. This is Hardin’s “tragedy of the commons.” See *supra* notes 38–39 and accompanying text. For a general discussion of the temporal depletion of physical resource commons, see Ostrom’s classic work *GOVERNING THE COMMONS*, *supra* note 20.

57. See, e.g., David Bollier, *Growth of the Commons Paradigm*, in *KNOWLEDGE AS A COMMONS*, *supra* note 18, at 27, 34 (“[M]any information commons exemplify what some commentators have called ‘the cornucopia of the commons,’ in which more value is created as more people use the resource . . . the operative principle is ‘the more the merrier.’”).

their variance over time could therefore be considered less noteworthy than that of physical resource commons. Nevertheless, as will be discussed in the next Section, information commons *do* change over time, and the analysis of these temporal changes is important to understanding and designing information commons.

2. *Knowledge, Rights, and Timing in Information Commons*

a) The Dynamic Character of Information Commons: An Example

Information commons, by their nature, are dynamic. Take, for example, the body of knowledge concerning Jupiter's moons. Before Galileo discovered Jupiter's four largest satellites in 1610, the accumulated body of knowledge concerning them was zero. No information at all existed, and thus there could be no information commons. When the noted astronomer published his discovery a few months after it was made, an item of public knowledge was created and the commons relating to Jupiter's moons was born. As telescopes improved and astronomers' understanding of the solar system advanced over the next four hundred years, this information commons expanded. Recently, the aptly named Galileo spacecraft has provided detailed images and measurements of the Jovian satellites and their atmospheres, resulting today in a body of public information concerning the moons of Jupiter that is quite large.⁵⁸

b) The Changing Bases of Knowledge and Rights in an Information Commons

The above example relating to Jupiter's moons is generalizable to most information commons: the pool of data constituting the commons may expand and contract over time.⁵⁹ For purposes of discussion, I will term the total pool of data constituting an information commons at any given time its

58. For a concise overview, see Joseph A. Burns, *The Four Hundred Years of Planetary Science Since Galileo and Kepler*, 466 NATURE 575 (2010).

59. Some information commons, of course, *are* relatively static. Take, for example, a compilation of the works of Shakespeare. This body of work, though subject to periodic minor additions and subtractions, and notwithstanding the occasional effort (by no less august personages than the justices of the U.S. Supreme Court, see Jess Bravin, *Justice Stevens Renders an Opinion on Who Wrote Shakespeare's Plays*, WALL ST. J., Apr. 18, 2009, at 1) to claim or refute the Bard's authorship of a particular work, has remained more or less constant for more than a hundred years. The same is less true of science commons, which, in nearly all fields, tend to change as new experiments are conducted, observations are made, and hypotheses are advanced, withdrawn, refuted, and validated.

“knowledge base.”⁶⁰ The size of the knowledge base and the rate at which it changes are related variables by which an information commons can be measured and evaluated.

Just as the pool of *data* within an information commons may expand and contract over time, so may the set of *rights* applicable to the information within the commons. That is, for a given commons, the nature and duration of the usage restrictions on each data element may evolve over time, and the aggregate pool of usable data within the commons will likewise change.⁶¹ For purposes of this discussion, I will term the portion of the knowledge base of an information commons, the use of which is materially encumbered,⁶² as its “encumbered knowledge base.” Similarly, the portion of its knowledge base,

60. It is not uncommon to refer to this total pool of data as the information commons itself. However, for purposes of this analysis, it is helpful to distinguish between (a) the information commons, which includes its knowledge base, its “exogenous” characteristics, and the rules and policies governing its constitution, and (b) the size of the knowledge base associated with this commons at distinct periods in time.

61. There is a tendency in the discourse of commons to equate, or at least strongly associate, information commons with the public domain. *See, e.g.*, Anupam Chander & Madhavi Sunder, *The Romance of the Public Domain*, 92 CALIF. L. REV. 1331, 1338 (2004) (explicitly equating “commons” with the public domain, both generally and specifically in the context of indigenous genetic resources and traditional knowledge). This linkage has some appeal, as enjoying the benefits of a traditional commons (e.g., an open pasture) can be characterized by having the liberty to avail oneself fully of the benefits afforded by the resources within the “commons” (e.g., allowing one’s cattle to consume as much grass as they wish). The analogy between traditional commons and information commons, however, is imperfect. Unlike eating grass, catching fish, or timbering a forest, the “consumption” of resources in an information commons may occur in many different modes. For example, a digital music file in a “music commons” may be listened to, it may be copied, it may be transmitted, it may be mixed with other works, it may be synchronized with video footage, and it may be publicly performed, to name just a few potential modes of consumption. Each of these uses implicates distinct rights of the rights owner, and it is not necessary that a user be permitted to exercise every one of these rights in order to consider the music file part of a “commons.” Thus, for purposes of this Article, I proceed on the basis that an “information commons” may include knowledge that is covered by intellectual property rights, the use of which is thereby encumbered. *See* Boyle, *supra* note 47, at 68 (recognizing multiple “public domains”); Hess & Ostrom, *supra* note 38, at 121–22 (addressing the confusion between common-property and open-access regimes); J.H. Reichman & Paul F. Uhler, *A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment*, 66 LAW & CONTEMP. PROBS. 315, at 318–19 n.10 (2003) (distinguishing between “public domain” and “open access”). Madison, Frischmann, and Strandburg refer to structures including both public domain and encumbered knowledge as “semi-commons.” Madison, Frischmann & Strandburg, *supra* note 23, at 668.

62. By “materially encumbered” I mean that one or more material restrictions on the use of the data exist. These might include a contractual or policy embargo on presentation or publication of further results based on that data. At the extreme end of the temporal spectrum, patent rights can be viewed as another variety of encumbrance.

the use of which is not materially encumbered, or which is generally accessible and usable by a relevant public community,⁶³ as its “unencumbered knowledge base.”

Accordingly, each information commons may be characterized both in terms of its overall knowledge base and its unencumbered knowledge base (together with its complementary encumbered knowledge base). The size of an information commons’ overall knowledge base may or may not be closely correlated with the size of its unencumbered knowledge base. Thus, there may be large bodies of publicly accessible information (large knowledge base) that are almost entirely encumbered (small unencumbered knowledge base). One such example would be the database of knowledge contained in issued U.S. patents. It contains much information, but little freedom to use it, at least in the near-term.⁶⁴ Other information commons, such as that relating to the moons of Jupiter, may contain less information, but may have very few limitations on use. Thus, each information commons has varying complexions of knowledge and rights. And, by extension, as the character of an information commons changes over time, the relationship between these two variables—knowledge and rights—also changes, providing the basis for the latency analysis presented in the next Section.

3. *A Proposed Latency Analysis for Information Commons*

a) Knowledge Latency

For any given data element that is intended for inclusion in an information commons, there will be a period of time between its creation and its entry into the commons, when it becomes accessible to the relevant community. I term this time period “knowledge latency,” signifying that there is a delay, ranging from quite short to quite long, before which a particular data element destined for the commons actually appears there.⁶⁵

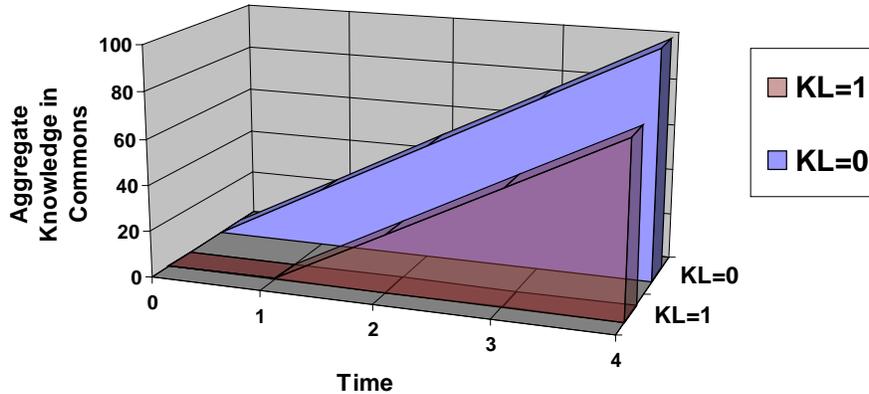
63. *See supra* note 6. Just as a traditional commons, such as a pasture or a forest, need not be accessible to the entire world, but only to the farmers and residents in the immediate area, I do not impose a requirement that information commons be accessible to the entire world. Rather, so long as the relevant community, be it of scientists, policy makers, or other interested users, has access to the commons, I view this as sufficient to qualify the body of information as accessible.

64. I have deliberately limited this example to “issued” patents. Of course, following expiration, the information contained in a patent is freely usable, as discussed *infra* notes 141–43 and accompanying text. I also assume, for purposes of this illustration, that the majority of inventive disclosures contained within a patent are protected by the patent claims and thus may not be used without infringement of these claims.

65. In information commons governed by policies with data deposit/release requirements, the maximum knowledge latency for the commons may be expressed as a

Knowledge latency therefore is an indicator of the rate of growth of the commons⁶⁶ and the amount of information available within the commons at any given time (Figure 1).

Figure 1: Effect of Knowledge Latency (KL) in an Information Commons⁶⁷



Knowledge latency in a given information commons may be expressed either as a *mandated* value derived from “rules in use” of the commons, or as an *actual* value. The *actual* value for knowledge latency may deviate from the *mandated* value for a number of reasons, including technical variations in data-deposit practices and intentional or inadvertent non-compliance by data generators. As with any set of policy-imposed timing requirements (e.g., time periods for making filings with governmental agencies), it is important to consider the mandated time delay for the deposit of data to an information commons. Because a mandated value is also, theoretically, the maximum amount of time that should elapse before a data element is deposited in the

single numerical value. In commons having data deposit requirements that are not quantified, such as those requiring that data be deposited as rapidly as possible following verification or based on other non-numerical criteria, knowledge latency may be expressed as a value range or estimated value, based on reasonable expectations and practice in the relevant field.

66. Note that knowledge latency shifts (delays) the time at which knowledge enters the commons, but not the *rate* at which knowledge enters the commons.

67. Figure 1 illustrates the aggregate Knowledge Base of an information commons over time, as determined by its Knowledge Latency. In the commons with KL=0, the Knowledge Base begins to grow at Time=0, whereas in the commons with KL=1, the Knowledge Base does not begin to grow until Time=1, resulting in a smaller Knowledge base at each subsequent time point.

commons, I generally refer to knowledge latency in terms of its *maximum* value.⁶⁸

b) Rights Latency

Just as there may be a delay between the generation of a data element and its deposit into an information commons, there may also be a delay between the appearance of data in the commons and its free usability, that is, its entry into the unencumbered knowledge base. I term this delay “rights latency.” As with knowledge latency, the term may be applied to an individual data element (i.e., representing the time before a particular data element becomes freely usable) or to the commons as a whole (i.e., representing the *maximum* time that it will take for data within the commons to become freely usable). The rights latency for a particular information commons may reflect a variety of factors, including policy-imposed embargos on the use of data and, in the extreme case, patent rights.⁶⁹ True public domain commons such as a compendium of Shakespeare’s works, in which no copyright or contractual encumbrances exist, would have rights latencies of zero. Commons that include data covered by patents would have rights latencies equal to the remaining patent term. Most information commons would fall somewhere between these two extremes.

c) Latency Analysis

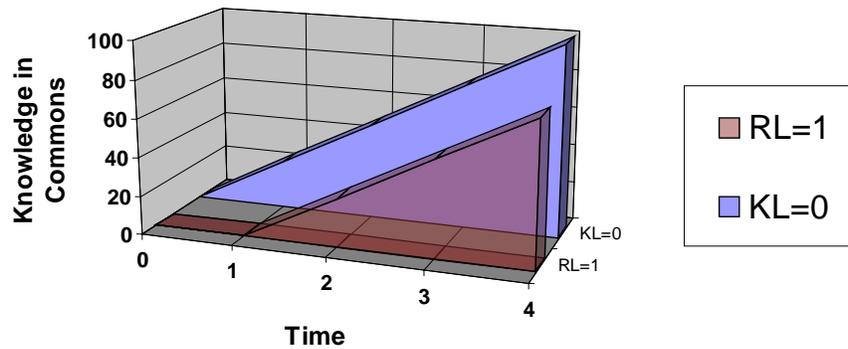
The related variables knowledge latency and rights latency describe important and complementary aspects of an information commons. If, as illustrated in Figure 2, rights latency is other than zero, the quantity of usable information in the commons will lag behind the total knowledge base.⁷⁰

68. Actual or observed knowledge latency values (both in terms of median values and the range of observed values) may be useful, however, in retrospective evaluation of user compliance with policy requirements or in assessing the efficiency of commons systems, and future analyses may find it useful to approach knowledge latency in this manner.

69. It may seem counter-intuitive to treat contractual limitations on use in a manner comparable to patent-based limitations on commercial exploitation. However, both of these “restrictions” serve to remove an item of information from public use. *See, e.g.*, Boyle, *supra* note 47, at 37. In each case, a user of the commons is aware of the restricted information and the knowledge that he or she obtains adds to his or her general store of knowledge, informs his or her view of the field, and apprises him or her of the developments and work of the data generator. In each case, however, the data user is unable to derive a direct benefit from the encumbered data, be that benefit a presentation or publication, or the development of a commercial product.

70. For purposes of this discussion, I assume that once established for a given information commons, knowledge latency and rights latency do not change substantially. This assumption, however, will not always hold true: commons “rules in use” may be

Figure 2: Effect of Knowledge Latency (KL) and Rights Latency (RL) in an Information Commons

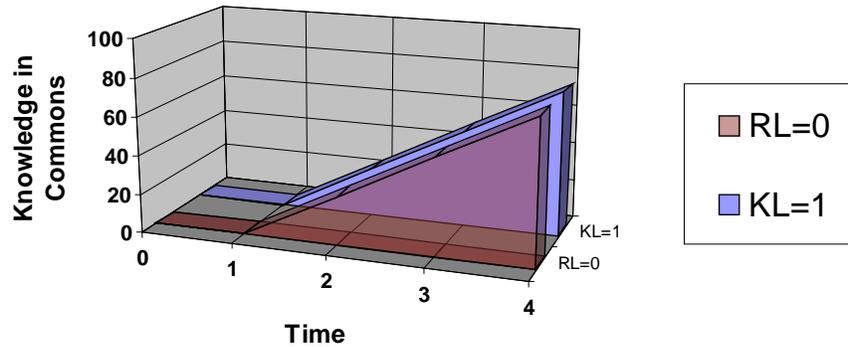


In the example illustrated by Figure 2, the rules of the commons establish that information be deposited into the commons as soon as it is created (KL=0), but a restriction, contractual or otherwise, limits its use until the passage of one time interval (RL=1). The line KL=0 shows the hypothetical increase of total knowledge within the commons beginning at the initial time (as there is no delay due to knowledge latency), and the area beneath line KL=0 represents the total knowledge within the commons. The line RL=1 shows the increase of unencumbered knowledge within the commons, and the area beneath the line RL=1 shows the total unencumbered knowledge within the commons.

An alternative scenario is shown in Figure 3. In this example, knowledge is withheld from the commons for one time interval after it is created (KL=1), but is released to the commons without encumbrance on future use (RL=0). The lines KL=1 and RL=0 are overlaid, resulting in a quantity of *unencumbered* knowledge that is equivalent to that illustrated in Figure 1, but less overall knowledge at any given time than the commons in Figure 2. The practical implications of these different design choices are illustrated by the case study in Section IV.A.

amended following the initial commons formation, and changes in law and policy may also impact the knowledge latency and rights latency of an information commons.

Figure 3: Effect of Knowledge Latency (KL) and Rights Latency (RL) in an Information Commons



d) Implementing Latency Analysis in Commons Design

As demonstrated by the above examples, latency variables are important elements in the analysis of information commons. However, the importance of these variables goes beyond their mere descriptive value. As discussed in Part V, the adjustment of rules that govern the latency characteristics of an information commons can enable policy designers to achieve varying policy goals. Moreover, the adjustment of latency variables can effectively mediate between the requirements of competing stakeholder interests and enable the creation of commons where disagreement might otherwise have precluded it.

In this Article, I propose a straightforward methodological approach for applying latency analysis to the evaluation and design of information commons. This approach consists of three steps which may be integrated into the frameworks of Ostrom and Hess,⁷¹ and Madison, Frischmann, and Strandburg⁷²: (1) determine the relevant communities of stakeholders of the information commons and their characteristics,⁷³ (2) determine the policy considerations relevant to each stakeholder group and the initial positions of each such group,⁷⁴ and (3) adjust the latency variables of the commons

71. See Ostrom & Hess, *supra* note 18, at 48–50.

72. Under both formulations, “rules in use” dictating these latencies will have an effect on the nature of the common resource. See Madison, Frischmann & Strandburg, *supra* note 23, at 680–82.

73. See Ostrom & Hess, *supra* note 18, at 48–50 (defining the “attributes of the community”); Madison, Frischmann & Strandburg, *supra* note 23, at 690 (prescribing the identification of constituencies and their part in commons formation).

74. See Ostrom & Hess, *supra* note 18, at 48–50 (considering goals and objectives among the attributes of the community); Madison, Frischmann & Strandburg, *supra* note 23, at 691–93 (identifying community goals and objectives).

(knowledge latency and rights latency) to mediate among the differing positions of the stakeholder groups.

III. COMMONS IN THE SCIENCES

The corpus of public scientific information has long been viewed as the quintessential information commons.⁷⁵ In this Part, I describe the norms and practices that have shaped the development of science commons, identify the various stakeholders invested in the development and use of such commons, and address some of the challenges that affect the growth of science commons today. In particular, I focus on three sets of exogenous factors that impact the addition of useful information to the commons: publication delay, copyright and other access-limiting rules, and patent-based restrictions on usage of information in the commons. The factors described in this Part form the basis of the policy objectives and stakeholder goals that will be addressed in Part IV. These competing objectives can effectively be mediated through the adjustment of latency variables.

A. CONSTRUCTING SCIENCE COMMONS

1. *Incentives to Share*

In the 1940s, sociologist Robert K. Merton famously identified the willingness of scientists to share their findings and experimental data as one of the fundamental norms that characterize both the practice and culture of science.⁷⁶ This norm finds its origin, however, far earlier than Merton's day; it is often traced to the late sixteenth and early seventeenth centuries—the age of Galileo, Bacon, and Newton. Prior to this time, specialized technical knowledge was typically guarded by trade guilds and individual practitioners who had little incentive to share with outsiders.⁷⁷ But beginning during what has traditionally been termed the “scientific revolution,”⁷⁸ new norms among scientists emerged that favored the open publication of ideas and the sharing of data. The causal factors that led to the adoption and spread of these “open science” norms are subject to debate among historians of science and are

75. See Benkler, *supra* note 4, at 1110.

76. See *supra* note 2 and accompanying text.

77. See Robert P. Merges, From Medieval Guilds to Open Source Software: Informal Norms, Appropriability Institutions, and Innovation 6–7 (Nov. 13, 2004), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=661543.

78. This term has been criticized by historians of science who contend that the emergence of modern science arose from a gradual process rather than a sudden “revolution.” See PETER J. BOWLER & IWAN RHYS MORUS, MAKING MODERN SCIENCE: A HISTORICAL SURVEY 23–25 (2005).

beyond the scope of this Article.⁷⁹ Nevertheless, the legacy of these significant social changes continues to influence the patterns and customs of scientific practice today.

Merton rationalized the sharing of scientific data on three functional bases. First, the advancement of science depends upon scientists' access to, and use of, previously collected data.⁸⁰ Little progress would be made if individual scientists were required to reproduce all the experiments of their predecessors. Thus, in order for science to advance, scientists must build upon earlier results in a cumulative fashion, resulting in an ever-expanding body of scientific knowledge that is available and accessible to the scientific community at large.⁸¹ Second, scientific data must be shared to enable scientists to validate and independently verify the analyses and conclusions of their colleagues.⁸² Recent instances of scientific fraud and the perception that misconduct among scientists is increasing underscore the need for critical

79. Paul David attributes the rise of "open science" to the intersection of late-Renaissance aristocratic patronage with the new class of experimental philosophers. Paul David, *Common Agency Contracting and the Emergence of "Open Science" Institutions*, 88 AM. ECON. REV. 15 (1998). Peter Bowler and Iwan Rhys Morus associate the emergence of modern scientific institutions and attitudes with the rise of universities. BOWLER & MORUS, *supra* note 78, at 322–26. Adrian Johns points to the success of the first scientific journal, the *Philosophical Transactions*, as indicative of the new style of collaborative experimental philosophy. ADRIAN JOHNS, *PIRACY: THE INTELLECTUAL PROPERTY WARS FROM GUTENBERG TO GATES* 62–63 (2009). And Elizabeth Eisenstein credits the printing press with the rise of scientific data sharing. ELIZABETH L. EISENSTEIN, *THE PRINTING PRESS AS AN AGENT OF CHANGE* 520–21 (1979).

80. Merton, *supra* note 5, at 274–75.

81. In this respect, Merton cites the well-known remark attributed to Sir Isaac Newton, "If I have seen farther it is by standing on the shoulders of giants." Merton, *supra* note 5, at 274–75. Merton's sentiment is echoed by contemporary analysts. See, e.g., JOINT DATA STANDARDS STUDY, *supra* note 8, at 11 ("[D]ata sharing contributes to a virtuous circle, where promoting effective sharing widens research and enhances scientific impact."). Today, the positive effects of data sharing are manifested both in the elimination of duplicative experiments and in the creation of large data sets that could not be created by individual scientists or groups. See Mike May, *Sharing the Wealth of Data*, SCIENTIFIC AMERICAN — WORLDVIEW, 88, 89–90 (2009).

82. See also NAT'L ACAD. OF SCIS., *ENSURING THE INTEGRITY, ACCESSIBILITY, AND STEWARDSHIP OF RESEARCH DATA IN THE DIGITAL AGE* 55 (2009) [hereinafter NAS, RESEARCH DATA] ("Only when a researcher shares data and results with other researchers can the accuracy of the data, analyses, and conclusions be verified."); Paul David, *The Economic Logic of "Open Science" and the Balance Between Private Property Rights and the Public Domain in Scientific Data and Information: A Primer*, in NRC, *PUBLIC DOMAIN*, *supra* note 7, at 21 ("Disclosure . . . creates an expectation that all claims to have contributed to the stock of reliable knowledge will be subjected to trials of verification, without insult to the claimant.").

and independent review of scientific claims.⁸³ Similar considerations have also contributed to calls for greater public access to research data supporting federal regulation.⁸⁴ Third and finally, scientists must share data in order to inform the scientific community of their own discoveries, accomplishments, and breakthroughs. This public notice serves to reward individual scientists through recognition and the esteem of their colleagues. Recognition has both intangible psychic benefits as well as tangible benefits, such as tenure advancement for junior investigators, increasingly prestigious appointments, endowed chairs, honorific positions for tenured investigators,⁸⁵ and perhaps most importantly in challenging economic times, scarce government grant funding.⁸⁶ According to the functionalist view, these three factors work in

83. See Dov Greenbaum, *Research Fraud: Methods for Dealing with an Issue that Negatively Impacts Society's View of Science*, 10 COLUM. SCI. & TECH. L. REV. 61, 75–77 (2009); Jocelyn Kaiser, *Data Integrity Report Sends Journals Back to the Drawing Board*, 325 SCI. 381 (2009) (citing several examples of scientific fraud, including a notoriously falsified South Korean stem cell paper in 2005, that prompted the commissioning of the NAS study and report cited *supra* note 82). The importance of making research data available for validation purposes was recently illustrated by the case of Dr. Marc Hauser at Harvard, whose conclusions regarding the behavior of tamarin monkeys were refuted by experts once they reviewed Hauser's original videotapes. Nicholas Wade, *In Harvard Lab Inquiry, a Raid and 3-Year Wait*, N.Y. TIMES, Aug. 13, 2010, at A12.

84. Omnibus Appropriations Act of 1999, Pub. L. 105-277, 112 Stat. 2681 (Data Access Act of 1999); Consolidated Appropriations Act of 2001, Pub. L. 106-554, § 515, 114 Stat. 2763, 2763A-154 (codified at 44 U.S.C. § 3516) (Data Quality Act of 2001). Despite the seemingly innocuous intent of this legislation, it has been criticized sharply for its potential to assist industry in challenging science-based governmental regulation. See, e.g., CHRIS MOONEY, *THE REPUBLICAN WAR ON SCIENCE* 102–20 (2005); Rick Weiss, *Data Quality Law Is Nemesis of Regulation*, WASH. POST, Aug. 16, 2004, at A1. See generally NAS, RESEARCH DATA, *supra* note 82, at 69.

85. Merton, *supra* note 5, at 274–75; see also NAS, RESEARCH DATA, *supra* note 82, at 55 (“Researchers receive intellectual credit for their work and recognition from their peers . . . when they publish their results and share the data on which those results are based.”); David, *supra* note 82, at 22.

86. In 2005, the National Institutes of Health (“NIH”), a major source of federal research funding, required that applicants have “demonstrated an ongoing record of accomplishments that have advanced their field(s).” Press Release, Nat’l Insts. of Health, *Enhancing Peer Review: The NIH Announces Enhanced Review Criteria for Evaluation of Research Applications Received for Potential FY2010 Funding*, Notice Number: NOT-OD-09-025 (Dec. 2, 2008), available at <http://grants.nih.gov/grants/guide/notice-files/not-od-09-025.html>. Commenting on the alterations to the NIH grant review process, a chairperson of an NIH grant review section noted that “applicants with robust publication histories, [and] proven track records of scientific accomplishment . . . may have the edge over their younger, less experienced counterparts.” Bob Grant, *New NIH Forms Raise Concerns*, THE SCIENTIST.COM (Dec. 8, 2009, 3:49 PM), <http://www.the-scientist.com/blog/display/56209/>. It should be noted, however, that the NIH does offer some types of grants specifically for less experienced, and thus less published, investigators. Nat’l Insts. of

concert, rewarding the individual scientist for sharing his or her data and yielding an overall benefit to the scientific community and society at large. As observed by the National Research Council, “the act of publishing [scientific findings] is a *quid pro quo* in which authors receive credit and acknowledgement in exchange for disclosure of their scientific findings.”⁸⁷

Recent commentators have questioned whether Merton’s norms accurately reflect the manner in which science is practiced today, arguing that competition and self-interest may motivate scientists far more than the larger social factors cited by Merton.⁸⁸ Patrick Taylor, who argues for broad data sharing in the biosciences on ethical and pragmatic grounds, worries that “research sharing is an aspiration vulnerable to compromise when it interferes with private incentives, whether researchers’ academic self-interest or potent market forces.”⁸⁹ Others have sought to explain scientific data sharing in economic terms by reference to the non-monetary incentives valued and earned by scientists (as individual economic actors) engaged in academic research.⁹⁰ Madison, Frischmann, and Strandburg, building on Ostrom’s institutional-choice foundation and functionalist intellectual property theory, posit that information commons “arise as solutions to collective action, coordination or transaction cost problems.”⁹¹

Health, *New and Early Stage Investigator Policies*, U.S. DEP’T OF HEALTH & HUMAN SERVS., http://grants.nih.gov/grants/new_investigators/ (last visited Sept. 17, 2010).

87. NAT’L RESEARCH COUNCIL, *SHARING PUBLICATION-RELATED DATA AND MATERIALS: RESPONSIBILITIES OF AUTHORSHIP IN THE LIFE SCIENCES* 34 (2003) [hereinafter NRC, *SHARING PUBLICATION-RELATED DATA*].

88. *See generally* HARRY COLLINS & TREVOR PINCH, *THE GOLEM: WHAT EVERYONE SHOULD KNOW ABOUT SCIENCE* (1993) (describing self-interested competition among scientists); IAN I. MITROFF, *THE SUBJECTIVE SIDE OF SCIENCE* (1974); Ian I. Mitroff, *Norms and Counter-norms in a Select Group of the Apollo Moon Scientists: A Case Study in the Ambivalence of Scientists*, 39 AM. SOC. REV. 579 (1974); Opderbeck, *supra* note 49, at 189–90 (summarizing various criticisms of Merton’s conceptualization of scientific norms). To his credit, Merton was not unaware of the intense competition and occasional skulduggery that characterize the scientific enterprise. In fact, he wrote extensively about these competitive factors, analyzing scientific rivalry and priority disputes since the great conflict between Newton and Leibnitz. *See* Merton, *supra* note 5, at 287; Robert K. Merton, *Behavior Patterns of Scientists (1968)*, in *THE SOCIOLOGY OF SCIENCE*, *supra* note 5, at 325, 336–37. The fact that individual scientists may act for reasons of self-interest, or that certain scientists may violate accepted norms, does not itself invalidate the observation of broader community norms.

89. Patrick I. Taylor, *Research Sharing, Ethics and Public Benefit*, 25 NATURE BIOTECHNOLOGY 398, 399 (2003).

90. *See, e.g.*, Benkler, *supra* note 47, at 433–35; Peter Lee, *Toward a Distributive Commons in Patent Law*, 2009 WIS. L. REV. 917, 942.

91. Madison, Frischmann & Strandburg, *supra* note 23, at 691–92.

Another factor that has influenced the sharing of scientific data and the creation of science commons in recent decades is governmental policy. The federal government is by far the largest funder of basic scientific research in the United States.⁹² As such, federal policy plays a significant role in determining the manner in which scientific research is conducted. In general, current federal policy promotes the sharing and open accessibility of scientific data through a variety of contractual and grant-based mechanisms.⁹³ Federal policy thus reinforces and institutionalizes existing norms of scientific data sharing.⁹⁴

2. *Modes of Data Sharing*

The most typical and widespread means of disseminating results in the sciences is, and has been for centuries, publication in peer-reviewed scientific journals.⁹⁵ The quantity of a scientist's publications is one of the most important factors used in assessing the quality of his or her research, advancing his or her career, and determining his or her stature within the scientific community.⁹⁶ Thus, the scientist has a significant personal incentive

92. In 2008–2010, the annual federal budget for basic scientific research was approximately \$30 billion and represented approximately 60% of total basic research spending. JOHN F. SARGENT, JR., CONG. RESEARCH SERV., R40710, FEDERAL RESEARCH AND DEVELOPMENT FUNDING: FY2010, at 10 (2010), available at <http://www.fas.org/sgp/crs/misc/R40710.pdf>. Total federal spending on R&D, including basic and applied research, development and facilities, exceeded \$140 billion per year in 2008–2010. NAT'L SCI. FOUND., FEDERAL R&D FUNDING BY BUDGET FUNCTION: FISCAL YEARS 2008–2010 (2010), available at <http://www.nsf.gov/statistics/nsf10317/pdf/nsf10317.pdf>.

93. See Lee, *supra* note 90, at 941–42; Reichman & Uhler, *supra* note 61, at 331–32 (citing National Science Foundation data sharing requirements). Federal policy relating to genomic research is discussed in greater detail *infra* Section IV.A.

94. See Reichman & Uhler, *supra* note 61, at 332. Peter Lee also envisions governmental science policy as a means to achieve the ends of distributive justice, particularly in providing disadvantaged populations with access to health-promoting technologies. Lee, *supra* note 90.

95. See ZIMAN, *supra* note 4, at 39 (arguing that the peer-reviewed publication process is “at the very core of academic science” and “inseparable from its other functions”); Merton, *supra* note 88, at 337 (“From its very beginning, the journal of science introduced the institutional device of quick publication to motivate men of science to replace the value set upon secrecy with the value placed upon the open disclosure of the knowledge they had created.”).

96. See, e.g., Robert K. Merton, *Priorities in Scientific Discovery (1957)*, in *THE SOCIOLOGY OF SCIENCE*, *supra* note 5, at 286, 316 (noting the “tendency, in many academic institutions, to transform the sheer number of publications into a ritualized measure of scientific or scholarly accomplishment”); Editorial, *When Blogs Make Sense*, 466 NATURE 8 (2010) (“[P]eer-reviewed publications continue to be the primary markers of academic achievement.”).

to publish his or her results as quickly as possible⁹⁷ in as “high-impact” a journal as possible.⁹⁸ While in some disciplines researchers have begun to share scientific articles prior to publication through trusted “preprint” systems such as the arXiv.org server, publication in peer-reviewed journals remains an important and necessary step for recognition and validation.⁹⁹

The results and data reported in scientific journal articles, however, must be distinguished from the much larger quantity of experimental and observational data generated in the course of research and upon which published results are based.¹⁰⁰ A journal article typically includes a brief presentation of significant experimental findings, often made in summary or tabular fashion, together with the scientist’s analysis and conclusions based upon those findings.¹⁰¹ While the published data is usually essential to

97. A scientist’s incentive to publish quickly arises not only from a desire to advance his or her career, but also from a genuine fear that, if work is not published as soon as possible, competing groups may develop the same or similar results and “scoop” one’s work. See, e.g., NRC, SHARING PUBLICATION-RELATED DATA, *supra* note 87, at 28; Elizabeth Pennisi, *Genomics Researchers Upset by Rivals’ Publicity*, 329 SCI. 1585 (2010) (citing incidents of “scooping” in publishing sequencing results for various non-human genomes); Robert K. Merton, *Making It Scientifically*, N.Y. TIMES BOOK REV., Feb. 25, 1968, at 1 (reviewing James Watson’s *The Double Helix* and noting that “multiple independent discoveries are one of [the] occupational hazards” of science); *supra* note 3 and accompanying discussion.

98. See RESEARCH INFO. NETWORK, TO SHARE OR NOT TO SHARE: PUBLICATION AND QUALITY ASSURANCE OF RESEARCH DATA OUTPUTS 25 (2008), available at <http://www.rin.ac.uk/data-publication/> (noting that the assessment of researchers is “perceived to value above all else the publication of papers in high-impact journals”); ZIMAN, *supra* note 4, at 180 (noting that, in terms of scientific success, “[o]ne paper with a hundred favourable citations is worth infinitely more than a hundred papers with one citation each”).

99. See Editorial, *supra* note 96 (noting the popularity of preprint sharing of articles among astronomers, and its unpopularity among biologists); Donald Siegel & Philippe Baveye, Commentary, *Battling the Paper Glut*, 329 SCI. 1466 (2010) (criticizing the current importance placed on quantity of publications).

100. See Nelson, *supra* note 4, at 299.

101. See generally Rebecca S. Eisenberg, *Patents and Data-Sharing in Public Science*, 15 INDUS. & CORP. CHANGE 1013, 1024 (2006). By way of example, one recently published study identifies the *fgf4* gene as a factor leading to short-leggedness in dogs such as the Welsh corgi and the dachshund. Heidi G. Parker, *An Expressed Fgf4 Retrogene Is Associated with Breed-Defining Chondrodysplasia in Domestic Dogs*, 325 SCI. 995 (2009). The association of *fgf4* with the physical or “phenotypic” trait of short-leggedness is an experimental *result*. A vast quantity of *data* had to be collected and generated in order to arrive at this result, including raw genetic sequence reads for numerous dogs across different breeds, associated phenotypic data for each of the subjects, and a complex of statistical analyses, associations and computations.

It is also worth noting the distinction drawn by behavioral scientists among the terms *data*, *information*, and *knowledge*. Under the hierarchy developed by Fritz Machlup in 1983, “data” are individual facts, measurements, and assumptions (encompassing both “data” and “results,” as I have defined them), “information” is data that has been organized

support the scientist's analysis, the data reported in a journal article seldom represents the entirety of the "raw" data collected or observed by the scientist, and is typically only a small fraction of the full data set.¹⁰²

Traditionally, a scientist who wished to inspect another scientist's raw data, whether to validate the experimenter's results or to build upon those results, had only a few informal options for obtaining access to this data. Sometimes data would be presented at scientific conferences or symposia, though these venues typically provided the same (or less) data than would otherwise be presented in a published paper, albeit sooner. Many requests for experimental data were, and are still, made directly by one scientist to another by letter, telephone or, more recently, e-mail.¹⁰³ Such informal requests have typically been fulfilled, if at all, on the basis of friendship or professional courtesy, conditioned on the availability, work schedule, inclination, and discretion of the scientist in possession of the data.¹⁰⁴

Today, powerful electronic databases and high-speed networks that enable the dissemination of scientific data on an unprecedented scale supplement these informal and inefficient methods of data sharing. In the

and contextualized, and "knowledge" is information that has been processed and assimilated by a human. See DONALD O. CASE, LOOKING FOR INFORMATION: A SURVEY OF RESEARCH ON INFORMATION SEEKING, NEEDS, AND BEHAVIOR 64 (2d ed. 2007); Charlotte Hess & Elinor Ostrom, *Introduction: An Overview of the Knowledge Commons*, in KNOWLEDGE AS A COMMONS, *supra* note 18, at 3, 8. Due to the distinction I draw between "data" and "results," and because the presentation of scientific data regularly includes some analysis and synthesis, I do not adhere strictly to the Machlup hierarchy in my terminology, but rather use the term "information" to connote both "data" and "results," whether as individual data points or in an aggregated, processed form.

102. For example, the full genomic sequence of an organism might require hundreds or thousands of pages to print, whereas most journal articles are in the range of ten or fewer printed pages.

103. See Reichman & Uhler, *supra* note 61, at 343–48 (discussing the characteristics of informal data sharing arrangements between scientists).

104. See *id.* at 405 ("[R]esearchers . . . tend to accommodate requests to share data in response to community norms, peer pressure, the expectation of reciprocity, and other factors shaped by perceived self-interest."). There is a growing body of empirical evidence demonstrating, however, that these requests for data sharing among scientists are often ignored or refused. See, e.g., Blumenthal et al., *National Survey*, *supra* note 17 (reporting that 8.9% of academic life scientists has refused to share research results with other scientists within the past three years); Blumenthal et al., *Prevalences and Predictors*, *supra* note 17 (concluding, on the basis of similar data to that presented in the authors' 2002 paper, that "data withholding is common in biomedical science"); Eric G. Campbell et al., *Data Withholding in Academic Genetics: Evidence from a National Survey*, 287 J. AM. MED. ASS'N 473 (2002) (reporting that 47% of geneticists who requested information relating to published research were denied at least once in the preceding three years, and 10% of all post-publication data results were denied); Editorial, *supra* note 3.

United States, the projects giving rise to these scientific commons have historically been “big science” initiatives in fields such as high-energy physics, astronomy, and geoscience, often using government facilities such as large telescopes, spacecraft, and particle accelerators.¹⁰⁵ More recently, however, scientific commons increasingly comprise data from projects that are generated by academic or research institutions funded in whole or in part by government grants.¹⁰⁶ In a typical arrangement of this nature, a government agency will fund these research centers to procure equipment and generate data in a coordinated or collaborative manner, either in fulfillment of a broader governmental program or as part of a research proposal made by the requesting institution.¹⁰⁷ The resulting data is often deposited in a government-operated database such as GenBank¹⁰⁸ and is thus made accessible to other scientists. This aggregation and availability of data in “science commons” enables the efficient, rapid, and cost-effective sharing of new knowledge and enables study and analysis that otherwise might have been impossible.¹⁰⁹

3. Stakeholders and Policy Considerations

In designing the “rules in use” for a science commons, policy makers must consider the interests of many different stakeholder groups, which may be overlapping, divergent, and sometimes contradictory.¹¹⁰ The principal stakeholder constituencies relevant to science commons are: (1) scientists who generate and contribute data to the commons (“data generators”); (2) academic or corporate institutions that employ these scientists;¹¹¹ (3)

105. See Reichman & Uhler, *supra* note 61, at 322.

106. Paul N. Schofield, et al., *Sustaining the Data and Bioresource Commons*, 330 SCI. 592, 592 (2010).

107. *Id.*

108. GenBank is operated by the National Library of Medicine of the National Institutes of Health.

109. See Eisenberg, *supra* note 101, at 1020. The system is not, however, without its flaws. See Keith Baggerly, Commentary, *Disclose All Data in Publications*, 467 NATURE 401 (2010) (noting that publicly available data is often insufficient to enable the reproduction of experiments in the field of genetics).

110. Both Ostrom and Hess, and Madison, Frischmann, and Strandburg emphasize the importance of identifying the various constituencies connected with a cultural commons. Madison, Frischmann & Strandburg, *supra* note 23, at 690; Ostrom & Hess, *supra* note 18, at 48–50.

111. While research scientists are often employees of their universities or other research institutions, individual scientists often exhibit different motivations and goals than their employer organizations. See Jennifer Carter-Johnson, *Unveiling the Distinction Between the University and Its Academic Researchers: Lessons for Patent Infringement and University Technology Transfer*, 12 VAND. J. ENT. & TECH. L. 473 (2010).

scientists who access and use the common data, and who in many cases are themselves data generators (“data users”); (4) academic or corporate institutions that employ the data users; (5) distributors and other intermediary handlers of the common data such as database operators, scientific journals, or in some cases, the data generators themselves; (6) funders of the research generating the data, such as government agencies, private foundations, or industry;¹¹² and (7) members of the general public, in their dual capacities as taxpayers who ultimately fund scientific research and as consumers of the products that eventually result from that research (and who, in the case of patient advocacy groups, may be highly conversant in the relevant scientific literature).¹¹³

Like the other “cultural commons” described by Madison, Frischmann, and Strandburg, science commons are constructs based on sets of rules.¹¹⁴ And as in other private ordering structures, the rules governing science commons can supplant baseline legal regimes such as intellectual property

112. See Lee, *supra* note 90, at 950–73 (addressing the interests and policy concerns of government and private funders of scientific research).

113. See *id.* at 986–90 (addressing the interests and policy concerns of disease advocacy groups); Lea Shaver, *The Right to Science and Culture*, 2010 WIS. L. REV. 121, 154–74 (proposing that public access to scientific and cultural information should be viewed as a basic human right within the framework of the Universal Declaration of Human Rights). Anupam Chander and Madhavi Sunder take issue with the fact that the “consumers” of information within scientific commons—and the genome commons, in particular—are “a self-selecting few who have chosen to participate in the new technology revolution.” Chander & Sunder, *supra* note 61, at 1343 (quoting Rodrigo Martinez et al., *The Geography of the Genome*, WIRED, June 2003, at 160). I do not suggest that the “commons” of scientific information is particularly valuable, or even comprehensible, to the majority of the world’s citizens. In many cases, the public interest in the broad dissemination of scientific information is represented by government or advocacy groups that possess the requisite qualifications to understand and use this data. This is not to say, however, that the public at large, which stands to benefit or lose based on the outcome of policy decisions, does not constitute an important “stakeholder” in the debate concerning the science commons.

114. Broadly speaking, information commons can arise organically or by design. Organic commons arise when informal groups share information about topics of common interest, as occurs in online communities that grow up around topics as diverse as cooking, sports, politics, music, open source software, childrearing, health, and online gaming. The information in these informal commons is generally contributed and used freely without restriction, though informal norms and collective behavioral expectations undeniably shape usage and production practices. See Benkler, *supra* note 47, at 381–82 (explaining how formal and informal norms structure collaboration on open source software development). Due to numerous organizational factors, including the capital-intensive nature of many data-generating scientific projects and the pervasiveness of governmental funding of scientific research, however, commons of scientific information are typically more structured.

protection to reflect the policy objectives of particular stakeholder groups.¹¹⁵ From the outset, it should be clear that the different stakeholder groups described above will have differing perspectives and incentives to act, and while each has a stake in the sharing of scientific data, no single group is motivated by all three of the Mertonian sharing incentives. Rather, each group responding to its own particular incentives and requirements ultimately results in the social sharing structures observed both by Merton and later functionalist scholars.

Data-generating scientists, for example, typically have a strong interest in publishing their work and obtaining grant funding to continue their research programs. The institutions that employ them, while concerned with the reputation and renown of their researchers, may have additional interests in obtaining patents or securing corporate funding (concerns that are often, but not always, shared by scientists, and not necessarily to the same degree). These interests may conflict with the interests of data users, who, as a gross generalization, typically wish to access and use generated data as soon as it is available with as little cost, inconvenience, and encumbrance as possible. Government funders may also seek rapid release of data to further the progress of science in the public interest, a concern that is often mirrored by patient advocacy groups. The conflicts that arise among these stakeholder groups are inevitable and will be addressed in greater detail in the case studies presented in Part IV, *infra*.

B. EXOGENOUS INFLUENCES ON LATENCIES WITHIN A SCIENCE COMMONS

As noted above, science commons are constructs developed through the application of rules. But privately ordered internal rule sets are not the only factors that influence the size and growth of science commons. In many cases, exogenous factors play a critical, if not deterministic, role in establishing the dynamic characteristics of a commons. In this Section, I examine three distinct exogenous factors that exert a substantial effect on knowledge and rights latency within a science commons¹¹⁶: (1) delays in the

115. See Peter Lee, *Contracting to Preserve Open Science: Consideration-Based Regulation in Patent Law*, 58 EMORY L.J. 889, 917 (2009) (discussing the increased control “shrinkwrap” licenses allow over content as compared to copyright and patent law).

116. There are, of course, numerous other factors that influence the design of information commons, particularly in the biomedical realm, including important considerations surrounding the protection of individual identity and health information and the proper means of obtaining informed consent. See, e.g., JOINT DATA STANDARDS STUDY, *supra* note 8, at 45–46; Kaye, *supra* note 36, at 333–35. While these considerations are critical

publication of data; (2) copyright and other legal constraints on the dissemination of data; and (3) patent encumbrances on the use of data.¹¹⁷

1. *Publication Delay*

The most common and regularized method of sharing scientific results is through publication in scientific journals.¹¹⁸ Typically, journals require that scientists provide sufficient detail regarding their methods and materials to permit a reader to reproduce and validate the experimental conclusions advanced in the article.¹¹⁹ In the past decade, however, an increasing number of scientific journals have required that authors make the underlying data supporting their published claims available to readers.¹²⁰ For example, in the case of genomic sequence data, journals often require deposit of the data into a public database at the time of publication.¹²¹ However, only a fraction of scientific journals have such requirements,¹²² and the requirements that do exist are often unclear about when, how and which data must be made available.¹²³ Thus, for most scientific papers, supporting data is published or

to the design and operation of information commons in the biomedical arena, they are relatively time-independent (i.e., subject privacy should be protected whenever data is released), and thus do not factor heavily into the latency analysis presented in this Article.

117. Interestingly, Madison, Frischmann, and Strandburg do not view science commons as particularly susceptible to encumbrance by intellectual property rights, suggesting that “commons of scientific research results and tools in the basic sciences” may be most appropriately characterized as “nonpropertized” realms that exhibit the characteristics of the “natural intellectual environment.” Madison, Frischmann & Strandburg, *supra* note 23, at 686. In this Part, I argue that intellectual property rights exert an extreme influence on science commons.

118. *See supra* Section III.A.2.

119. NRC, SHARING PUBLICATION-RELATED DATA, *supra* note 87, at 32.

120. *Id.* at 32–33; *see also* GUIDE TO PUBLICATION POLICIES OF THE NATURE JOURNALS (Apr. 30, 2009), *available at* <http://www.nature.com/authors/gta.pdf> (“[A] condition of publication in a Nature journal is that authors are required to make materials, data and associated protocols promptly available to readers without preconditions.” (emphasis removed)).

121. *See* Stephen Hilgartner, *Potential Effects of a Diminishing Public Domain in Biomedical Research Data*, in NRC, PUBLIC DOMAIN, *supra* note 7, at 137.

122. *See* NRC, SHARING PUBLICATION-RELATED DATA, *supra* note 87, at 33 (according to a 2003 study, of the fifty-six most frequently cited life-science and medical journals, only 41% had policies expressly requiring the deposit of data supporting published articles).

123. A journal’s data sharing requirement may be based on whether the data is *central* to the published results, *integral* to the published results, or simply useful *background* information. *See* NRC, SHARING PUBLICATION-RELATED DATA, *supra* note 87, at 36 (describing the “uniform principle for sharing integral data and materials expeditiously (UPSIDE)”).

made publicly available, if at all, no earlier than the time of publication.¹²⁴ The result, in most cases, is that a significant period of time elapses between the time that experimental data is generated and the time that it becomes publicly available. This delay may occur during one or more of the myriad steps in publishing results. Investigators must analyze their results, gather additional data, refine their analysis, prepare a paper based on their findings, and submit the paper to journals. Journals must conduct peer review and the editorial process, and wait for investigators to make any revisions required by the journals, including, at times, conducting additional experiments. If a journal rejects the paper, investigators must revise and submit it to different journals. Finally, the journal must edit, format, and prepare accepted papers for publication. One recent study reports that the period from completion of scientific work until publication is typically between twelve and eighteen months.¹²⁵ Older studies have found comparable or longer publication delay periods in other fields of research;¹²⁶ the seventeen-year delay in the case of the *Ardi* researchers represents an extreme example.¹²⁷

In terms of the latency analysis described in Section II.B.3, publication delay lengthens knowledge latency, often by years. The effect of this delay is complex. Though publication delay may hinder the overall progress of science (i.e., retarding other scientists' ability to build upon an investigator's findings), it does not slow the work of the investigators who initially collected the data—one of the criticisms leveled against the *Ardi* researchers. They, unlike their peers who lack access to this data during the publication process, are free to continue to analyze, refine, and conduct follow-up studies based on the data. Thus, even in an environment in which an ethic of data sharing has been widely accepted, this publication-based delay provides the original investigator with a substantial head-start in the analysis and

124. See, e.g., Reichman & Uhler, *supra* note 61, at 335; Toronto Int'l Data Release Workshop Authors, *Prepublication Data Sharing*, 461 NATURE 168, 168 (2009) [hereinafter Toronto Authors].

125. Carlos B. Amat, *Editorial and Publication Delay of Papers Submitted to 14 Selected Food Research Journals. Influence of Online Posting*, 74 SCIENTOMETRICS 379 (2008).

126. See William D. Garvey & Belver C. Griffith, *Scientific Information Exchange in Psychology*, 146 SCI. 1655, 1656 (1964) (reporting that in the psychology field, the time between hypothesis and publication is between 30 and 36 months, and the time between reportable results and publication is between 18 and 21 months); Charles G. Roland & Richard A. Kirkpatrick, *Time Lapse Between Hypothesis and Publication in the Medical Sciences*, 292 J. AM. MED. ASS'N 1273, 1274 (1975) (finding delays of 20 and 24 months between the completion of research and publication, respectively, for medical laboratory research and clinical research studies). Anecdotally, the author has been informed that publication delays are typically even longer in the social sciences.

127. See *supra* notes 1–3 and accompanying text.

publication of results based on his or her own experimental data. This distinction is important as it illustrates the dichotomy between two critical stakeholders in the science commons: data generators and data users. The divergent interests of these two constituencies played a major role in the development of the genome commons, as described in Section IV.A, *infra*.

2. *Copyright and Other Restraints on the Dissemination of Data*

Intellectual property protection has figured prominently in the ongoing debate over access to scientific data. Intellectual property can have two principal effects on an information commons: it can prevent the entry of data into the commons (impeding the growth of the overall knowledge base and thereby increasing knowledge latency), and it can limit the ability of users to utilize data that is already in the commons (impeding the growth of the unencumbered knowledge base and thereby increasing rights latency). Broadly speaking, intellectual property rights such as copyrights and trade secrets typically, though not always, fall into the first category and patents typically fall into the second category.

The effect of trade secret protection is the most straightforward to analyze. Scientific work that is sponsored by industry is often subject to written confidentiality obligations or other trade secret restrictions that explicitly prevent scientists from sharing resulting data and materials with others and, in some cases, delaying or even prohibiting the publication of their results.¹²⁸ BP's recent attempt to suppress data collected from the *Deepwater Horizon* oil spill is a prominent example of this practice.¹²⁹ With such restrictions in place, data cannot, practically speaking, enter a commons.¹³⁰ But corporate restrictions are not the only sources of secrecy in

128. See NAS, RESEARCH DATA, *supra* note 82, at 67; Allen C. Nunnally, *Intellectual Property Perspectives in Pharmacogenomics*, 46 JURIMETRICS J. 249, 255–58 (2006); Rai, *supra* note 4, at 111 (“Commercial involvement in academic research has . . . undermined norms governing the sharing of research materials and tools.”); INST. OF MEDICINE, EXTENDING THE SPECTRUM OF PRECOMPETITIVE COLLABORATION IN ONCOLOGY RESEARCH: WORKSHOP SUMMARY 36 (Margie Patlak, et al., eds. 2010) [hereinafter INST. OF MED.] (“Competing companies often compel their employees to keep silent about their endeavors, and the sharing of information is often frowned on lest information be divulged that might compromise the company’s competitive advantage.”).

129. See Mascarelli, *supra* note 16, at 538 (noting both the public outcry at BP’s initial attempts to “gag” scientific findings surrounding the leaking oil rig and subsequent attempts by the company to relax these restrictions).

130. At least not by legitimate means. Confidential data may be leaked or otherwise disclosed in violation of applicable confidentiality obligations and thereby enter the commons. This route toward public dissemination, however, is sporadic, unpredictable, and should not typically be relied upon by designers of information commons in the sciences.

the scientific community. Academic researchers themselves often have strong incentives to keep scientific data confidential, at least until the time of publication, and these incentives are supported, if not mandated, by university policies and procedures.¹³¹

The analysis of copyright restrictions is more subtle. Copyright law, unlike the law of trade secrets, does not have as its goal the concealment of information. Rather, copyright law grants the owner of a copyrighted work certain exclusive rights to exploit that work, including, for example, the exclusive rights to display, distribute, and reproduce.¹³² In the context of information commons, copyright principles are typically raised when discussing limitations on access to scientific data that has already been published. That is, even though scientific facts and conclusions are not themselves copyrightable, the articles (including text, diagrams and illustrations) in which they are presented are subject to copyright and thus controlled by the publishers of the journals carrying those articles.¹³³

In some cases, even data that might otherwise be in the public domain (such as mapping and geographic data developed under contract to the U.S. Federal Government) may be stored in proprietary databases that are accessible only to paid subscribers.¹³⁴ In several areas the “privatization” of governmental data is proceeding at a rapid pace due to (not entirely unjustified) perceptions of inefficiency and poor quality of governmental databases.¹³⁵ This situation, however, has led to fears that increasing amounts of data will be “enclosed” from public view; as Reichman and Uhler write, “the private sector simply cannot duplicate the government’s public good functions and still make a profit.”¹³⁶ This sentiment echoes a general criticism that the proprietary nature of scientific journals and electronic databases in which scientific results are published, and the high cost of subscribing to those journals and databases, is severely limiting the dissemination of

131. See INST. OF MED., *supra* note 128, at 36–37.

132. 17 U.S.C. § 106 (2006).

133. In many instances, authors are required to assign copyright in their articles and other works to journal publishers, or at least to grant publishers the exclusive right to publish their work in most relevant channels. See NAS, RESEARCH DATA, *supra* note 82, at 67.

134. See *id.* at 65. While the United States does not offer specific intellectual property protection for databases, such protection is available in Europe and elsewhere. See Reichman & Uhler, *supra* note 61, at 355.

135. See Reichman & Uhler, *supra* note 61, at 396.

136. *Id.* at 397.

knowledge.¹³⁷ The issue of scientific journals and the recent “open access” publishing movement in the sciences is discussed in greater detail in Section IV.B.

3. *Patents and Restraints on the Use of Data*
a) Effect of Patents on Scientific Data

Patents may be obtained in most countries to protect novel and inventive articles of manufacture, compositions of matter, and processes. Excluded from patentable subject matter are laws of nature and natural phenomena.¹³⁸ Thus, the information included within a science commons may be more or less susceptible of patent protection depending on its character. Pure scientific discoveries, such as observational data regarding the moons of Jupiter and conclusions regarding the giant planet’s mass and rotation drawn therefrom are unlikely to be suitable subject matter for patent protection. The same is not necessarily true in the biosciences, where the patentability of genetic, biological, and physiological associations has previously been affirmed by the U.S. Patent and Trademark Office and various courts (although the patentability of such subject matter remains the subject of substantial litigation and uncertainty).¹³⁹ In this respect, the significance of a

137. See, e.g., Susan R. Poulter, *Legal Pressures on the Public Domain: Licensing Practices*, in NRC, PUBLIC DOMAIN, *supra* note 7, at 101–03; Boyle, *supra* note 47, at 38–40; Reichman & Uhler, *supra* note 61, at 319–22. For a discussion of similar database protection issues in Europe and their potential effect on the dissemination of scientific knowledge, see NAT’L RESEARCH COUNCIL, BITS OF POWER — ISSUES IN GLOBAL ACCESS TO SCIENTIFIC DATA 150–53 (1997) [hereinafter NRC, BITS OF POWER]. See also Julie Cohen, *The Challenge of Digital Rights Management Technologies*, in NRC, PUBLIC DOMAIN, *supra* note 7, at 109 (discussing additional deleterious effects that DRM technologies may have on scientific research).

138. See, e.g., EUROPEAN PATENT OFFICE, EUROPEAN PATENT CONVENTION art. 52 (2000), available at <http://www.epo.org/patents/law/legal-texts/epc.html> (excluding from the meaning of invention, and thus patentable subject matter, “discoveries, scientific theories and mathematical methods”); JAPAN PATENT OFFICE, PART II: REQUIREMENTS FOR PATENTABILITY, CHAPTER 1: INDUSTRIALLY APPLICABLE INVENTIONS art. 1.1, available at http://www.jpo.go.jp/tetuzuki_e/t_tokkyo_e/Guidelines/2_1.pdf (noting that “a law of nature as such” is not considered to be a statutory invention). In the United States, patentable subject matter includes “any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof.” 35 U.S.C. § 101 (2006). This has been broadly interpreted by the United States Supreme Court to include “anything under the sun that is made by man.” *Diamond v. Chakrabarty*, 447 U.S. 303, 309 (1980). However, the Court has also recognized exclusions to this broad interpretation, including “laws of nature, natural phenomena, and abstract ideas.” *Diamond v. Diehr*, 450 U.S. 175, 185 (1981).

139. See, e.g., *Bilski v. Kappos*, 130 S. Ct. 3218 (2010) (affirming the Federal Circuit’s finding that a simplistic business method claim of hedging risk in commodities was not patentable subject matter); *Prometheus Labs., Inc. v. Mayo Collaborative Servs.*, No. 09-490,

human genetic mutation is fundamentally different from, say, the eccentricity of Io's elliptical orbit: the scientific community's ability to utilize this information, even if it is publicly known, may differ greatly depending on the existence or absence of relevant patents. As such, the ability to use information resident within a commons of DNA cancer markers may be more constrained than the use of data within the commons of Jovian satellite information.¹⁴⁰ That is, when a valid patent claims an "invention" such as a means of diagnosing or treating a disease based on the presence of a specified genetic marker, the patent prevents any unauthorized person from "using" that information for such a purpose, even if the information is known to that person via the commons.

b) Patents and Latency

Unlike copyright and trade secret restrictions, which, in the contexts described above, tend to limit access to scientific data, patents are required to teach the "best mode" of enabling the inventions that they claim and thus, by their nature, disclose knowledge to the public. The U.S. Supreme Court has long held that the "ultimate goal of the patent system is to bring new . . . [ideas] and technologies into the public domain through disclosure."¹⁴¹ In the United States, most patent applications are published and become publicly available eighteen months after filing.¹⁴² The quid pro quo for public disclosure of the invention is the exclusive right to exploit the invention during the patent term (twenty years from filing in the United

2010 WL 2571881 (U.S. June 29, 2010) (remanding to the appellate court for reconsideration of whether simple medical diagnosis method claims meet the subject matter requirements of 35 U.S.C. § 101 in light of the Supreme Court's decision in *Bilski v. Kappos*, *supra*); Ass'n for Molecular Pathology v. U.S. Patent & Trademark Office, No. 09 Civ. 4515, 2010 WL 1233416 (S.D.N.Y. Mar. 29, 2010) (ruling that patent claims which claimed isolated human DNA sequences (genes) or methods of comparison of these genes to those of a human sample were ineligible for patent protection under 35 U.S.C. § 101 as they were products of nature and mental processes, respectfully).

140. Cf. INST. OF MED. & NAT'L RESEARCH COUNCIL, LARGE-SCALE BIOMEDICAL SCIENCE 27 (2003) [hereinafter LARGE-SCALE SCIENCE] (noting greater willingness among scientists working on "big-science" projects in astronomy and similar areas to share data due to the smaller commercial value of such discoveries as compared to that in the biosciences).

141. *Bonito Boats v. Thundercraft*, 489 U.S. 141, 151 (1989); see also LANDES & POSNER, *supra* note 54, at 294–99 (noting that the knowledge disclosed in a patent document may be used by competitors seeking to "invent around" the patented invention, but also observing that the competitive value of this knowledge decreases as the scope of patent protection increases); Eisenberg, *supra* note 101, at 1022–23 (explaining the argument that "disclosure of unprotected information is not an incidental byproduct of a process that aims to motivate something more worthwhile, but the whole point of the patent system").

142. 35 U.S.C. § 122 (2006).

States).¹⁴³ During this term, no person may make, use, sell, or import the patented invention without the permission of the patent owner, and even academic research activities have been found to constitute infringement.¹⁴⁴ Thus, in terms of latency analysis, patents create a “commons” of public knowledge in which knowledge latency may be expressed as the period of time between knowledge generation and publication of the patent application, and rights latency may be expressed as the period of time between publication of the application until expiration of the patent. Represented numerically, a “patent commons” would yield the following latency values:

$$\textit{knowledge latency (KL)} = \textit{preparation time} + 18 \textit{ months}$$

$$\textit{rights latency (RL)} = 240 \textit{ months} - 18 \textit{ months} = 222 \textit{ months}$$

These values are presented here principally for comparative purposes in the discussion that follows in Part V.

c) Patents and Anticommons

Significant debate exists concerning the effect that patents may have on scientific research and whether patents promote or chill innovation. While many commentators trace this controversy to the early 1980s and the enactment in the United States of the Bayh-Dole Act,¹⁴⁵ its roots extend at least to the 1850s and virulent opposition in Britain and continental Europe to the then-newly formalized patent systems.¹⁴⁶ Among the charges leveled

143. *Kewanee Oil v. Bicron*, 416 U.S. 470, 481 (1974). In this case, the Court states: [S]uch additions to the general store of knowledge are of such importance to the public weal that the Federal Government is willing to pay the high price of 17 years of exclusive use for its disclosure, which disclosure, it is assumed, will stimulate ideas and the eventual development of further significant advances in the art.

Id. (referring to the then-current seventeen-year patent term); *see also* *Sinclair & Carroll Co. v. Interchemical Corp.*, 325 U.S. 327, 330–31 (1945); *Griffith Rubber Mills v. Hoffar*, 313 F.2d 1 (9th Cir. 1963). *But see* Alan Devlin, *The Misunderstood Function of Disclosure in Patent Law*, 23 HARV. J.L. TECH. 401, 403–04 (arguing that patent disclosures seldom provide technically useful information to the public); Note, *The Disclosure Function of the Patent System (or Lack Thereof)*, 118 HARV. L. REV. 2007 (2005) (challenging the notion that in practice the patent system serves to disseminate knowledge).

144. *See* *Madey v. Duke Univ.*, 307 F.3d 1351 (Fed. Cir. 2002).

145. Bayh-Dole Act of 1980, Pub. L. No. 96-517, 94 Stat. 3015 (codified at 35 U.S.C. §§ 200–212). This controversial piece of legislation rationalized the previously chaotic rules governing federally sponsored inventions and allowed researchers to obtain patents on inventions arising from government-funded research.

146. *See* JOHNS, *supra* note 79, at 247–48, 267–75 (describing the mid-Victorian patent abolition movement).

against the patent system by Victorian critics was the deleterious effect that such grants of exclusive rights might have on the “commons” of scientific and technical knowledge.¹⁴⁷ Similar fears were expressed in the United States in the 1930s following the Federal Communications Commission’s investigation of AT&T and its massive patent holdings.¹⁴⁸

Anxiety over the effect of patents on the science commons continues today, the debate often being reduced to the oversimplified tradeoff between “access” and “incentive.”¹⁴⁹ In their influential 1998 paper, Michael Heller and Rebecca Eisenberg envision a scenario in which a proliferation of ownership claims within fields of study make it increasingly difficult to conduct research in those fields either because rights owners are unwilling to license patents on acceptable terms or because the sheer number of rights holders make it impractical or cost-prohibitive for researchers to procure the necessary rights.¹⁵⁰ They describe this effect as a “tragedy of the anticommons”¹⁵¹ and argue that such a state of affairs could effectively “privatize” research results that otherwise would have been freely available and thereby reduce the quantity of socially useful biomedical research.¹⁵² Other commentators have echoed Heller’s and Eisenberg’s concerns.¹⁵³ Arti Rai, in particular, argues that the increased availability of patents on inventions developed at academic laboratories (due in part to the enactment of the Bayh-Dole Act) has eroded the traditional “open science” norms of academic research and, consequently, has negatively affected the advancement of science.¹⁵⁴

147. *Id.* at 271.

148. *Id.* at 402–09 (calling the AT&T investigation “the era’s principal venue for debating the consequences of patents in general for society, science, and industry”).

149. See LANDES & POSNER, *supra* note 54, at 11.

150. Heller & Eisenberg, *supra* note 32. Heller and Eisenberg do not address the public-oriented approach of the Bermuda Principles described in Section IV.A, but focus instead on increasing patent-generating activity in other sectors of the biomedical research arena.

151. Echoing Hardin’s “tragedy of the commons.” See *supra* Section II.A.1.

152. Heller & Eisenberg, *supra* note 32, at 698.

153. See, e.g., Lori Andrews et al., *When Patents Threaten Science*, 314 SCI. 1395 (2006); Benkler, *supra* note 4, at 1110; Hilgartner, *supra* note 121, at 137 (“Excessive concern with the protection of intellectual property can erect barriers to establishing new public domains.”); Lee, *supra* note 115, at 903–05 (arguing that patents on basic biomedical research information have the potential to hinder research).

154. Rai, *supra* note 4, at 109 (“[U]niversities and individual researchers soon began to respond to the financial incentives of Bayh-Dole by rejecting communalism and increasing efforts to seek patents.”). Rai’s comprehensive analysis of norms in the biosciences is based, in part, on Eisenberg’s article. Rebecca S. Eisenberg, *Proprietary Rights and the Norms of Science in Biotechnology Research*, 97 YALE L.J. 177 (1987); see also DANIEL S. GREENBERG, SCIENCE

The effects and very existence of the “anticommons,” however, have been challenged by scholars who question many of the assumptions underlying Eisenberg’s and Rai’s construction of the “science commons” and counter that patents promote, rather than dampen innovation.¹⁵⁵ In the past few years, researchers have begun to apply empirical techniques to test the “tragedy of the anticommons.”¹⁵⁶ By and large these studies have been inconclusive, showing either no perceptible anticommons effect or only a modest effect.¹⁵⁷

FOR SALE: THE PERILS, REWARDS, AND DELUSIONS OF CAMPUS CAPITALISM 61 (2007); JENNIFER WASHBURN, UNIVERSITY, INC.: THE CORPORATE CORRUPTION OF HIGHER EDUCATION 73–75 (2005); John V. Frangioni, *The Impact of Greed on Academic Medicine and Patient Care*, 26 NATURE BIOTECHNOLOGY 503 (2008); Lee, *supra* note 115, at 941–42.

155. See, e.g., David E. Adelman, *A Fallacy of the Commons in Biotech Patent Policy*, 20 BERKELEY TECH. L.J. 985 (2005) (arguing that the public commons model incorrectly assumes that the commons for biomedical science is “finite and congested,” thus overstating the value of a commons approach); Shubha Ghosh, *How to Build a Commons: Is Intellectual Property Constrictive, Facilitative or Irrelevant?*, in KNOWLEDGE AS A COMMONS, *supra* note 18, at 209, 212–24 (arguing that the effect of intellectual property on commons development is context-dependent and, as such, may hinder, facilitate, or be irrelevant to commons building); F. Scott Kieff, *Facilitating Scientific Research: Intellectual Property Rights and the Norms of Science—A Response to Rai and Eisenberg*, 95 Nw. U. L. REV. 691 (2001) (rebutting the commons-based arguments posited by Rai and Eisenberg); Opderbeck, *supra* note 49, at 215–18 (arguing that private Coasian negotiations are likely to lead to optimal results in the market for biotechnology innovation).

156. See, e.g., John P. Walsh, Ashish Arora & Wesley M. Cohen, *Effects of Research Tool Patents and Licensing on Biomedical Innovation in Patents*, in PATENTS IN THE KNOWLEDGE-BASED ECONOMY 285 (Wesley M. Cohen & Stephen A. Merrill eds., 2003) (finding, based on interviews and archival data, little evidence that academic research has been impeded by concerns about patents on research tools); John P. Walsh, Charlene Cho & Wesley M. Cohen, *View from the Bench: Patents and Material Transfers*, 309 SCI. 2002 (2005) (finding that patents generally are not used to deny access to knowledge inputs, but that access to research materials may be more difficult); Karim R. Lakhani et al., *The Value of Openness in Scientific Problem Solving* (Harvard Bus. Sch., Working Paper No. 07-050, 2006), available at <http://www.hbs.edu/research/pdf/07-050.pdf> (offering empirical evidence in support of the proposition that the solution of scientific problems is facilitated by free and open information sharing); Fiona Murray & Scott Stern, *Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the Anti-Commons Hypothesis* (NBER Working Paper Series, Paper No. 11465, 2005), available at <http://www.nber.org/papers/w11465/> (finding some empirical support for a modest anti-commons effect).

157. See, e.g., David, *supra* note 82, at 29 (“There is not much empirical evidence as to how altering the legal conditions and terms of intellectual property rights translates into change in the overall strength of economic incentives for the producers.”); Eisenberg, *supra* note 101, at 1018 (“It is less clear that patents are dampening incentives to engage in further basic research in the non-profit sector.”); Ann E. Mills & Patti M. Tereskerz, *Changing Patent Strategies: What Will They Mean for the Industry*, 25 NATURE BIOTECHNOLOGY 867, 867 (2007) (citing “diminished concern in the biotech industry over the possible occurrence of an anticommons”); Michael S. Mireles, *An Examination of Patents, Licensing, Research Tools, and the*

But even if an anticommons has not arisen as a result of patenting in the biosciences, the prospect of obtaining patent protection covering the fruits of scientific discoveries has had an effect on the rapidity with which scientific data is disclosed to the public. To ensure that disclosure and publication do not defeat patenting rights, many universities require that researchers submit proposed articles that disclose potentially patentable inventions to a university technology office for review prior to submitting the article to a journal in order to give the university sufficient time to prepare and file a patent application if so desired.¹⁵⁸ These express policies, together with increasing pressure from universities and research institutions to commercialize their scientific discoveries, have led to additional delays in the release of data, both through publication and other channels. One study has found that approximately twenty percent of life sciences faculty delayed publication of research results for more than six months due to patent or other commercial issues.¹⁵⁹ Another study reports that scientists with an

Tragedy of the Anticommons in Biotechnology Innovation, 38 U. MICH. J.L. REFORM 141, 171–94 (2005) (finding previous studies inconclusive); Charles R. McManis & Suheol Noh, *The Impact of the Bayh-Dole Act on Genetic Research and Development: Evaluating the Arguments and Empirical Evidence to Date* 48 (Aug. 13, 2006) (unpublished draft), available at <http://www.law.berkeley.edu/institutes/bclt/ipsc/papers2/mcmanis.doc> (“[V]ery little empirical evidence has been produced to date to support the argument that granting patents on the results of ‘upstream’ genetic research undermines the norms of the biological research community or retards biomedical innovation, technology transfer, or the development of downstream commercial products and processes.”). Indeed, some economics scholars even point out that the historical “enclosure” movement that fenced-off common pastures in England from the sixteenth to nineteenth centuries, and which is often cited by commons proponents as a negative development, see, for example, Boyle, *supra* note 47, significantly enhanced agricultural output. LANDES & POSNER, *supra* note 54, at 12.

158. Frangioni, *supra* note 154, at 504 (describing the rise of academic invention disclosure policies geared toward patent filings); see, e.g., *About WUSTL: Compliance & Policies: Intellectual Property Policy*, WASHINGTON U. IN ST. LOUIS sec. I.3.d (Feb. 15, 2008), <http://wustl.edu/policies/intelprop.html>. Section I.3.d states:

Nothing in this Policy shall limit or restrict the right of University faculty and students to publish results of their research, subject to reasonable delays to preserve patent or other intellectual property rights. Delays in publication required by the University . . . as a general rule, shall not exceed ninety days from initial disclosure of the intellectual property to the Research Office

Id.; see also UNIV. OF N.C. AT CHAPEL HILL, PATENT & INVENTION POLICY sec. III (2009), available at <http://www.unc.edu/campus/policies/patent%20policy%2000015747.pdf> (“[T]his Patent & Invention Policy does not limit the right to publish, except for short periods of time necessary to protect patent rights.”).

159. Blumenthal et al., *National Survey*, *supra* note 17, at 1224. This data has been supported by numerous subsequent studies, including one reporting that the disclosure of a gene associated with the condition hemochromatosis (an iron overload disease) was delayed

interest in patenting their results regularly withhold presentation of their data at scientific meetings.¹⁶⁰

In this Article, I do not seek to stake out a normative position regarding the effect, positive or negative, that patents and other intellectual property rights have on innovation and scientific advancement. I do, however, wish to draw attention to the very real and measurable effect that patents and other intellectual property have on science commons. That is, these forms of intellectual property protection, whether they promote or deter scientific research, undeniably reduce the rate at which the results of such research are contributed to the store of public knowledge and the speed at which such information becomes generally usable. Whether or not this delay yields a net benefit to society, policy makers designing information commons in the sciences will likely act in accordance with their own perceptions regarding this normative question. In the case studies that follow, I examine two well-known narratives of commons formation in the sciences with a particular view toward the timing-based adjustments and compromises that policy designers implemented in furtherance of their own institutional goals.

IV. TIMING AND DATA SHARING: TWO CASE STUDIES

A. THE GENOME COMMONS¹⁶¹

1. *Development of Genomics Data Release Policies*

a) The Human Genome Project (HGP)

The Human Genome Project, one of the most ambitious scientific undertakings in history, has been compared in scope and importance to the Manhattan Project and the manned space program.¹⁶² The project, which

by more than a year after the first patent application was filed by the sponsoring institution. Jon F. Merz et al., *Diagnostic Testing Fails the Test: The Pitfalls of Patents are Illustrated by the Case of Haemochromatosis*, 415 NATURE 577 (2002).

160. Jeremy M. Gruschcow, *Measuring Secrecy: A Cost of the Patent System Revealed*, 33 J. LEGAL STUD. 59, 73 (2004) (measuring the gap between patent filing and presentations at cancer research meetings).

161. A more detailed analysis of the policy considerations surrounding the development of the “genome commons” can be found in Contreras, *supra* note 35.

162. See, e.g., FRANCIS S. COLLINS, *THE LANGUAGE OF LIFE* 2 (2010); ARTHUR M. LESK, *INTRODUCTION TO GENOMICS* 22 (2007); James D. Watson, *The Human Genome Project: Past, Present and Future*, 248 SCI. 44, 44 (1990).

spanned fifteen years and involved over a thousand scientists worldwide, has resulted in major advances in biochemistry, bioinformatics, and genetics.¹⁶³

From the outset, scientists expected that the HGP would generate large quantities of data regarding the genetic make-up of humans and other organisms. Thus, in 1988 the National Research Council was commissioned to study issues likely to be raised by the HGP. It recommended that all data generated by the project “be provided in an accessible form to the general research community worldwide.”¹⁶⁴ Other influential groups and policy makers soon took up the call for the commitment of all human sequence information to the public domain.¹⁶⁵ Perhaps most importantly, many leading scientists involved in the project urged that the output of the HGP be released to the public in the service of science.¹⁶⁶ They may have been influenced by pre-existing data sharing norms in the research communities in which they worked,¹⁶⁷ or held a genuine belief that genomic information, as the common biological heritage of the human species, was qualitatively *different* than other scientific data. As expressed by Dr. Ari Patrinos, the Department of Energy’s (DOE) Associate Director for Biological and Environmental Research, “the genome belongs to everybody.”¹⁶⁸

163. See generally COLLINS, *supra* note 36, at 674 (describing the “profound impact on scientific progress” achieved by the HGP); Int’l Hum. Genome Sequencing Consortium, *Initial Sequencing and Analysis of the Human Genome*, 409 NATURE 860, 911–13 (2001) [hereinafter Genome Sequencing Consortium].

164. NAT’L RESEARCH COUNCIL, MAPPING AND SEQUENCING THE HUMAN GENOME 8 (1988) [hereinafter NRC, HUMAN GENOME].

165. See NAT’L RESEARCH COUNCIL, REAPING THE BENEFITS OF GENOMIC AND PROTEOMIC RESEARCH 27 (2006) [hereinafter NRC, GENOMIC AND PROTEOMIC RESEARCH] (citing a 1988 NIH advisory council report stating that “the data must be in the public domain, and the redistribution of the data should remain free of royalties”).

166. James Watson, then director of the National Center for Human Genome Research, wrote in 1990 that “making the sequences widely available as rapidly as practical is the only way to ensure that their full value will be realized.” Watson, *supra* note 162, at 48; see also COLLINS, *supra* note 162, at 301; Boyle, *supra* note 47, at 37.

167. See Robert Cook-Deegan, *The Urge to Commercialize: Interactions Between Public and Private Research Development*, in NRC, PUBLIC DOMAIN, *supra* note 7, at 89 (“There were [] communities doing molecular biology at the same time on yeast and *Drosophila* that had ‘open-science’ norms. Those norms were the ones adopted as the models for the Human Genome Project.”); Genome Sequencing Consortium, *supra* note 163, at 864. The evolution of the open science culture among *c. elegans* researchers is described in some detail in NRC, GENOMIC AND PROTEOMIC RESEARCH, *supra* note 165, at 54–56.

168. Eliot Marshall, *Bermuda Rules: Community Spirit, With Teeth*, 291 SCI. 1192 (2001).

In 1992, the federal co-managers of the project, NIH¹⁶⁹ and DOE, approved joint guidelines for the sharing of data by investigators performing HGP-funded activities, acknowledging that the rapid sharing of HGP materials and data was “essential for progress toward the goals of the program,” and that such sharing would avoid unnecessary duplication of effort and expedite research in other areas.¹⁷⁰ Genomic sequence data generated by the HGP would be deposited in the NIH-managed GenBank database, and would be accessible to all scientists.¹⁷¹ However, the 1992 guidelines counterbalanced the desire for rapid public release of data against a perceived need to give researchers “some scientific advantage from the effort they have invested.”¹⁷² In recognition of this effort, the 1992 guidelines established a six-month maximum period from the time that HGP data is generated until the time that it must be made publicly available (i.e., knowledge latency of six months).¹⁷³ During this six-month latency period, HGP researchers could analyze their data and prepare publications, and only after the end of the six-month period were they required to release the data into public databases.¹⁷⁴

169. NIH formed the National Center for Human Genome Research (NCHGR) in 1989, under the direction of James Watson, to carry out its component of the HGP. In 1997, the Department of Health and Human Services elevated NCHGR to the status of a full “institute” within the NIH system, forming the National Human Genome Research Institute (NHGRI). *About the Institute: A History and Timeline*, NAT’L HUM. GENOME RES. INST., <http://www.genome.gov/10001763/> (last visited Sept. 17, 2010).

170. *DOE-NIH Guidelines Encouraging Sharing of Data and Resources*, HUM. GENOME NEWS, Jan. 4, 1993, at 4 [hereinafter *DOE-NIH Guidelines*].

171. See NRC, GENOMIC AND PROTEOMIC RESEARCH, *supra* note 165, at 4–5.

172. *DOE-NIH Guidelines*, *supra* note 170, at 4.

173. This six-month period was significantly shorter than release periods for other government-funded projects, which often allowed researchers to retain their data privately either until publication of results or for some specified “exclusivity period,” usually in the neighborhood of twelve to eighteen months. See, e.g., NAS, RESEARCH DATA, *supra* note 82, at 64 (noting that NASA and the European Southern Observatory Administration impose twelve-month proprietary periods and the U.S. National Optical Astronomy Observatory imposes an eighteen-month proprietary period on the release of data); NRC, SHARING PUBLICATION-RELATED DATA, *supra* note 87, at 75 (describing the one-year “hold allowance” on the deposition of crystallography data into the Protein Data Bank); NRC, BITS OF POWER, *supra* note 137, at 80–82 (describing data release policies of NASA and Global Change Research Program); Reichman & Uhler, *supra* note 61, at 335 (“In most cases, publication of research results marks the point at which data produced by government-funded investigators should become generally available.”).

174. *DOE-NIH Guidelines*, *supra* note 170.

b) The Bermuda Principles

By 1996, the HGP, together with its international collaborators, was ready to begin sequencing the human genome. In February, approximately fifty scientists and policy makers met in Hamilton, Bermuda to discuss the sharing and release of HGP-funded genomic data.¹⁷⁵ Among the issues debated was the speed at which data should be released to the public, and whether the six-month “holding period” approved in 1992 should continue.¹⁷⁶ Several arguments for eliminating the holding period were presented. From a pragmatic standpoint, scientists argued that gene sequencing centers working on the HGP required regularly updated data sets in order to avoid duplication of effort and to optimize coordination of the massive, multi-site project.¹⁷⁷ Waiting six months to obtain data was simply not practical if the project were to function effectively. But perhaps more importantly, the concept of rapid data release became endowed with an ideological character: the early release of data was viewed as necessary to accelerate the progress of science.¹⁷⁸ The result of the meeting was the

175. The meeting, sponsored by the Wellcome Trust, was designated the International Strategy Meeting on Human Genome Sequencing and held on February 25–28, 1996. Participants included representatives of NIH and DOE, the Wellcome Trust, UK Medical Research Council, the German Human Genome Programme, the European Commission, the Human Genome Organisation (HUGO) and the Human Genome Projects of France and Japan. See David Smith & Anthony Carrano, *International Large-Scale Sequencing Meeting*, HUM. GENOME NEWS, Apr.–June 1996, at 19.

176. See Marshall, *supra* note 168; Robert Mullan Cook-Deegan & Stephen J. McCormack, *A Brief Summary of Some Policies to Encourage Open Access to DNA Sequence Data*, 293 SCI. 217 (2001) (supplemental material), available at <http://www.sciencemag.org/cgi/content/full/293/5528/217/DC1>.

177. David R. Bentley, *Genomic Sequence Information Should Be Released Immediately and Freely in the Public Domain*, 274 SCI. 533, 533 (1996); see also Adam Bostanci, *Sequencing Human Genomes*, in FROM MOLECULAR GENETICS TO GENOMICS 174 (Jean-Paul Gaudillière & Hans-Jörg Rheinberger eds., 2004).

178. See, e.g., Bentley, *supra* note 177, at 534; Francis Collins et al., *A Vision for the Future of Genomics Research*, 422 NATURE 835, 846 (2003) (“Scientific progress and public benefit will be maximized by early, open and continuing access to large data sets.”); Cook-Deegan & McCormack, *supra* note 176 (“[W]ithout [the Bermuda Principles] the wait for information sufficient to meet patent criteria from high throughput sequencing programs would lead to long delays, and thus be a serious drag on science, undermining the publicly funded sequencing programs’ very purpose.”); Genome Sequencing Consortium, *supra* note 163, at 864 (“We believed that scientific progress would be most rapidly advanced by immediate and free availability of the human genome sequence. The explosion of scientific work based on the publicly available sequence data in both academia and industry has confirmed this judgment.”); Lee, *supra* note 115, at 905 (referring to “NIH’s aggressive intervention to enhance access to taxpayer-financed research tools”); Toronto Authors, *supra* note 124, at 169–70.

“Bermuda Principles,” a set of guidelines endorsed by each participant and intended to prevent any genome sequencing center from establishing a “privileged position in the exploitation and control of human sequence information.”¹⁷⁹ Most significantly, the Bermuda Principles required that genomic data be released to GenBank within twenty-four hours following assembly, the fastest data release requirement in history.¹⁸⁰

c) Bermuda, Data Release, and Patents

Not all observers viewed the adoption of the Bermuda Principles as an unmitigated success. Some critics questioned the effect that rapid release of data would have on scientists’ ability to publish the results of their work before others who had accessed it via public means.¹⁸¹ And, more pragmatically, some claimed that rapid data release was not even required in order to achieve the scientific goals of the HGP.¹⁸² But the most persistent skepticism regarding the Bermuda Principles came from those who believed that they subverted researchers’ ability to protect their discoveries with patents. In particular, the Bermuda Principles and their adoption by NIH in 1997 ensured that genomic data from the HGP and other large-scale sequencing projects would be made publicly available before data generators had an opportunity to file patent applications covering any “inventions” arising from that data, and in a manner that ensured its availability as prior art against third party patent filings at the earliest possible date.¹⁸³ As a result, the ability of both data generators and third parties to obtain patents on HGP

179. Bermuda Principles, *supra* note 35.

180. *Id.*

181. Deanna M. Church & LeDeana W. Hillier, *Back to Bermuda: How Is Science Best Served?*, 10 GENOME BIOLOGY 105, 105.1 (2009) (“[T]here was some concern that [the policy] would jeopardize the genome center’s ability to analyze and publish the data they had produced.”).

182. Bostanci, *supra* note 177, at 175 (arguing that the yeast genome project, a cooperative effort of seventy-nine different laboratories, operated successfully with a delayed data release policy).

183. In jurisdictions such as the European Union and Japan that have so-called “absolute novelty” requirements, an invention may not be patented if it has been publicly disclosed prior to the filing of a patent application. *See* JOHN GLADSTONE MILLS III ET AL., PATENT LAW FUNDAMENTALS sec. 2:30 (perm. ed., rev. vol. 2009). In such countries, a description of the invention in a scientific journal could preclude the inventor from obtaining patent protection for his or her invention. In the United States, a patent application may be filed with respect to an invention that has been disclosed in a printed publication, but only if the publication occurred less than one year before the filing of the patent application. 35 U.S.C. § 102(b) (2006). Thus, an inventor seeking patent protection for his or her invention must file a patent application prior to the disclosure of the invention in a publication (or, in the United States, no more than one year following publication).

data was severely curtailed, both on grounds of “novelty” (one cannot patent material that is “known . . . by others”)¹⁸⁴ and “nonobviousness” (one cannot patent material that is obvious in view of the prior art).¹⁸⁵ Rapid release of HGP data under the Bermuda Principles thus achieved a second policy goal of project leaders: limiting patent-related encumbrances on HGP data and the fruits thereof.¹⁸⁶ Critics of this approach charged that the NIH’s support of the Principles contravened the requirements of the Bayh-Dole Act, which expressly permits federally funded researchers to patent their discoveries.¹⁸⁷ However, scientists and NIH leaders dismissed these arguments regarding patents and Bayh-Dole, arguing instead that patenting of raw DNA sequences is inappropriate and should be impossible.¹⁸⁸

184. 35 U.S.C. § 102(a). This approach has also been used in the case of traditional knowledge released by the Indian government through a publicly accessible digital library for the purpose of constituting prior art “to prevent erroneous patent grants.” John Swinson & Laura Pearson, *Protection of Traditional Yoga-Related Knowledge in a Modern World*, BNA WORLD INTELLECTUAL PROP. RPT., Sept. 3, 2010.

185. 35 U.S.C. § 103 (2006). For example, the publication of a full genetic sequence would likely prevent the patenting of a gene included in that sequence even though the gene is disrupted by gaps (introns) in the genetic code that are excised during gene expression. Alexander K. Haas, *The Wellcome Trust’s Disclosure of Gene Sequence Data into the Public Domain & the Potential for Proprietary Rights in the Human Genome*, 16 BERKELEY TECH. L.J. 145, 158–59 (2001).

186. *See supra* note 166. As a corollary effect, the requirement of the Bermuda Principles that sequence data be deposited into GenBank and other public databases also had the effect of eliminating the copyright barriers to data sharing that are discussed *supra* notes 122–37 and accompanying text. Interestingly, Rebecca Eisenberg suggests that, in some cases, the early release of experimental data may actually encourage *more* patent filings by third parties who are thereby enabled to combine public data with proprietary improvements and patent the combination thereof. Eisenberg, *supra* note 101, at 1026.

187. *See* JAMES SHREEVE, *THE GENOME WAR* 46 (2004) (“Strictly speaking, the policy directly contradicted the Bayh-Dole Act.”); Arti K. Rai & Rebecca S. Eisenberg, *Bayh-Dole Reform and the Progress of Biomedicine*, 66 LAW & CONTEMP. PROBS. 289, 308 (2003) (“Arguably, NIH has acted outside the scope of its statutory authority . . . at least with respect to patentable inventions.”).

188. Bentley, *supra* note 177, at 534; *see also* JAMES D. WATSON ET AL., *RECOMBINANT DNA* 295 (3d ed. 2005); Marshall, *supra* note 168. The NIH policy implementing the Bermuda Principles states that “raw human genomic DNA sequence, in the absence of additional demonstrated biological information, lacks demonstrated specific utility and therefore is an inappropriate material for patent filing.” *NHGRI Policy Regarding Intellectual Property of Human Genomic Sequence*, NAT’L HUM. GENOME RES. INST., Apr. 9, 1996, <http://www.genome.gov/10000926/> [hereinafter *NHGRI 1996 Policy*]. NIH’s position regarding the patenting of genetic material can be traced to its abortive attempt to seek patents on short genetic sequences known as expressed sequence tags (ESTs). In early 1991, NIH filed patent applications covering hundreds of ESTs and announced its intention to continue seeking such patents in the future. *See* Thomas Barry, *Revisiting Brenner: A Proposed Resolution to the Debate Over the Patentability of Expressed Sequence Tags Using the Concept of Utility*

d) Growing Tension Between Data Generators and Users

Despite the success and broad adoption of the Bermuda Principles, in the years immediately following the completion of the HGP a number of large-scale, publicly funded genomics projects began to adopt data release policies reflecting an increasing recognition of the inherent tension between data *generators*, the scientists at large-scale gene sequencing centers that produce the bulk of raw genomic data, and data *users*, those scientists who access and use genomic data to investigate specific disease-gene associations and conduct other analyses that rely on this data. In 2003, the Wellcome Trust sponsored a meeting in Ft. Lauderdale, Florida to revisit issues of pre-publication data release in the “post-genome” world.¹⁸⁹ The most significant result of the Ft. Lauderdale meeting was an agreement that a broad category of scientific projects, termed “community resource projects” (CRPs), should be subject to the rapid, pre-publication data-release rules established in Bermuda.¹⁹⁰ NIH’s 2003 genomic data release policy, issued just a few weeks after the Ft. Lauderdale meeting, reiterated the agency’s commitment to the Bermuda Principles and recognized the need for data generators to publish analyses of their data.¹⁹¹ Yet despite this acknowledgement, the agency

Control, 35 AIPLA Q.J. 1, 11 (2007); Rebecca Eisenberg, *Intellectual Property at the Public-Private Divide: The Case of Large-Scale cDNA Sequencing*, 3 U. CHI. L. SCH. ROUNDTABLE 557, 558–59 (1996). This announcement sparked what Robert Cook-Deegan refers to as “an international firestorm” and led to threats by other HGP collaborators to refuse to share their own data with NIH. ROBERT COOK-DEEGAN, *THE GENE WARS: SCIENCE, POLITICS, AND THE HUMAN GENOME* 330–31 (1994). In addition to negative reactions from abroad, the NIH’s and DOE’s own advisory committees were “unanimous in deploring the decision to seek such patents.” *Id.* at 317. Even James Watson decried the EST patenting plan as “sheer lunacy.” SHREEVE, *supra* note 187, at 84–85. The EST debate marked a turning point in NIH’s attitude toward patents on genetic material. Despite its initial enthusiasm for such patents, in 1994 NIH elected not to appeal the Patent and Trademark Office’s rejection of its EST patent applications. *See* LARGE-SCALE SCIENCE, *supra* note 140, at 36–37. After that, the PTO adopted a consistently lukewarm, if not outright averse, attitude toward the patenting of genetic sequences. *See* NRC, *GENOMIC AND PROTEOMIC RESEARCH*, *supra* note 165, at 52–53.

189. THE WELLCOME TRUST, *SHARING DATA FROM LARGE-SCALE BIOLOGICAL RESEARCH PROJECTS: A SYSTEM OF TRIPARTITE RESPONSIBILITY* (2003), available at <http://www.genome.gov/Pages/Research/WellcomeReport0303.pdf> [hereinafter Ft. Lauderdale Principles].

190. As defined in the Ft. Lauderdale Principles, a “community resource project,” like the original HGP, is “a research project specifically devised and implemented to create a set of data, reagents or other material whose primary utility will be as a resource for the broad scientific community.” *Id.* at 2–3.

191. *Id.* at 4 (“[T]he sequence producers have a legitimate interest in publishing peer-reviewed reports describing and analyzing the sequence they have produced.”).

declined to impose any explicit limitation on users of the released sequence data. Instead, the policy reaffirmed NIH's position that DNA sequence data "should be available for all to use without restriction" and urged data users to act within "standard scientific norms" and to acknowledge data generators in published analyses based on their data.¹⁹² Data generators were left with the unsatisfying conclusion that potential loss of their publication priority was "a *necessary risk* set against the considerable benefits of immediate data release."¹⁹³

e) GWAS and Embargo Policies

Between 2003 and 2006, the technologies available for genomic research continued to improve in quality and decrease in cost, resulting in an expansion of the scope of genomic research and the advent of so-called genome-wide association (GWA) studies.¹⁹⁴ These studies differ from pure sequencing projects in that their goal is not the generation of large data sets (such as the genomic sequence of a particular organism), but the discovery of disease markers or associations hidden within the genome. They thus have greater potential clinical utility and lie further along the path to ultimate commercialization as therapeutics or diagnostics. Moreover, as the types of data involved in large-scale genomics projects expanded, the community of researchers participating in these projects became more diverse. Today, many scientists with backgrounds outside of genomics, including medical researchers, medical geneticists, clinicians, and epidemiologists, actively lead and participate in GWA projects. Yet these scientists do not necessarily share the norms of rapid pre-publication data release embraced by the genomics community since the early days of the HGP.¹⁹⁵ In many cases, particularly when patient data is involved, these scientists are accustomed to an

192. *Id.* For a general critique of the NIH's "hortatory" approach to this issue, see Rai & Eisenberg, *supra* note 187, at 293–94.

193. Ft. Lauderdale Principles, *supra* note 189, at 4 (emphasis added). In addition, the Ft. Lauderdale Principles suggest that data generators publish "project descriptions" (sometimes referred to as "marker papers" or "statements of intent") in order to establish a published priority claim to the work they are conducting. *See id.* at 3; Toronto Authors, *supra* note 124, at 169. Based on informal discussions with genomic scientists, it appears that few of these project descriptions have been published, though efforts are emerging to increase this practice. *See Editorial: Cite Site*, 42 NATURE GENETICS 729 (2010) (describing the *Nature Preprints* archive that is intended to offer a place for data generators to list citable project descriptions (marker papers) and receive citation credit).

194. *See* Monya Baker, *Genome Studies: Genetics by Numbers*, 451 NATURE 516 (2008).

195. *See, e.g.*, Cook-Deegan, *supra* note 167, at 90 (noting "thousands of labs" outside the traditional genomics disciplines that perform gene sequencing, but few of which observe the Bermuda Principles).

environment in which data is tightly guarded and released only after publication of results, and then only in a limited, controlled manner.

Accordingly, when the federally funded Genetic Association Information Network (GAIN) was established in 2006 to conduct GWA studies of six common diseases, the program's data release policies reflected a compromise among data generators and data users.¹⁹⁶ Data generators agreed to "immediate" release of data generated by the project, but for the first time a temporal restriction was also placed on *users* of the data.¹⁹⁷ That is, in order to secure a period of exclusive use and publication priority for data generators, data users were prohibited from submitting abstracts and publications and making presentations based on GAIN data for a specified "embargo" period¹⁹⁸ generally fixed at nine months.¹⁹⁹ This embargo period is an example of rights latency.

Shortly thereafter, NIH adopted a similar embargo-based approach in its institute-wide policy regarding the generation, protection and sharing of data generated by federally funded GWA studies (NIH GWAS Policy).²⁰⁰ The NIH GWAS Policy states that users of GWA data should refrain from submitting their analyses for publication, or otherwise presenting them publicly, during an "exclusivity" period of up to twelve months from the date that the data set is first made available.²⁰¹ While the agency expresses a "hope" that "genotype-phenotype associations identified through NIH-supported and NIH-maintained GWA data sets and their obvious implications will remain available to all investigators, unencumbered by

196. *See generally* The GAIN Collaborative Research Grp., *New Models of Collaboration in Genome-Wide Association Studies: The Genetic Association Information Network*, 39 NATURE GENETICS 1045 (2007).

197. *Id.* at 1048 box 1.

198. GENETIC ASS'N INFO. NETWORK (GAIN), DATA USE CERTIFICATION AGREEMENT 4–5 (Dec. 3, 2008) [hereinafter GAIN DATA USE AGREEMENT], *available at* http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?view_pdf&stacc=phs000016.v1.p1 (Terms of Access, no. 8).

199. *Id.*; The GAIN Collaborative Research Grp., *supra* note 196, at 1049.

200. *See* Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS), 72 Fed. Reg. 49,290 (Aug. 28, 2007) [hereinafter NIH GWAS Policy].

201. *Id.* at 49,296. This exclusivity period was originally proposed to be nine months, but was subsequently lengthened following NIH's receipt of public comments from data generators and others. *See* Request for Information (RFI): Proposed Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS), 71 Fed. Reg. 51,629 (Aug. 30, 2006).

intellectual property claims,”²⁰² the agency stops short of prohibiting the patenting of resulting discoveries.

Embargo mechanisms have also been implemented in subsequent genomic and related data release policies.²⁰³ At a 2009 summit in Toronto, more than one hundred scientists, scholars, and funding agency representatives met to assess the current state of rapid pre-publication data release and the applicability of the Bermuda Principles to projects well beyond the generation of genomic sequence data.²⁰⁴ The participants reaffirmed a general commitment to rapid pre-publication data release, but at the same time acknowledged the acceptability of a “protected period” (rights latency) of up to one year during which data users may be restricted from publishing analyses of released data sets.²⁰⁵

f) Private Sector Policies

A noteworthy parallel to the government-sponsored projects discussed above is that of private-sector initiatives in the genome sciences. Pharmaceutical giant Merck organized the first of these in 1994, when it established a large public database of expressed sequence tags (ESTs). The stated purpose of the so-called Merck Gene Index was to increase the availability of basic knowledge, and, in part, the likelihood of discovery in support of proprietary therapeutic innovations.²⁰⁶ Another important, but less publicized, motivation for placing the EST data into the public domain was reputedly to preempt the patenting of these genetic sequences by private companies such as Incyte Pharmaceuticals and Human Genome Sciences.²⁰⁷

202. NIH GWAS Policy, *supra* note 200, at 49,296.

203. *See, e.g.*, ENCODE CONSORTIA DATA RELEASE, DATA USE, AND PUBLICATION POLICIES, *available at* <http://www.genome.gov/Pages/Research/ENCODE/ENCODEDAtaReleasePolicyFinal2008.pdf> (implementing a nine-month embargo period). *But see Human Microbiome Project (HMP): Overview*, THE NIH COMMON FUND, <http://nihroadmap.nih.gov/hmp/index.asp> (last visited Oct. 23, 2010) (implementing a data release policy for a recently initiated project to generate genomic sequence and related data for the thousands of microorganisms residing within the human body which contains no “embargo” period, but permitting data generators to withhold certain “analyses”—excluding sequence or related data—for a period of twelve months following generation).

204. *See* Toronto Authors, *supra* note 124.

205. *Id.* at 170.

206. *See* Press Release, Merck & Co., First Installment of Merck Gene Index Data Released to Public Databases: Cooperative Effort Promises to Speed Scientific Understanding of the Human Genome (Feb. 10, 1995), *available at* <http://www.bio.net/bionet/mm/bionews/1995-February/001794.html>.

207. *See* DON TAPSCOTT & ANTHONY D. WILLIAMS, WIKINOMICS: HOW MASS COLLABORATION CHANGES EVERYTHING 166 (2006); Marshall, *supra* note 168, at 1192; Rai, *supra* note 4, at 134 (describing the commercial strategy of Merck in releasing the EST data

A similar effort known as the SNP Consortium was formed in 1999 by a group of companies and the Wellcome Trust to identify and map genetic markers referred to as “single nucleotide polymorphisms” (SNPs) and to release the resulting data to the public domain, unencumbered by patents.²⁰⁸ The consortium accomplished this goal by filing U.S. patent applications covering the SNPs it discovered and mapped, and then ensuring that these applications were contributed to the public domain prior to issuance.²⁰⁹ This approach ensured that the consortium’s discoveries would act as prior art defeating subsequent third-party patent applications, with a priority date extending back to the initial filings. The SNP Consortium’s innovative “protective” patenting strategy has been cited as a model of private industry’s potential to contribute to the public genome commons.²¹⁰

Since the successful completion of the SNP Consortium project, numerous other privately funded research collaborations have adopted data release models that are similarly intended to place large quantities of genomic data into the public domain. In recent years, however, these efforts have included timing mechanisms in their data release policies. For example, the International SAE Consortium (SAEC) was formed in 2007 to fund research toward the identification of DNA markers for drug-induced serious adverse events. The SAEC adopted a “defensive” patent filing strategy similar to that of the SNP Consortium, which secures for data-generating scientists a period of exclusivity during which they have the sole ability to analyze data and prepare papers for publication.²¹¹ Like the policies adopted by government-

to the public); *see also* discussion of ESTs and the patenting debate surrounding them, *supra* note 188 and accompanying text. By 1998, the Merck Gene Index had released over 800,000 ESTs through GenBank. *See* TAPSCOTT & WILLIAMS, *supra*, at 166.

208. *See* TAPSCOTT & WILLIAMS, *supra* note 207, at 168 (noting consortium members’ concerns about biotech companies’ plans to patent SNPs and “sell them to the highest bidder”); Arthur L. Holden, *The SNP Consortium: Summary of a Private Consortium Effort to Develop an Applied Map of the Human Genome*, 32 *BIO TECHNIQUES* 22, 26 (2002) (“The overall IP objective is to maximize the number of SNPs that (i) enter the public domain at the earliest possible date, and, (ii) are free of third-party encumbrances such that the map can be used by all without financial or other IP obligations.”).

209. The SNP Consortium’s patenting strategy included filing patent applications covering all mapped SNPs and then converting those applications into statutory invention registrations (SIRs) or abandoning the applications after publication.

210. *See, e.g.*, Cook-Deegan & McCormack, *supra* note 176, at 217 (describing the consortium’s “unusual and sophisticated approach to keeping data in the public domain”); Marshall, *supra* note 168, at 1192 (noting the consortium’s “defensive move” deriving from the Merck Gene Index’s earlier strategy); Nunnally, *supra* note 128, at 252–53.

211. INT’L SAE CONSORTIUM LTD., DATA RELEASE AND INTELLECTUAL PROPERTY POLICY (last amended Nov. 5, 2009) (on file with author).

funded projects, the SAEC policy imposes a nine-month embargo on publication or presentation of publicly released data. But in addition the SAEC implements a delayed-release principle, allowing the data-generating group to retain data internally (i.e., to share it only among the SAEC participants) for a period of up to twelve months following data validation. Similarly, pharmaceutical manufacturer Pfizer and the non-profit Sage Bionetworks have recently entered into a collaboration in which Pfizer will reportedly make large amounts of its proprietary data available through Sage's "open access" mechanisms, but may retain data generated from the collaboration for a period of one year after the project is completed.²¹² The significance of these data "retention" approaches and their means of addressing policy objectives using a latency-based approach that differs significantly from the "embargo" approach will be discussed in Section V.A, *infra*.

B. SCIENTIFIC PUBLISHING

1. *Scientific Publishing and the Open-Access Debate*

The data release policies relevant to the genome commons focus on the release of experimental data prior to the publication of analytical results in a scientific journal. But an equally if not more contentious debate exists regarding the dissemination of scientific knowledge *after* it has been published. The debate centers on the accumulated scientific knowledge that is accessible only in proprietary scientific journals. According to one estimate, there were approximately 50,000 different scientific journals in print at the end of 2003,²¹³ many of which are published by commercial entities that charge significant subscription fees.²¹⁴ The proliferation of scientific journals coupled with high subscription costs has led to the cancelation of multiple journal subscriptions by academic libraries, often depriving researchers at

212. Stephen Strauss, *Pharma Embraces Open Source Models*, 28 NATURE BIOTECHNOLOGY 631, 632 (2010).

213. JOHN WILLINSKY, THE ACCESS PRINCIPLE: THE CASE FOR OPEN ACCESS TO RESEARCH AND SCHOLARSHIP 14 (2006) (citing Carol Tenopir).

214. According to one study, the average subscription cost of commercially published journals in the field of economics in 2001 was over \$1,600. Theodore C. Bergstrom, *Free Labor for Costly Journals?*, 15 J. ECON. PERSP. 183, 183 (2001). Specialist publications, particularly in the medical literature, can cost in the range of \$20,000 per year. Pamela Burdman, *A Quiet Revolt Puts Costly Journals on Web*, N.Y. TIMES, June 26, 2004, at B7 (citing the annual subscription rates of *The Journal of Comparative Neurology* (\$17,995) and *Brain Research* (\$21,269)).

those institutions of access to the current literature in their fields.²¹⁵ And because scientific publishers typically acquire the copyright in articles that they publish,²¹⁶ authors have traditionally been constrained in their ability to disseminate their published work outside of the publishers' established channels, whether in print or online.

Opposition to the influence that publishers have exerted over the dissemination of scientific data has led to a vocal and powerful "open access" movement among academic scholars and librarians.²¹⁷ They argue that the increasing control exerted by publishers over scientific results, and the decreasing number of scientists who can access these results in proprietary databases and publications, hinders the progress of science.²¹⁸ Accordingly, in 2003, a number of European academic and scientific institutions published the "Berlin Declaration," encouraging researchers to use open-access publishing methods in lieu of traditional subscription-based publication.²¹⁹ The largest private funders of biomedical research in the United States and Britain, the Howard Hughes Medical Institute and the Wellcome Trust, made a similar plea in the so-called "Bethesda Statement."²²⁰ Shortly thereafter, several major research institutions including Harvard University, the Massachusetts Institute of Technology, and University College London implemented policies requiring their faculty to deposit all research publications in open-access databases, even if previously published in a commercial journal.²²¹ This trend has since spread to institutions across the United States, Europe, Canada, Australia, and India.²²² The open-access

215. See WILLINSKY, *supra* note 213, at 24–25; NAS, RESEARCH DATA, *supra* note 82, at 78.

216. NAS, RESEARCH DATA, *supra* note 82, at 78.

217. "Open access" publishing typically connotes distribution through an online repository that is accessible to the public without charge. See generally Nancy Kranich, *Countering Enclosure: Reclaiming the Knowledge Commons*, in KNOWLEDGE AS A COMMONS, *supra* note 18 (explaining that the number of institutions globally that have adopted open-access publishing policies has doubled in the past year); WILLINSKY, *supra* note 213, at app. A; Peter Suber, *Creating an Intellectual Commons Through Open Access*, in KNOWLEDGE AS A COMMONS, *supra* note 18, at 172.

218. See sources cited *supra* note 217.

219. BERLIN DECLARATION ON OPEN ACCESS TO KNOWLEDGE IN THE SCIENCES AND HUMANITIES (2003), available at http://oa.mpg.de/files/2010/04/berlin_declaration.pdf.

220. Press Release, Peter Suber, Dir., Open Access Project, Bethesda Statement on Open Access Publishing (June 20, 2003), available at <http://www.earlham.edu/~peters/fos/bethesda.htm>.

221. See *Open-Access Publishing Gains Another Convert*, 459 NATURE 627 (2009); John Timmer, *MIT to Make All Faculty Publications Open Access*, ARS TECHNICA, Mar. 24, 2009.

222. NAS, RESEARCH DATA, *supra* note 82, at 79; Jocelyn Kaiser, *Free Journals Grow Amid Ongoing Debate*, 329 SCI. 896, 897 (2010) [hereinafter Kaiser, *Free Journals*].

directives adopted by these institutions directly contradict the publication policies of most scientific journals, and as such have led to difficult negotiations between publishers and universities adopting this approach. In many cases, publishers have reached a compromise with universities providing for a six- to twelve-month “embargo” period following publication before an article can be deposited in an open-access repository.²²³

2. *Open Access and Government-Funded Research*

An even more contentious debate exists surrounding the publication of government-funded research. According to one estimate, research funded by NIH alone results in approximately 60,000 scientific papers per year.²²⁴ Beginning in 2003, a growing number of scientists, archivists, and policy makers began to express the view that it was inappropriate for taxpayer-funded research to inure solely to the financial benefit of publishers, and that the public should have free access to taxpayer-funded research.²²⁵ Publishers, however, contend that federally funded research is separate from the subsequent publication of research results, and that the publication process, which encompasses both peer review and editorial oversight, is not subsidized by federal grant monies.²²⁶ They argue that “open access” publishing models, in which oversight and peer review may be less rigorous, permit less deserving research to be published.²²⁷

Nevertheless, in June 2004, the U.S. House Appropriations Committee instructed NIH to make all scientific publications generated by NIH-funded research available online without charge.²²⁸ The NIH implemented a voluntary “open access” policy in 2005, which was made mandatory in 2008.²²⁹ The current NIH policy provides that all publications arising out of

223. NAS, RESEARCH DATA, *supra* note 82, at 79.

224. WILLINSKY, *supra* note 213, at 2.

225. *Id.* It is useful here to differentiate between the information commons created by information accessible to the scientific community that subscribes to a given scientific journal and the broader public and scientific community. As noted by Reichman and Uhler, *supra* note 61, and others, the former community may be significantly smaller than the latter.

226. See Jeannie Baumann, *House Oversight Hearing Panelists Debate Open Access to Federally Funded Research*, BNA BIOTECH WATCH, Aug. 4, 2010 (citing Allan Adler of the Association of American Publishers).

227. Kaiser, *Free Journals*, *supra* note 222, at 896–97.

228. See Suber, *supra* note 217, at 203 n.41.

229. Nat'l Insts. of Health, Revised Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research, NOT-OD-08-033 (Apr. 7, 2008), available at <http://grants.nih.gov/grants/guide/notice-files/not-od-08-033.html> (implementing Consolidated Appropriations Act, 2008, Pub. L. No. 110-161, § 218, 121 Stat. 1844, 2187 (2007)).

NIH-funded research data must be submitted to the National Library of Medicine's publicly accessible PubMed Central digital archive within one year following publication.²³⁰ The one-year period established by the NIH policy provides the publishing industry with a period of exclusive article access, to avoid the loss of paid subscription revenue.²³¹ This period reflects a compromise among publishers who opposed the measure in general, NIH, which initially proposed a six-month period, and open-access advocates, who argued for immediate public release of articles.²³² Funding agencies in Europe have enacted mandates similar to the NIH policy.²³³ Legislation recently proposed in the U.S. Congress would expand the NIH's open-access mandate to all federal agencies with a shortened embargo period of six months.²³⁴ Not surprisingly, legislation supported by the publishing industry and opposing both initiatives has also been proposed in Congress.²³⁵

3. *Alternative Open-Access Models*

The scientific publishing industry has not, however, uniformly opposed open-access initiatives and a few publishers have even embraced them. Journals such as the *New England Journal of Medicine (NEJM)* and *Molecular Biology of the Cell (MBC)* now voluntarily make their contents publicly available after a waiting period (six months in the case of *NEJM*; two months in the case of *MBC*).²³⁶ *NEJM* and *MBC*, however, are both published by scientific trade associations,²³⁷ and thus serve their members through multiple channels, of which journal publication is only one. The largest publishers of scientific journals, Reed Elsevier (approximately 1,800 titles), Taylor and Francis (more than 1,000 titles), and Springer Verlag (more than 500 titles), which collectively control sixty percent of scientific research content, are

230. It has been estimated that, as of April 2010, PubMed Central contains two million scholarly articles. Declan Butler, *US Seeks to Make Science Free for All*, 464 NATURE 822, 822 (2010).

231. *Id.*

232. WILLINSKY, *supra* note 213, at 3; Editorial, *Open Sesame*, 464 NATURE 813 (2010).

233. Butler, *supra* note 230, at 822.

234. Federal Research Public Access Act (FRPAA), H.R. 5037, 111th Cong. (2010).

235. Fair Copyright in Research Works Act, H.R. 801, 111th Cong. (2009) (would prohibit federal agencies from adopting open-access publication policies). For a discussion of this legislative proposal, see, for example, James Boyle, *Misunderestimating Open Science*, FIN. TIMES, Feb. 24, 2009. A representative of publisher Reed Elsevier recently called the proposed open-access requirements "a means for facilitating international piracy." Butler, *supra* note 230, at 822.

236. WILLINSKY, *supra* note 213, at 68.

237. *NEJM* is published by the Massachusetts Medical Society and *MBC* by the American Society for Cell Biology.

commercial interests with significant subscription and reprint revenues at stake.²³⁸

One open-access approach that could potentially satisfy the revenue requirements of commercial publishers is the so-called “author pays” model, in which articles are made freely accessible to readers, but the journal charges authors a fee to cover editorial, publishing, and peer review costs.²³⁹ The author-pays model was pioneered by the non-profit Public Library of Science (PLOS), which operates seven journals covering biomedical topics such as genetics, biology, and medicine, and which was co-founded by Nobel laureate and former NIH Director Harold Varmus.²⁴⁰ Although the author-pays model has still not made significant inroads in most scientific disciplines, PLOS’s flagship journal *PLoS ONE* has seen significant increases in submissions since its foundation in 2006.²⁴¹ Publication fees for PLOS journals range from \$1,350 (for *PLoS ONE*) to \$2,900 (*PLoS Biology* and *PLoS Medicine*),²⁴² and an increasing number of funding agencies and foundations appear to be willing to pay open-access publication fees for research that they fund.²⁴³

Commercial publishers, perhaps sensing the prevailing wind in support of open access, have also taken tentative steps toward open-access

238. WILLINSKY, *supra* note 213, at 18.

239. This approach should be distinguished from the customary practice of many professional societies (such as the American Physiological Society) to charge authors publication fees in order to defray member subscription costs. See Dale J. Benos, L. Gabriel Navar & Margaret Reich, *Publishing in the Journals of the APS: Why are Authors Charged Fees?*, 278 AM. J. PHYSIOLOGY GASTROINTESTINAL LIVER PHYSIOLOGY 663, 663 (2000) (“Like many other association publishers, APS is able to keep subscription prices low by sharing some of the cost of publishing the journals with the authors who submit manuscripts. . . . Many commercial publishers do not charge authors for publication (i.e., page charges) but have much higher subscription prices.”); see also *Information for Authors*, PROCEEDINGS NAT’L ACAD. SCI. (Oct. 2010), <http://www.pnas.org/site/misc/iforc.shtml> (detailing fees per page and per color figure); *Instructions to Authors*, J. VIROLOGY, http://jvi.asm.org/misc/journal-ita_pub.dtl (last updated Sept. 2010) (detailing fees per page and per color figure).

240. See PUB. LIBR. SCI., <http://www.plos.org> (last visited Nov. 8, 2010); Kaiser, *Free Journals*, *supra* note 222, at 897; Kranich, *supra* note 217, at 96.

241. See Butler, *supra* note 230, at 823.

242. *Publication Fees for PLoS Journals*, PUB. LIBR. OF SCI., <http://www.plos.org/journals/pubfees.php> (last visited July 4, 2010).

243. Butler, *supra* note 230, at 823. A group of major research universities including Harvard, M.I.T., Dartmouth, Cornell, Berkeley, Columbia, Memorial Sloan-Kettering, and the University of Ottawa have formed a group called the Compact for Open-Access Publishing Equity to advocate for greater payment of open-access publication fees by research funders. COMPACT FOR OPEN-ACCESS PUBLISHING EQUITY, <http://www.oecompact.org/compact/> (last visited Oct. 23, 2010).

publishing. In 2008, for example, the German publisher Springer Verlag acquired BioMed Central, which publishes more than 200 peer-reviewed open-access journals.²⁴⁴ Other commercial publishers, such as Nature Publishing Group, have experimented with a hybrid approach under which authors have the option either to publish in the traditional manner or to pay to make individual articles available on an open access basis.²⁴⁵ Nevertheless, the largest scientific publishers continue to oppose legislation and policies designed to mandate open access for government-funded research.²⁴⁶

V. APPLYING THE LATENCY FRAMEWORK

The two case studies presented in Part IV illustrate well-known examples of the construction of science commons. I have chosen to highlight them because they exemplify the use of temporal features by policy designers to achieve specific policy outcomes in a private ordering context and to reach compromise among stakeholders with competing interests. In this Part, I will apply the latency analysis introduced in Part III to these case studies.

A. LATENCY IN THE GENOME COMMONS

1. *Stakeholders and Policy Design Considerations*

The primary stakeholders engaged in the policy debate over data release in the genome commons are funding agencies (primarily the U.S. NIH and the Wellcome Trust in the UK), data generators, and data users.²⁴⁷ More recently, the critical role of scientific journals has been recognized, particularly with respect to the publication of articles that may be in violation of embargo requirements and the need to offer meaningful and career-enhancing publication opportunities to data generators.²⁴⁸ While data release policies are typically drafted and adopted by funding agencies, NIH in particular has given substantial deference to the views and opinions of the scientific community when developing policy, while also seeking to represent

244. See BIOMED CENT.: THE OPEN ACCESS PUBLISHER, <http://www.biomedcentral.com> (last visited Nov. 8, 2010).

245. See Press Release, Nature Publ'g Grp., Open Access Options on 7 More Nature Publishing Group Journals (May 19, 2010), available at http://www.nature.com/press_releases/openaccess.html.

246. See Baumann, *supra* note 226.

247. See Ft. Lauderdale Principles, *supra* note 189, at 2–3 (referring to the “tripartite” sharing of responsibility among these groups). Because most large-scale genomic databases are operated by governmental agencies (e.g., NIH’s GenBank and dbGaP), the interests of such data intermediaries are typically aligned with those of governmental funders.

248. See *id.* at 4; Toronto Authors, *supra* note 124, at 170.

the interests of the general public.²⁴⁹ Thus, the role and influence of other stakeholder groups is not to be underestimated: the development of data release policies in the genome sciences has been a process of negotiation and compromise.

The evolution of the genome commons illuminates three principal policy considerations that motivated these stakeholder groups: (1) promoting the advancement of science by making genomic data as widely available as possible (“scientific advancement,” typically espoused by funders and public-interest advocates); (2) giving data generators priority in the publication of papers based on their data (“publication priority,” typically espoused by data generators); and (3) minimizing patent-related encumbrances on genomic data sets (“minimizing encumbrances,” espoused by both funders and data users).²⁵⁰ The interplay of these three design considerations, and the latency-based compromises that were effected to satisfy competing requirements of relevant stakeholders, resulted in the policies that are in effect today.

249. See, e.g., Collins et al., *supra* note 178, at 846 (2003) (describing community involvement in setting goals for national genomics research program). NIH has most recently requested public comment and feedback on data release policies in October 2009. Press Release, Nat’l Insts. of Health, Notice on Development of Data Sharing Policy for Sequence and Related Genomic Data, NOT-HG-10-006 (Oct. 19, 2009), available at <http://grants.nih.gov/grants/guide/notice-files/NOT-HG-10-006.html>.

250. Other factors such as protection of human subject data also played a substantial role in the development of genomic data release policies. See *supra* note 113 and accompanying text. These factors, however, have remained relatively invariant and not subject to compromise, and are thus not addressed in detail here.

Table 1: Latency Characteristics of Genomics Data Release Policies

	Knowledge Latency (KL)	Rights Latency (RL)
U.S. Federally Funded Genomics Policies²⁵¹		
NIH-DOE 1992 Guidelines ²⁵²	6 months	0
Bermuda Principles (1996) ²⁵³	24 hours	0
NIH 1996 Policy ²⁵⁴	Short ²⁵⁵	0
NIH 1997 Policy ²⁵⁶	24 hours	0
NIH 2000 Policy ²⁵⁷	24 hours–7 days	0
NIH 2003 Policy ²⁵⁸	24 hours	0
Intl. HapMap Project (2003) ²⁵⁹	Short	0
ENCODE Pilot (2003) ²⁶⁰	Short	0
GAIN (2006) ²⁶¹	Short	9 months (generally)
NIH GWAS Policy (2007) ²⁶²	Short	12 months
ENCODE + modENCODE (2008) ²⁶³	Short	9 months
Non-Federal Genomics Policies		
SNP Consortium (1998) ²⁶⁴	30/90 days	0
Intl. SAE Consortium (2007) ²⁶⁵	12 months	9 months

251. As discussed *supra* Section IV.A.1, many of the NIH and other policies phrase timing requirements as encouragements or suggestions. While the enforceability of such “soft” requirements is open to question, for purposes of this analysis, the timing features of these policies are listed notwithstanding potential issues with enforcement.

252. *See supra* note 170.

253. *See supra* note 35.

254. *See supra* note 188.

255. A number of policies include requirements that data be released as rapidly as possible following validation or the satisfaction of some other initial condition. This requirement can be translated into a numerical value depending on the particular type of data involved and customary practice in the field. For purposes of this Article, however, these policies should simply be viewed as having quite short knowledge latencies, probably on the order of one month or less.

256. *See supra* note 172 and accompanying text.

257. *NHGRI Policy for Release and Database Deposition of Sequence Data*, NAT’L HUM. GENOME RES. INST. (Dec. 21, 2000), <http://www.genome.gov/page.cfm?pageID=10000910>.

258. *Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-Scale Sequencing and Other Community Resource Projects*, NAT’L HUM. GENOME RES. INST. (Feb. 2003), <http://www.genome.gov/10506537>.

259. The Int’l HapMap Consortium, *The International HapMap Project*, 426 NATURE 789, 793 (2003) (“All data . . . will be released rapidly.”).

260. *ENCODE Project Data Release Policy (2003–2007)*, NAT’L HUM. GENOME RES. INST., <http://www.genome.gov/12513440> (last updated Apr. 20, 2010).

261. *See supra* notes 198–201.

262. *See supra* note 200.

263. *ENCODE CONSORTIA DATA RELEASE, DATA USE, AND PUBLICATION POLICIES*, <http://www.genome.gov/Pages/Research/ENCODE/ENCODEDataReleasePolicyFinal2008.pdf> (last visited Oct. 23, 2010).

264. *See supra* notes 208–09.

265. *See supra* note 211.

Pfizer – SAGE (2010) ²⁶⁶	12 months	0
-------------------------------------	-----------	---

2. Latency as a Policy Design Tool—Retention versus Embargo

As demonstrated by Table 1, in every NIH data release policy following the adoption of the Bermuda Principles, genomic data must be released rapidly to public databases. The twofold policy motivations for this requirement have been discussed above: there is the explicit motivation of accelerating the progress of science, and a less explicit, but strongly implied, motivation of limiting patent encumbrances on genomic data. However, once concerns regarding publication priority between data generators and data users arose, there was a need for policy change. Most other federal agencies and private initiatives in the genome sciences have addressed this conflict by permitting data generators to withhold their data from the public for a specified time period, generally in the range of twelve months,²⁶⁷ after which it would be released unencumbered (this can be called a “retention strategy”). In terms of latency analysis:

Retention Strategy

knowledge latency (KL) = 12 months

rights latency (RL) = 0

NIH policy makers, however, elected to take an alternative approach. Instead of allowing data generators to retain their data for a protected period and then releasing it unencumbered, each of the new NIH policies continues to require rapid data release, but imposes a publication embargo on users (an “embargo strategy”). Thus, in the case of the GAIN NIH GWAS Policy and other recent NIH policies:

Embargo Strategy

knowledge latency (KL) = 0/short

rights latency (RL) = 12 months

These two strategies are summarized in Table 2.

266. See *supra* note 212.

267. See *supra* note 173 and accompanying text.

Table 2: Latency Analysis of Genomic Data Release Policies

	Knowledge Latency (KL)	Rights Latency (RL)
Bermuda	~0	0
Embargo (GAIN/GWAS)	~0	12 mo.
Retention (SAGE, SAEC)	12 mo.	0+

On paper, the result for data generators appears to be the same under both an embargo and a retention strategy: in either case they have a period of twelve months during which they retain exclusive rights to publish papers analyzing the data. But in practice there are material differences between the two strategies.²⁶⁸ These differences are driven by material externalities that distinguish government-funded projects from privately funded projects.

A retention strategy lengthens knowledge latency and, by definition, extends the time before data is released to the public. NIH has repeatedly stated that genomic data should be released as rapidly as possible for the advancement of science and the public good.²⁶⁹ The embargo approach accomplishes this goal by minimizing knowledge latency while still protecting the data generators' publication interests. However, the embargo strategy involves a significant tradeoff in terms of effectiveness. Usage embargos in NIH's recent data release policies are typically embodied in click-wrap agreements or online certifications that must be made with a data request.²⁷⁰ A discussion of the enforceability of these mechanisms (whether in strictly legal terms or in terms of the ability and willingness of journals and funders to police compliance) is beyond the scope of this Article.²⁷¹ However, it is self-evident that even the *most* robust contractual embargo provides the data generator with *less* protection than withholding the data from the public (i.e., if the user has no data, he or she cannot breach his or her obligation to

268. The distinction between these two approaches is illustrated *supra* Figures 2 and 3.

269. See *supra* note 178 and accompanying text.

270. A "click-wrap" agreement (alternatively referred to as a "click-through" or "click-to-accept" agreement or license) is "an electronic form agreement to which [a] party may assent by clicking an icon or a button or by typing in a set of specified words." Christina L. Kunz et al., *Click-Through Agreements: Strategies for Avoiding Disputes on Validity of Assent*, 57 BUS. LAW. 401, 401 (2001).

271. For a general discussion of the enforceability of click-wrap agreements, see GEORGE B. DELTA & JEFFREY H. MATSUURA, LAW OF THE INTERNET, § 10.05 (2d ed. Supp. 2010–2012). Rebecca Eisenberg, who analogizes click-wrap agreements for genomics projects to the open source software General Public License (GPL), raises questions about the enforceability of such agreements. Eisenberg, *supra* note 101, at 1028. In addition to legal enforceability, serious issues of detection and policing arise when considering the efficacy of obligations imposed in this manner.

refrain from publishing). Moreover, a retention strategy gives the data generator a true head start with respect to the data, during which no third party may analyze or build upon it (witness the seventeen-year holdback of *Ardi* data). In contrast, an embargo strategy enables third parties both to analyze and to build upon the data *during* the embargo period, putting them in a position to submit papers for publication the moment the embargo expires. Finally, under a rapid-release embargo system, data generators run the risk that journals less familiar with rapid pre-publication data release will consider the initial release of data in a public database to constitute a prior publication that would disqualify the data generator from subsequently publishing his or her analysis.²⁷²

With all of these comparative disadvantages to the data generator, why do the NIH policies adopted since 2006 emphasize rights latency over knowledge latency? The answer may lie in NIH's approach to intellectual property. As discussed above, NIH must operate within the constraints of the Bayh-Dole Act. Thus, while NIH's various post-Bermuda data release policies all acknowledge the requirements of the Bayh-Dole Act, they evidence a general bias against patent encumbrances on genomic data.²⁷³ The enforceability, however, of policy provisions that merely "urge" or "encourage" data generators and users not to seek patents on inappropriate subject matter is open to some doubt.²⁷⁴ Lacking a strong policy tool with which to limit expressly the patenting of genomic information, NIH policy makers have employed rapid pre-publication data release requirements as a surrogate for achieving the same result.²⁷⁵ The Bermuda Principles, in

272. Most scientific journals will only accept articles for publication if they have not previously been published or released. While journals specializing in genomics and allied fields, or which are otherwise actively engaged in the conversation regarding pre-publication data release, would ordinarily not consider pre-publication data release to constitute a prior publication for this purpose, the author has encountered anecdotal evidence suggesting that journals that are less familiar with the practice have rejected articles by data generators on this basis.

273. See, e.g., *NHGRI 1996 Policy*, discussed *supra* notes 187–88 and accompanying text; *ENCODE Pilot Policy*, discussed *supra* note 203; *NIH GWAS Policy* discussed *supra* Section IV.A.1.e.

274. See Rai & Eisenberg, *supra* note 187, at 309.

275. An interesting early exception is the International HapMap Project, which was organized in 2002 by scientists and funding agencies from Japan, the United Kingdom, Canada, China, Nigeria, and the United States to develop a "haplotype" map of the human genome. See The Int'l HapMap Consortium, *supra* note 259. Data generated by the project were released "rapidly" into publicly accessible databases, but a click-wrap agreement that purportedly prohibited data users from filing patent applications on project results was also used. *Id.* at 793. Though this precise history is somewhat unclear, Rebecca Eisenberg reports

particular, and their adoption by NIH in 1997 and reaffirmation in 2003, ensured that genomic data from the HGP and other large-scale sequencing projects would be made publicly available before data generators had an opportunity to file patent applications covering any “inventions” arising from that data, and in a manner that ensured its availability as prior art against third party patent filings at the earliest possible date.²⁷⁶

In contrast, non-governmental groups such as SAEC adopted lengthier knowledge latency periods to protect the publication priority of their researchers more effectively, but they did so in conjunction with explicit patent-defeating strategies.²⁷⁷ These groups, unlike federal agencies, have the freedom to impose express contractual limitations on patenting and to adapt defensive patenting strategies without running afoul of the requirements of the Bayh-Dole Act. Thus these policies may be most optimal with respect to the policy goals of minimizing encumbrances and protecting data generators’ publication priority, but are less optimal than the government-led policies in terms of broad disclosure of knowledge and scientific advancement. The trade-offs may reflect the different policy objectives of government NIH funders versus private commercial-consortium organizers. Table 3 summarizes the differing policy outcomes achieved by adjustment of latency variables in the genome commons.

that a formal Data Release Policy was adopted by the project as late as 2004 and was intended to supersede the click-wrap structure. Eisenberg, *supra* note 101, at 1026. It is possible that this policy was implemented in response to objections by federal funding agencies to the earlier contractual patenting prohibitions.

276. See discussion *supra* notes 177–82 and accompanying text.

277. See discussion *supra* notes 210–11 and accompanying text.

Table 3: Policy Outcomes in of Genomic Data Release Policies

	Scientific Advancement (SA) (1)	Minimizes Encumbrances (ME) (2)	Publication Priority (PP) (3)
Bermuda (KL=0/RL=0)	High	Medium	Low
Embargo (KL=0/RL=1)	High	Medium	Medium
Retention (KL=1/RL=0)	Medium	High	High

- (1) Scientific Advancement (SA) is highest under policies that minimize knowledge latency (KL). While data is retained by data generators in private, overall scientific advancement cannot occur. Thus, under a retention strategy, SA is classified as “Medium” (as opposed to “Low,” given that data *is* released to the public, albeit after a waiting period).
- (2) Minimization of encumbrances (e.g., patent protection on data) is achieved by two different means. Private groups implementing retention strategies have employed contractual anti-patenting policies coupled with “protective” patent filings, resulting in a high degree of freedom from patent encumbrances. Government-funded projects, which cannot avail themselves of these techniques, are left with “hortatory” encouragements against patenting, coupled with early disclosure of information as prior art (low KL). While the effectiveness of these measures is debatable, they are likely not as strong as the approach taken under the retention strategy, hence a “Medium” rating.
- (3) Publication priority (PP) for data generators was explicitly sacrificed under public projects such as the HGP in the service of scientific advancement. While embargo policies have attempted to improve priority for data generators, the enforceability of contractual embargo provisions is less certain than simple withholding of information under a retention policy.

B. LATENCY AND SCIENTIFIC PUBLISHING

1. Stakeholders and Policy Design Considerations

In contrast to the policy negotiations over the genome commons, which, despite their far-reaching implications for biomedical research, directly affect a relatively discrete community of scientists and funders, the debate over scientific publishing implicates all scientific disciplines and many other fields of scholarly research including the humanities and arts. As such, the community of stakeholders relevant to this debate is large and varied. Nevertheless, the stakeholder categories identified in Section III.A.3 apply to this analysis as well.

As in every information commons, data generators play a central role. In this context, data generators are the scientists who have an interest in publishing their results for purposes of career and reputational enhancement

(i.e., the vast majority of scientists).²⁷⁸ Unlike the genome commons, in which the communities of data generators and data users are more or less distinct (though converging as the cost of genomic sequencing continues to decrease), data-generating scientists in the context of scientific publishing are nearly always data users as well. Thus, in addition to the need to publish their own work, scientists must also access and utilize the published work of others. Research institutions are typically responsible for ensuring that their scientists have access to sufficient information resources, typically through their library function. The library community has been particularly vocal in the scientific publishing debate, and has decried rising subscription costs, publishing practices, and legal regimes that have led to so-called “enclosures” of the science commons and the inaccessibility of current scientific information to researchers.²⁷⁹ They argue that the advancement of science is imperiled by this enclosure and the decreasing number of individuals having access to potentially important scientific discoveries and data.²⁸⁰ This view has also been expressed by government funding agencies that argue for greater public access to taxpayer-supported research. As in the case of the genome commons, NIH has taken a leading role in representing the public interest in this debate.²⁸¹ Publishers—commercial publishers in particular—are on the opposite side of this debate and have argued that the services they provide in terms of selection, peer review, editing, and distribution are costly; that financial incentives must exist if scientific publication is to continue; and that the movement toward open-access publication is likely to result in scientific articles that are less reliable and of lesser quality.²⁸² Thus, where their principal policy objective is profit-making (or, in more sympathetic terms, financial survival), publishers also invoke a policy argument sounding in scientific advancement.

2. *Compromises Effected Through Latency Choices*

While the differences among stakeholders described in the preceding section may seem insurmountable at first glance, these differences have been addressed, at least for the moment, through a series of temporal policy

278. See *supra* Section III.A.1.

279. See, e.g., Kranich, *supra* note 217, at 87–107.

280. Empirical evidence supporting this claim, however, is still somewhat sparse and inconclusive. See Kaiser, *Free Journals*, *supra* note 222, at 898.

281. See *supra* notes 228–35 and accompanying text.

282. See Editorial, *supra* note 232, at 813. By one estimate, publication costs for a single article in the most prestigious scientific journals can run up to \$10,000. Kaiser, *Free Journals*, *supra* note 222, at 897.

compromises that adjust knowledge latency²⁸³ in the relevant science commons. To recapitulate: (1) universities and publishers have negotiated limited exclusivity periods of *six-to-twelve months* before university researchers are required to release their work to the public, (2) membership organizations that publish scientific journals such as the New England Journal of Medicine and Molecular Biology of the Cell, in response to member pressures, voluntarily permit open-access release of articles following an exclusivity period of up to *six months*, (3) NIH has mandated that all publications arising from NIH-funded research be released to the PubMed Central database *one year* after publication, and (4) additional legislation that has been proposed in Congress would extend the NIH mandate to all federal agencies and reduce the holding period to *six months*. Table 4 summarizes these compromises in terms of the exclusivity periods granted to publishers.

Table 4: Latency-Based Compromises in Scientific Publishing

Policy	Exclusive Period (knowledge latency)
Publisher-University Negotiated Embargo Periods	6–12 months
New England Journal of Medicine (NEJM)	6 months
Molecular Biology of the Cell (MBC)	2 months
NIH Open Access Policy (2008)	12 months
Proposed Federal Research Public Access Act (2010)	6 months

In each of these policies, scientific publishing is conducted in its current form and the publisher enjoys a limited-duration period of exclusivity. Following this period of exclusivity, release of the published work via open-access channels is authorized. It is significant that the policies illustrated above were developed through a wide range of processes, from unilateral adoption (in the case of NEJM and MBC), to bilateral negotiation (in the case of publisher-university embargos), to executive agency action (in the case of the NIH policy), to legislative action (in the case of the proposed FRPAA).

It is also significant that publishers, scientists, government, and research institutions have seemingly accepted the adjustment of knowledge latency as a suitable vehicle for addressing the differing concerns and objectives of

283. Because no substantial limitations on the use of data contained in scientific articles are imposed by journals (other than typical prohibitions on reproduction existing under copyright law), the adjustment of rights latency has not played a significant role in the debate over scientific publishing. Of course, information published in journals may be protected by patents, but patenting is currently not addressed by journal publication policies.

these diverse stakeholder groups. Though positions regarding the desired length of holding or embargo periods still differ, it appears that the scientific publishing community is converging on a period in the range of six to twelve months as an appropriate compromise.²⁸⁴ This convergence resembles that observed with respect to the genome commons. In both instances, competing stakeholder groups have addressed seemingly intractable policy differences through the adjustment of latency variables.

C. SYNTHESIS AND BROADER APPLICATIONS

Three conclusions about the design of scientific information commons can be drawn from the above case studies. First, these examples demonstrate that multiple stakeholder groups with competing and divergent interests can, and do, achieve compromise when constructing information commons. Second, once these commons are constructed, they can and do achieve desired social benefits. Third, the resolution of many of the conflicts among stakeholder groups in constructing information commons can be addressed through the adjustment of latency variables, whether knowledge latency and rights latency in tandem (as in the genome commons), or knowledge latency alone (as in scientific publishing). Why is this the case? Why do latency variables, among the many other characteristics of information commons,²⁸⁵ exert such a strong effect on the outcome of rules negotiations, whether privately ordered or governmental, and why are they the variables, among so many others, that have attracted the attention of policy makers and other stakeholder groups?

Several potential explanations present themselves. First, latency variables are fundamental to information commons, as they define the rate at which knowledge enters, and becomes usable within, the commons. Just as the rate of depletion of a physical resource commons such as a forest, pasture, or fishery is critical to understanding and preserving such a resource, latency is critical to understanding and building a commons of information. As such, latency should play a central role in rulemaking and design with respect to

284. To be sure, some publishers still argue that a twelve-month “embargo” period is too short to enable them to recoup their publication costs and that no single time period will be uniformly acceptable. Baumann, *supra* note 226.

285. Arun Agrawal has identified twenty-four separate variables that have been considered in analyses of common pool resources. Arun Agrawal, *Sustainable Governance of Common-Pool Resources: Context, Methods, and Politics*, 32 ANN. REV. ANTHROPOLOGY 243, 254 (2003). In his response to Madison, Frischmann, and Strandburg, *supra* note 23, Gregg Macey also points to the “many-variable” problem as a weakness of functionalist accounts of commons structures. Gregg P. Macey, *Cooperative Institutions in Cultural Commons*, 95 CORNELL L. REV. 757, 768–69 (2010).

information commons and, at a minimum, should be considered in the analysis of common information resources, whether under the IAD framework, the revised framework of Madison, Frischmann, and Strandburg, or future analytical frameworks.

Second, latency variables are structurally suited to effect compromise among competing interests. They are readily quantifiable, infinitely divisible, and intuitively understandable to experts and non-experts alike.²⁸⁶ Unlike other commons characteristics that could potentially be used to achieve similar policy goals (e.g., scientific advancement, publication priority, and minimization of encumbrances) such as, for example, the limited scope of intellectual property protection and the exemptions for “fair use” and experimental use,²⁸⁷ the adjustment of well-defined temporal periods is both efficient and definite.

Third, latency variables implicitly establish that the *time* during which information resides within (or is usable within) a commons and, conversely, the time during which information is held outside of the commons (or is subject to usage limitations), has a quantifiable *value*.²⁸⁸ When a value for a commodity exists, parties can negotiate based on their own economic assessments and, as has been shown in the above case studies, strike a bargain. In this sense, latency variables are satisfyingly reductionist. They enable stakeholders to distil complex policy positions into simple numerical values that are susceptible to numerical compromise, and to negotiate on the

286. Negotiation over timing begins at an early age, from requests to postpone bedtime to the length of periods allotted for play, naps, and homework.

287. These concepts have proven to be notoriously vague and difficult foundations on which to base behavior.

288. Interestingly, the *value* of commons-based information has not been widely studied. In contrast is the value of encumbered information, notably that covered by patents and copyrights, which has been studied extensively. See, e.g., Gerald J. Mossinghoff, *Overview of the Hatch-Waxman Act and Its Impact on the Drug Development Process*, 54 FOOD & DRUG L.J. 187 (1999) (describing the compromises and motivations involved in crafting the Hatch-Waxman Act, which included patent term extensions for pioneer drugs undergoing long FDA approval processes); Avishalom Tor & Dotan Oliar, *Incentives to Create Under a “Lifetime-Plus-Years” Copyright Duration: Lessons from a Behavioral Economic Analysis for Eldred v. Ashcroft*, 36 LOY. L.A. L. REV. 437, 439–40 (2002) (arguing that the value created by copyright term extensions are not always offset by incentives to produce copyrighted works); Sabra Chartrand, *Patents; Congress Has Extended Its Protection for Goofy, Gershwin and Some Moguls of the Internet*, N.Y. TIMES, Oct. 19, 1998, at C2 (noting that the Copyright Term Extension Act of 1998 was passed with “intense lobbying for it from powerful copyright holders”); Andrew Pollack, *A Company Races to Keep a Drug Patent*, N.Y. TIMES, Mar. 19, 2010, at B1 (noting the large value of a patent extension for a drug patent, loss of which would “cost[] the company hundreds of millions of dollars in lost sales”).

basis of these ostensibly objective criteria rather than more value-laden positions such as preferences for intellectual property protection versus the public domain. The introduction of latency variables into the commons formation discourse thus enables parties to sidestep (usually fruitless) debate over entrenched values and reach agreement based on simple numerical calculation.²⁸⁹ Consequently, the use of latency variables in policy design helps to reduce negotiation transaction costs, those perennial hobgoblins of economic analysis, and hopefully to achieve efficient and equitable results for all stakeholders.²⁹⁰ And, perhaps more importantly, the achievement of agreement on the basis of latency-based compromise may enable socially beneficial information commons to be formed where otherwise they would not, or where their formation otherwise would be substantially delayed due to negotiation roadblocks.

Finally, one cannot help but speculate regarding the significance of the convergence of latency variables in the two case studies presented in this Article. In the example of the genome commons, both the “retention” approach (via knowledge latency) and the “embargo” approach (via rights latency) have resulted in twelve-month waiting periods before data is freely

289. See *supra* note 29. Roger Pielke divides political conflicts regarding science-based policy into two broad categories: “tornado politics” (conflicts that can be resolved through the application of more and better scientific data, such as the decision regarding the most prudent course of action in the face of an approaching tornado) and “abortion politics” (conflicts that are characterized by differing values systems that are unlikely to be resolved through the application of additional data, such as the debate regarding the legalization of abortion). ROGER A. PIELKE, JR., *THE HONEST BROKER: MAKING SENSE OF SCIENCE IN POLICY AND POLITICS* 40–44 (2007). The use of latency variables as surrogates for value-laden positions regarding, for example, openness and intellectual property, transforms the debate regarding information commons to one that more closely resembles Pielke’s “tornado politics” than “abortion politics.”

290. This is not to say, of course, that the reduction of deeply held values to numerical surrogates is necessarily a desirable approach in all circumstances. In the case of commons formation, however, this technique has proven to be a useful tool in overcoming negotiation roadblocks and ensuring that commons are created. It is also worth noting that the “negotiation” strategies enabled by the use of latency analysis may only be effective for stakeholder groups that are adequately represented at the negotiating table. That is, when stakeholder groups are wholly or partially disenfranchised or underrepresented (in the case, for example, of citizens of developing countries), it may be difficult for such stakeholders to represent their collective position effectively, with or without the availability of latency techniques. Fortunately, in the cases examined in this article and most other cases involving commons of scientific information, governmental policy makers have typically sought to represent the interests of the public at large. See discussion *supra* note 113 and accompanying text.

usable in the commons.²⁹¹ In the case of scientific publishing, though negotiations and debate are ongoing, it appears that a six- to twelve-month “exclusivity” period may emerge as a satisfactory resolution for most parties. It is tempting to speculate whether these real-world latency periods represent, or at least begin to approach, values that may be deemed to be “efficient” or “optimal,” and whether the final values of these latency periods can tell us something about the intrinsic value of scientific knowledge in the market economy. But even if one postulated that an optimal latency value may exist for a particular scientific commons, it is likely that this value would be difficult, if not impossible, to translate to other scientific disciplines, as norms, practices, and expectations are likely to vary among scientific disciplines that are as diverse as, for example, genomics, astrophysics, botany, and paleoanthropology (as demonstrated by the case of *Ardi*). It is also likely that the balance between knowledge latency and rights latency, and the efficiencies achieved by adjusting these twin variables, will vary according to the particulars of the scientific discipline and the commons being formed, just as the approaches and policy considerations that arose in the case studies of genomics and open-access publishing differ.²⁹² I am hopeful that further

291. Upon observing this result, one is drawn to Coase’s famous theorem which postulates that, absent transaction costs, the final allocation of resources among transacting parties will be the same, no matter how legal rules may define initial entitlements. Ronald Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1 (1960). Shubha Ghosh invokes Coase’s theorem in discussing the formation of public commons of copyrighted material, explaining that it may be “largely irrelevant who has the legal rights in the creation (the creator or the public at large) as long as bargaining is allowed.” Ghosh, *supra* note 155, at 222–23. Thus, in Ghosh’s estimation, information commons will be created, whether legal rules impose protective intellectual property or public domain regimes. In the case of the genome commons, one is tempted to wonder whether it is meaningful that twelve-month data waiting periods have arisen in two different contexts: that of private consortia, which explicitly prevent patenting of results, and that of NIH-funded programs, in which patenting cannot legally be prohibited. Though further study and economic analysis are required, such an outcome suggests itself as a natural proof of Coase’s theorem in the spirit of Robert Ellickson’s landmark study of rancher-landowner interactions in Shasta County, California. ROBERT C. ELLICKSON, *ORDER WITHOUT LAW: HOW NEIGHBORS SETTLE DISPUTES* 2–3 (1991).

292. One extreme example of the importance of temporal factors to an emerging information commons can be found in nineteenth century attempts to extend copyright protection to published news and, in particular, the proposed 1884 Bill Granting Copyright to Newspapers. S. 1728, 48th Cong. (1884). The so-called News Copyright Bill reflected a realization by the Associated Press and other metropolitan news sources that significant economic value arose from being the first to print the news. See Robert F. Brauneis, *The Debate Over Copyright in News and its Effect on Originality Doctrine*, in *INTELLECTUAL PROPERTY PROTECTION OF FACT-BASED WORKS* 39, 54–56 (Robert F. Brauneis, ed. 2009). Accordingly, the Bill sought to extend to newspaper publishers an eight-hour exclusive right

application of the latency analysis presented herein to additional commons of scientific information may shed further light on the complex economic dynamics of commons formation and better enable stakeholders and policy makers to develop socially-valuable commons structures.

VI. CONCLUSION

While the sharing of data among scientists is necessary for the advancement of science and reflects a long-held norm of scientific practice, barriers and delays, legal, institutional, and interpersonal, have hindered the rapid and free sharing of scientific data. Stakeholder groups that are involved in the generation, funding, dissemination, and use of scientific data have policy objectives that are often at odds with one another and that frequently impact the design of scientific “commons.” Such objectives include the desire to use and access the findings of others, the desire to achieve recognition from one’s discoveries, the protection of commercial and financial interests, and the enablement of scientific advancement through broad availability of data.

The “latency analysis” that I propose in this Article is a tool that can be used to mediate among divergent stakeholder policy goals by defining two key variables that describe the temporal characteristics of information commons: the speed at which data enters the commons (knowledge latency) and the speed at which data within the commons becomes freely usable (rights latency). By placing latency analysis within larger theoretical frameworks for analyzing information commons, including the IAD framework championed by Ostrom and Hess, and the refined framework of Madison, Frischmann, and Strandburg, I provide a theoretical basis for the introduction and use of this methodology.

The three-part latency analysis methodology is useful not only in the analysis of existing commons structures, but in the design of new commons as well. In commons design, the introduction of latency variables to the negotiation over policy terms can effectively mediate among divergent stakeholder interests and objectives in a value-neutral manner, thus

to print and distribute the news reported in their papers. While the 1884 Bill failed, it laid the groundwork for the Supreme Court’s decision in *International News Service v. Associated Press*, 248 U.S. 215 (1918), recognizing a non-copyright tort of “hot news” misappropriation and leading to numerous cases in which originators of factual compilations (sports scores, stock quotes, etc.) are afforded an initial “exclusivity” period in which to enjoy the fruits of their labor before otherwise unprotectable information enters the public domain. *See, e.g.*, Richard A. Posner, *Misappropriation: A Dirge*, 40 HOUSTON L. REV. 621, 629–34 (2003).

facilitating compromise and commons formation. The case studies described in this Article—the genome commons and scientific publishing—exemplify the use of latency variables to achieve resolution of otherwise difficult value-based positions. Through the use of timing mechanisms, competing stakeholder interests in publication priority versus scientific advancement (in the case of the genome commons), and broad information accessibility versus protection of financial interests (in the case of scientific publishing) have been successfully mediated. I propose that latency analysis and design techniques should be adopted more broadly and explicitly to other information commons, particularly in the sciences, and that the general application and recognition of this methodology can improve both the efficiency of commons formation and the ultimate utility of the resulting commons for all stakeholders.

OPERATING EFFICIENTLY POST-*BILSKI* BY ORDERING PATENT DOCTRINE DECISION-MAKING

Dennis Crouch[†] & Robert P. Merges^{††}

TABLE OF CONTENTS

I.	INTRODUCTION.....	1673
II.	PATENTABLE SUBJECT MATTER: <i>BILSKI</i> AND BEYOND.....	1676
III.	LEXICAL PRIORITY AND PATENTABILITY DOCTRINES	1679
	A. PATENT VALIDITY AS A CHAIN OR CIRCUIT	1680
	B. A PRAGMATIC RATIONALE FOR AVOIDING § 101 DECISIONS.....	1681
IV.	AGAINST A SIMPLE ORDERING: THE FUNDAMENTALLY OVERLAPPING NATURE OF CLAIMS AND DOCTRINES	1683
	A. APPLYING VALIDITY TESTS TO CLAIM SETS, NOT “INVENTIONS”	1684
	B. PRACTICAL AND CONCEPTUAL OVERLAP IN VALIDITY DOCTRINES.....	1686
	C. STOPPING RULES: WHEN TO TERMINATE VALIDITY TESTING.....	1689
V.	CONCLUSION.....	1691

I. INTRODUCTION

Now that the Supreme Court has decided *Bilski v. Kappos*,¹ there is an enormous amount of speculation about the case’s impact on patent applicants, litigants, and other participants in the patent system. Most of the

© 2011 Dennis Crouch & Robert P. Merges.

[†] Associate Professor of Law, University of Missouri School of Law.

^{††} Wilson Sonsini Goodrich & Rosati Professor of Law and Technology, Co-Director, Berkeley Center for Law and Technology. The authors wish to thank the staff of the Berkeley Technology Law Journal for their helpful editing.

1. 130 S. Ct. 3218 (2010).

commentary is concerned with the holding in *Bilski*, how this holding will be applied by courts and the Patent Office, and ultimately, the effect of the holding on inventors, and those who hold and seek patents.

We take a different approach. Rather than try to cut through the complexity of *Bilski*, or predict how it will be applied, we talk about how to avoid it. We are interested in minimizing the cost and confusion that accompany a review of patents for § 101 subject-matter eligibility.² We propose that the § 101 issue of *Bilski* be considered only when doing so is absolutely necessary to determine the validity of a claim or claims in a patent. We believe any claim that can be invalidated under one of the less controversial and less complex requirements for patentability—§§ 102, 103, and 112, for instance—ought to be disposed of without considering subject matter patentability.³ In other words, the *Bilski* issue should be avoided wherever it is not strictly necessary. To support this conclusion, we present a set of empirical data that indicates that the vast majority of patent claims challenged on subject matter eligibility grounds were also challenged on other patentability grounds.

We set the stage for our proposal in Part II, which briefly reviews the history behind *Bilski* and explains its open-ended holding and individualized approach. The difficulty of applying the *Bilski* ruling to different types of patent claims leads us to Part III, in which we call into question an accepted (if largely implicit) principle of patent law—that the lexical priority of statutory provisions in the 1952 Patent Act dictates a necessary logical sequence of invalidity tests. We reject this widespread assumption. There is nothing in the statute that requires this. Indeed, in Part III we argue that in many ways the very idea of a sequence of discrete patentability requirements is conceptually misleading. Claims can be and often are rejected by the Patent Office for multiple reasons, suggesting that at least certain claims suffer from defects that transcend specific statutory validity requirements. We argue further that the policy underpinnings of various requirements overlap in complex ways, so that in reality patentability doctrine does not test for a series of discrete and independent qualities that are distinct from and mutually exclusive of each other. In the same way, transcendent qualities of an invention can influence multiple doctrines simultaneously, with pioneering

2. 35 U.S.C. § 101 (2006).

3. These disqualifying doctrines include, inter alia, anticipation and pre-filing disclosures under 35 U.S.C. § 102 (2006); obviousness under 35 U.S.C. § 103(a) (2006); enablement, written description, and best mode under 35 U.S.C. § 112 ¶ 1 (2006); and indefiniteness under 35 U.S.C. § 112 ¶ 2 (2006).

inventions (due to both a liberal treatment under enablement, and a broad reach under infringement doctrines) being a prime example. This demonstrates again that there is not and should not be a strong separation between various patent law doctrines.

Another argument along these lines recognizes that while patentability doctrines are not discrete entities, neither is the validity of the “the invention” that is being considered. Patent applicants routinely present multiple, overlapping claims, all of which cover fine-grained variations on a central inventive insight or advance. So it is inaccurate to visualize patentability as a stepwise series of tests applied to a single “invention.” It is not true, for example, that “invention X” passes § 101 and should thus proceed in logical sequence to be tested under § 102. One claim growing out of inventive insight X might present no § 101 problems at all, yet another claim in the same patent application might raise a difficult issue under this provision. Each claim, being a unique slice of the overall inventive insight, ought to be considered on its own terms, and in whatever order makes the most sense. Put another way, the mental model of a stepwise sequence of patentability determinations overlooks the highly granular nature in which different slices of the inventive concept are presented for validity testing.

This analysis is further developed in Section III.A. When a claim fails to pass muster under any single test of validity, that claim should be invalidated. *No further tests should be applied.* We describe this as “chain” theory of validity: once one link in the chain is broken, the claim fails, and there is no reason to proceed further. Beyond that point, any expenditure of resources on validity questions is inefficient. Pragmatic considerations enter at this point. Issues of cost, justiciability, and spillover effects are perfectly appropriate in determining the actual sequence in which validity tests are applied with respect to any particular patent claim. The non-linearity of patent validity tests, together with the principle of efficient administration, yields a simple rule: start with chain links that are, in general, easiest and cheapest to test, and when the chain fails, stop the process. That way, the costliest and most complex doctrines—the trickiest “links in the chain”—are often avoided, and in any event are put off until later. Therefore, § 101 should be avoided, both at the Patent Office and in the courts. We justify this not only on efficiency grounds, but also by analogy to the Supreme Court rules of avoidance.

In Part IV, we apply this simple principle. It leads to several recommendations. First, though the PTO has good reasons for its longstanding practice of rejecting claims for multiple reasons, we recommend that § 101 be used only as an exception or last resort even at the PTO. Next, we contend that the courts should proceed in a stepwise fashion, beginning

with §§ 102, 103, and 112, changing the order of doctrines as dictated by pragmatic considerations, and stopping as soon as a claim is conclusively invalidated. In all cases, the difficult process of deciding whether a claim presents patentable subject matter under § 101 should be deferred until very late in the process. Therefore, we recommend, courts should in effect hold off on the difficult task of evaluating claims under § 101—ideally deploying the full § 101 analysis only when that is essential, i.e., when a claim passes muster under the other validity doctrines.

II. PATENTABLE SUBJECT MATTER: *BILSKI* AND BEYOND

The issue of patentable subject matter was given fresh prominence when the Supreme Court handed down the *Bilski v. Kappos* decision in June, 2010.⁴ In *Bilski*, the Supreme Court sustained the invalidity of claims to a financial hedging method that allows commodity users and producers to fix their costs by shifting the risk of supply and demand fluctuations onto other market participants.⁵ The rationale behind the ruling was that Bilski's "claims are not patentable processes because they are attempts to patent abstract ideas."⁶ The Court rooted the exclusion of "abstract ideas" in a long line of precedent stretching back to the nineteenth century, which, the opinion said, confers legitimacy despite the absence of any definition or even mention of the term "abstract" in the text of the Patent Act.⁷ The primary source of the Court's holding was three earlier opinions, all touching on the patentability of software-related claims. In explaining its ruling, the Court explicitly rejected several comprehensive standards for patentability that had been proposed by academics and practitioners. These more comprehensive standards were attempts to create order in the complex area of patentable subject matter.

4. *Bilski*, 130 S. Ct. 3218.

5. Dependent claims of the patent identified energy as the commodity being hedged, identified the other market participants as energy distributors, and identified a statistical method for calculating the fixed costs.

6. *Bilski*, 130 S. Ct. at 3229–30 (citing *Le Roy v. Tatham*, 55 U.S. 156, 175 (1853)).

7. According to the *Bilski* Court:

The Court's precedents provide three specific exceptions to § 101's broad patent-eligibility principles: "laws of nature, physical phenomena, and abstract ideas." [*Diamond v. Chakrabarty*, 447 U.S. 303, 309 (1980).] While these exceptions are not required by the statutory text, they are consistent with the notion that a patentable process must be "new and useful." And, in any case, these exceptions have defined the reach of the statute as a matter of statutory *stare decisis* going back 150 years.

Id. at 3221.

Instead of relying on these, however, the Court chose to rest its holding primarily on its case law from the decade 1970–1980.

Rather than adopting categorical rules that might have wide-ranging and unforeseen impacts, the Court resolved this case narrowly on the basis of its decisions in *Gottschalk v. Benson*, *Parker v. Flook*, and *Diamond v. Diehr*, which showed, according to the Court, that Bilski's claims are not patentable processes because they are attempts to patent abstract ideas.⁸

While it is understandable that the Court would choose to reject “atextual” tests, the organizing impulse behind these proposed standards was that they would transcend the Court's software cases, which were widely thought to be too fact-specific and outdated to serve as an effective source for a workable test.

One problem with the Court's approach is the lack of guidance in the words of the statute. The literal terms of § 101—“process, machine, manufacture and composition of matter”—are so open-ended that they provide very little traction for a court that feels the need to reign in the scope of patentable subject matter. This is where the historical non-textual exclusions from patentable subject matter come in. Long ago a sort of gloss was placed on the Patent Act: despite the statute's open-ended terms,⁹ the Court decreed that “laws of nature, physical phenomena, and abstract ideas” were implicitly excluded from the realm of what is patentable.¹⁰ Because none of these exclusions have ever been defined in legislation, their contours have taken shape over the years through the traditional way of common law principles. The chief advantage of this approach is well understood: flexibility and adaptability.¹¹ By the same token, the lack of a comprehensive definition can also create uncertainty—especially where the volume of case law is relatively low. When cases are few and far between, those who must rely on a

8. *Bilski*, 130 S. Ct. at 3229–30 (citing *Diamond v. Diehr*, 450 U.S. 175 (1981); *Parker v. Flook*, 437 U.S. 584 (1978); *Gottschalk v. Benson*, 409 U.S. 63 (1972)).

9. The statute has undergone both major and minor revisions since its first enactment in 1793, but the provision on patentable subject matter—§ 101 under the current, 1952, Act—has hardly changed at all. See ROBERT P. MERGES AND JOHN F. DUFFY, *PATENT LAW AND POLICY: CASES AND MATERIALS* 68 (4th ed. 2007) (citing Patent Act of 1793, ch. 11, § 1, 1 Stat. 318 (current version at 35 U.S.C. § 101 (2006)) (authorizing patents for “any new and useful art, machine, manufacture or composition of matter, or any new and useful improvement on any art, machine, manufacture or composition of matter”).

10. *Chakrabarty*, 447 U.S. at 309 (citing *Flook*, 437 U.S. 584; *Benson*, 409 U.S. at 67; *Funk Bros. Seed Co. v. Kalo Inoculant Co.*, 333 U.S. 127, 130 (1948); *O'Reilly v. Morse*, 56 U.S. 62, 112–21 (1854); *Le Roy*, 55 U.S. at 175 (1853)).

11. Prior to *Bilski*, the last subject matter eligibility question decided by the Supreme Court was *Diamond v. Diehr*, 450 U.S. 175 (1981).

common law rule are placed in the difficult position of reading the specific facts of the few decided cases for whatever clues and signals they might throw off.

This is precisely the situation that now confronts patent lawyers, inventors, and everyone else who is interested in the scope of patentable subject matter under § 101. The hard kernel of legal authority after *Bilski* is this: the risk-hedging claims in *Bilski* were too abstract. Therefore, claims that are similar enough to those at issue in *Bilski* will also be unpatentable. It is easy enough to state this general principle, but—because of the great variety and complexity of patent claims that will be subject to the post-*Bilski* standard—very difficult in practice to apply it to a specific case. The virtue of the *Bilski* opinion is that it does not tie the patent system down to a restrictive test for patentability. But there is a matching vice: a bedeviling lack of guidance over what patent applicants and patentees can expect when § 101 is applied to a specific patent claim.

We have a particular concern that application of this flexible “no rules” standard to patentable subject inquiry will be difficult at the bureaucratic level of a patent examiner. Patent examiners are typically non-lawyer technology experts trained to judge technical questions of newness and sufficiency of disclosure. Examiners are likely not similarly prepared to pursue the more philosophical inquiries associated with patentable subject matter.¹²

If the volume of Supreme Court case law was higher, this guidance might emerge relatively quickly. But typically, the Court takes few cases in this area. Consider also that in the *Bilski* case itself, the Court was deeply split (5–4) along traditional conservative-liberal lines. Because of this, the Court might well choose to avoid the issue for some period of time. This probably leaves the Federal Circuit to sort through this area on its own. In the next Part, we argue that the Federal Circuit, as well as the district courts and the PTO, ought to follow a prudential rule of simply avoiding the § 101 issue whenever possible.

12. See Reply Brief for Petitioners at 3, *Bilski v. Kappos*, 130 S. Ct. 3218 (2010) (No. 08-964), 2009 WL 3453657, at *3 (“Rather than struggling to determine whether a machine is ‘particular’ enough or whether a claim falls within the ever-changing definition of ‘technology,’ the question of patentability should instead focus on the underlying substance of an invention and whether it is novel, nonobvious, particularly described, and properly claimed.”); Brief for Amicus Curiae, Roberta J. Morris, Esq., Ph.D., In Support of Appellants and thus Supporting Reversal at 5–6, *In re Bilski*, 545 F.3d 943 (Fed. Cir. 2008) (No. 07-1130), 2008 WL 1842256, at *6–7 (explaining that patent examiners are better qualified to judge technical questions of novelty and obviousness than the “philosophical” questions of patentable subject matter).

III. LEXICAL PRIORITY AND PATENTABILITY DOCTRINES

The lexical priority of § 101 is often assumed to dictate the order in which validity issues must be addressed. So it is often assumed that when a § 101 issue arises in a patent case, the relevant decision maker should deal with that issue first, before other requirements for patentability are even considered. Some dicta in Supreme Court cases support this idea. In *Diamond v. Diehr*, the Court focused on whether the claim at issue was “barred at the threshold by § 101.”¹³ And in *Parker v. Flook*, the Court said “[t]he obligation to determine what type of discovery is sought to be patented must precede the determination of whether that discovery is, in fact, new or obvious.”¹⁴ So too at the Federal Circuit, which said in *State Street Bank & Trust Co. v. Signature Financial Group, Inc.* that “[t]he first door which must be opened on the difficult path to patentability is § 101.”¹⁵

As might be expected, this approach has filtered down to the Board of Appeals at the PTO, which has found that lexical priority dictates a rigid order for evaluating validity. In *Ex parte Christian*,¹⁶ for example, the examiner rejected Christian’s claims as anticipated by a prior publication. On appeal, the Board of Appeals refused to evaluate the merits of the prior art rejection, and instead instituted a new ground of rejection solely focusing on subject matter eligibility grounds under § 101.

Surely the idea that the patent validity provisions are meant to be applied in the order that they appear in the Patent Act has some appeal. If we take the linguistic details of our statutes seriously, why then not take the lexical ordering of the statutory provisions seriously as well? Section 101 appears before §§ 102 and 103, after all—and surely there is some good reason for this. Who would advocate skipping willy-nilly around the statute when a complex patent case is being decided? Surely the result would be something like chaos.

To begin, it is important to see that the various sections of the Patent Act certainly do proceed in a logical order. It is certainly not *illogical* to start the substantive requirements of patent law with patentable subject matter (after

13. 450 U.S. at 188.

14. 437 U.S. at 593; *see also In re Bergy*, 563 F.2d 1031 (C.C.P.A. 1977), *vacated sub nom. Parker v. Bergy*, 438 U.S. 902 (1978), *remanded to In re Bergy*, 596 F.2d 952, 960 (C.C.P.A. 1979) (“Achieving the ultimate goal of a patent under th[e] statutory provisions involves, to use an analogy, having the separate keys to open *in succession* the three doors of sections 101, 102, and 103” (emphasis added)).

15. 149 F.3d 1368, 1372 n. 2 (Fed. Cir. 1998) (quoting *Bergy*, 596 F.2d at 960); *see also In re Comiskey*, 554 F.3d 967, 973 (Fed. Cir. 2009).

16. *Ex Parte Christian*, No. 2009-6589, 2010 WL 3389297 (B.P.A.I. Aug. 23, 2010).

the administrative details of the very first sections of the Patent Act). In a way, it makes sense to begin with a general definition of the types of things that are patentable, and then proceed to the question whether a particular invention is novel; then nonobvious; then enabled; and so on. But we argue that although the statute unfolds in a logical order, this is not the only order that makes sense. And, most importantly, it is not essential to *apply* the statutory requirements in the precise order they are set out in the Patent Act.

It is black letter law, for example, that novelty under § 102 is determined strictly on the basis of all the elements recited in a given claim. But if one or more of those elements cannot be pinned down, it makes sense to start with a consideration of the claim itself. Thus, for example, the definiteness provision of § 112 ¶ 2 might logically be considered first in some cases.¹⁷ Furthermore, sometimes other claim-related requirements must also be considered first, before non-claim-related validity doctrines can be applied. For example, a claim whose only limitation is a single means plus function element—invalid under § 112 ¶ 1—need not first pass through the analysis required under §§ 102 and 103.¹⁸

It could be argued from all this that a series of provisions setting forth what is required for a valid claim might logically appear first in the Patent Act. That would certainly make sense, given the primacy of claims in contemporary patent analysis. But even under a patent statute so ordered, it would not make sense to always start with claim-related issues. Consider a case involving a claim with no apparent defects that read directly on a comprehensive piece of prior art. In this scenario, it would be logical to move directly to § 102. The point is therefore not that some optimal ordering of patent validity doctrines is available; it is that no such ordering will make sense in every case. Which leads us back the simple starting point: though § 101 comes first in the Patent Act, it need not always be considered first in a particular case.

A. PATENT VALIDITY AS A CHAIN OR CIRCUIT

We illustrate our point here with a simple analogy. To do its job, each link in a chain must be sound. If any one link breaks, the chain will not work. When thinking about its overall performance, it is the overall effectiveness of the chain that counts; there is no reason to focus on any particular link, or

17. *Cf.* *Datamize, LLC v. Plumtree Software, Inc.*, 417 F.3d 1342 (Fed. Cir. 2005) (deciding case on the basis of indefiniteness, without reaching other issues of patent validity).

18. *See In re Hyatt*, 708 F.2d 712 (Fed. Cir. 1983) (rejecting “single means” claim under § 112 ¶ 1, enablement requirement, without reaching other validity issues).

any particular order. Put another way, each link can potentially be the subject of the first test. As soon as a single link fails, the chain is broken, so other links need not be tested or considered.

Another useful analogy is to a series circuit. If two or more electrical components are connected in a series circuit, a malfunction in any component (such as a light bulb that burns out) disrupts the circuit, and none of the components will work.¹⁹ In testing a series circuit that has failed, the obvious procedure to follow is to start with the weakest component, whichever one that gives the greatest evidence of having a problem. It would make far less sense to follow a rigid testing process determined by some arbitrary ordering, such as testing the closest component to the power source first, the second-furthest second, and so on. It is precisely this rigid and unthinking ordering that we are opposed to when it comes to applying doctrines of patent validity.

These analogies bolster our initial point: in some cases, such as those where patentability is rejected on other grounds, *Bilski*-type subject matter questions can be entirely avoided. This is a good starting point for our conception of patent validity testing. But, there are a number of more practical reasons to believe that these § 101 issues *should* be avoided when possible.

B. A PRAGMATIC RATIONALE FOR AVOIDING § 101 DECISIONS

If an “easier” issue—one involving less controversy and requiring lower resource expenditures to correctly resolve—would conclusively resolve a case, the courts should decide the case on the basis of that issue, and express no opinion on § 101. This approach makes sense for two related reasons. It is more efficient, obviously. And also, it preserves the scarce currency of court legitimacy. These goals are furthered in the Supreme Court’s doctrine of avoidance, under which the Court resolves cases on non-constitutional grounds whenever possible.²⁰ By analogy to the doctrine of avoidance, our approach conserves the courts’ legitimacy by reserving consideration of § 101 issues for only those cases in which it is absolutely essential.²¹

19. This is not true of a parallel circuit, in which each component is attached to a power source separately.

20. *Ashwander v. Tenn. Valley Auth.*, 297 U.S. 288, 345–48 (1936); HART & WECHSLER’S, *THE FEDERAL COURTS AND THE FEDERAL SYSTEM* 76–77 (Richard H. Fallon, Jr. et al eds., 6th ed. 2009).

21. To be clear, we are not proposing that courts apply an aggressive form of avoidance that would alter the statutory construction of other Patent Act provisions in order

So for example, a questionable claim under § 101, which is clearly invalid because of a statutory bar event or very clear prior art under § 102, should be resolved on the § 102 ground. The same for validity under the utility requirement, under § 103 or § 112. Any validity requirement that has generated an extensive body of case law will in general be a firmer and less controversial basis for invalidating a patent than the philosophical post-*Bilski* inquiry under § 101. We say this in full recognition that the analysis of some validity provisions—in particular § 103, and in some cases the written description requirement under § 112—can themselves be arduous undertakings.

Although clearly not a constitutional question, we see some parallels between our vision for patentable subject matter procedural jurisprudence and the Supreme Court's doctrine of avoidance. As with the Constitution, the text of the § 101 has remained virtually unchanged for over two hundred years.²² During that time, the statute has served as the fundamental core defining our patent system, and, in that role, the Supreme Court has continued to interpret the provision with its own gloss that does not appear to be fundamentally based on the words of the statute.²³ And, perhaps more than any other provision in the Patent Act, § 101 decisions tend to be policy based and politically minded.²⁴

Rules of constitutional avoidance are deeply-seated in U.S. judicial practice. As Professor Adrian Vermeule wrote, “[a]voidance is perhaps the preeminent canon of federal statutory construction.”²⁵ Many rules of avoidance exist—some that aid in construing federal laws in ways that avoid constitutional conflicts and others that suggest a jurisprudential approach that avoids directly addressing constitutional questions unless absolutely

to avoid deciding a § 101 issue. Rather, we propose only that the Court reserve any decision on § 101 issues until other patentability doctrines have been resolved.

22. See Patent Act of 1793, Ch. 11, 1 Stat. 318–23 (Feb. 21, 1793) (patents available for any “new and useful art, machine, manufacture or composition of matter”).

23. *Bilski v. Kappos*, 130 S. Ct. 3218 (2010) (noting that historic subject matter exclusions are “not required by the statutory text”).

24. See Orin S. Kerr, *Rethinking Patent Law in the Administrative State*, 42 WM. & MARY L. REV. 127, 190 (2000) (“PTO discretion to set the scope of patentable subject matter would transform control over the PTO into a valuable political chip.”).

25. Adrian Vermeule, *Saving Constructions*, 15 GEO. L.J. 1945, 1948 (1997); see *Ashwander*, 297 U.S. at 347 (Brandeis, J., concurring) (“The Court will not pass upon a constitutional question although properly presented by the record, if there is also present some other ground upon which the case may be disposed of.”); see also Alexander M. Bickel, *The Supreme Court 1960 Term Foreword: The Passive Virtues*, 75 HARV. L. REV. 40, 58 (1961); Lisa A. Kloppenberg, *Avoiding Constitutional Questions*, 35 B.C. L. REV. 1003, 1025 (1994); Cass R. Sunstein, *Constitutionalism After the New Deal*, 101 HARV. L. REV. 421, 468–69 (1987).

necessary.²⁶ There are also many different but related grounds for the principle of avoidance,²⁷ but one core idea is that the principle helps courts conserve their legitimacy, the scarce but essential resource that justifies judicial review.²⁸

We do not argue that issues under § 101 of the Patent Act rise to the level of constitutional questions. Our invocation of the avoidance rule is therefore by analogy only, which we find convincing, for the same reasons that avoidance makes sense in the Supreme Court context. Validity under § 101 presents issues to the courts that are complex, difficult, and saturated by fundamental policy considerations. Deciding a § 101 case necessarily involves a judgment about whether a particular invention falls into a particular category of inventions—and ultimately about the patentability of that category as a whole. This is an issue that often far transcends the inherently bounded questions of patentability under §§ 102 and 103, or enablement under § 112. The courts in this area are given rather thin material to work with: some fairly ancient general phrases (machine, manufacture, composition of matter), and some quite general categories of exception (products of nature, natural phenomena, abstract ideas). To fashion a holding in a § 101 case out of these materials will often require a court to stretch the available authority in an effort to apply it to specific facts. It is therefore to be expected that rulings in these cases will often engender controversy and strain the credibility of the courts in charge. Therefore, the same logic that leads the Supreme Court to avoid its most delicate subject matter ought to apply in the case of patents and § 101.

IV. AGAINST A SIMPLE ORDERING: THE FUNDAMENTALLY OVERLAPPING NATURE OF CLAIMS AND DOCTRINES

In this Part, we explain in more detail why a simple lexical ordering approach fits poorly with basic features of the patent system. The problem with rigid ordering is that it is based on a very simplistic model of patent claims and validity doctrines. In particular, a rigid approach assumes a single discrete invention that is processed through a linear series of discrete and independent validity doctrines. The truth of the matter is that this process is

26. See *Ashwander*, 297 U.S. at 346–48.

27. See, e.g., *Rescue Army v. Mun. Ct. of L.A.*, 331 U.S. 549, 571 (1947) (reciting a non-exhaustive list of grounds supporting the avoidance doctrine).

28. *Id.* (citing “the delicacy of [the judicial review] function” and “the limited resources of enforcement” available to courts as rationales for avoidance).

much more complex. Inventions are “granularized” and “unpacked” by patent applicants, who construct a series of claims that cover different aspects and dimensions of the inventive concept. So a single patent may present individual claims that trigger close calls under several different validity doctrines. This means that no single, linear procedure will work in a foolproof way, even on a single patent. In addition, when it comes to validity doctrines, there is a good deal of conceptual overlap between the discrete requirements of patent validity. Thus, at the conceptual level, even with respect to a single patent claim, there is no foolproof ordering that moves along a logical sequence in which each concept is discrete and separate from the others.

A. APPLYING VALIDITY TESTS TO CLAIM SETS, NOT “INVENTIONS”

We tend to think in terms of testing “inventions” for patentability under various validity doctrines, but this is not really an accurate portrayal of the law. In a strict sense, inventions are neither valid nor invalid; only claims are. Patent applicants know this all too well, which is why they almost invariably include multiple claims when seeking legal protection for their inventions. For purposes of validity, then, it makes more sense to speak in terms of claim sets rather than inventions.

For example, suppose an inventor comes up with a new way to access stored data files that creates a unique, encrypted identifier based on the contents of a file. This allows users to search for and retrieve files without knowing their precise physical locations or specific file names.²⁹ Conventional claiming strategy for an invention like this would start with one or more broad, independent claims and then refine each independent claim, through various narrowing limitations recited in a series of dependent claims. The independent claims might include a “computer system” or simply “system,” meant to be an open-ended claim to a broad class of computer hardware implementations of the data storage idea. Another independent claim that makes sense in this situation is one that covers a method for storing data. Additional independent claims might cover slightly less open-ended versions of these basic claims, such as an implementation of the storage technique in a local or distributed network (a slightly narrower

29. This example is loosely based on the development of content-addressable memory in computer systems. See *Content Addressable Storage*, WIKIPEDIA, http://en.wikipedia.org/wiki/Content-addressable_storage (last visited Oct. 31, 2010); U.S. Patent No. 6,807,632 (filed Jan. 21, 1999) (issued Oct. 19, 2004); U.S. Patent No. 7,475,432 (filed July 19, 2004) (issued Jan. 6, 2009).

embodiment than the broadest “system” claim, which would include embodiments running on a single computer). Some method claims in the claim set might include a general algorithmic approach (i.e., first assign a unique identifier to a file, then put this identifier into a table, then note the location of the table, etc.), while others might well be more specific about the algorithm. One can also imagine a claim to the method of using this storage technique for retrieval of title documents in a computerized real estate database. Finally, the method claims might culminate in a recitation of a specific series of steps or commands written in a certain programming language.

The claims in this hypothetical patent application raise many patent validity issues. Depending on how broadly the computer system and system claims are drafted, there could well be prior art that raises serious validity issues under § 102 or § 103. A claim to a system that includes “retrieval of encrypted file identifying data,” without further limitations, might be anticipated by an earlier computer system designed for high-security applications. Clearly it would make sense to first examine such a claim under § 102 or § 103. By the same token, other claims might well raise other issues. For example, a claim that includes a broadly-worded software element—such as a “data access module,” or “means for retrieving data”—might raise indefiniteness issues. Validity testing of a claim like this might sensibly start with a consideration of § 112 ¶ 2.³⁰

Now consider the claim mentioned earlier, covering a method of using the storage retrieval technique for title documents in the real estate industry. This claim might implicate § 101 patentable subject matter. Under *Bilski*, the relevant question would be whether the claim was too abstract. This would be a complex and difficult question, given the open-ended nature of the “abstractness” inquiry and the difficulty of fitting a case such as this into the framework of earlier Supreme Court cases.

If there is solid prior art in the area of real estate title storage and retrieval, the same claim might also raise validity issues under § 102 or § 103. The essence of our proposal in this Article is that in such a case novelty and

30. See *Blackboard, Inc. v. Desire2Learn, Inc.*, 574 F.3d 1371, 1382–83 (Fed. Cir. 2009) (invalidating a claim including the element “means for assigning a level of access and control,” where specification failed to disclose any structure for this element); *Aristocrat Tech. Austl. PTY Ltd. v. Int’l Game Tech.*, 521 F.3d 1328 (Fed. Cir. 2008) (invalidating a claim including “control means” element where specification failed to disclose any structure corresponding to the element); *Datamize, LLC v. Plumtree Software, Inc.*, 417 F.3d 1342 (Fed. Cir. 2005) (examining indefiniteness with regard to software disclosure).

nonobviousness should be considered first, before addressing the § 101 requirement. If prior art is identified that invalidates the claim, there is no need to move on to the difficult § 101 issue. As we have been arguing, there is no necessary order in which validity issues must be considered in every instance. Each claim must be taken on its own. In some cases, it makes sense to consider one doctrine first; in others, it makes sense to start somewhere else.

With the hypothetical claim we have just described, the place to start is with §§ 102 and 103. Just as judgment and prudence dictate that with other claims in our example it might make sense to start with § 112 or § 102, we think the title retrieval system claim ought to be examined first under provisions other than § 101. Only if the claim passes muster under these other provisions will it be necessary to take up the question of patentable subject matter.

B. PRACTICAL AND CONCEPTUAL OVERLAP IN VALIDITY DOCTRINES

Our argument thus far has been based on an assumption of doctrinal overlap. Namely, that a substantial number of patent claims lacking subject matter eligibility under of § 101 also fail to satisfy at least one other validity test. Although we have not completed an exhaustive empirical study, this assumption appears reasonable and is supported by two recent studies. In the first study, Professors Mark Lemley, Christopher Cotropia, and Bhaven Sampat examined the prosecution history files of over 1,500 recently issued US patents. When parsing through their sample, the authors found that 84% of the patent applications that had been rejected for lacking subject matter eligibility were also rejected as either anticipated or obvious.³¹ In a separate study conducted for this article, we reviewed a set of 117 recently released opinions of the Board of Patent Appeals and Interferences (BPAI) that decided an issue of subject matter eligibility. In 110 (94%) of the BPAI opinions in our sample, each claim questioned on subject matter eligibility grounds also stood rejected on at least one other ground.³² This data appears

31. Christopher A. Cotropia, Mark Lemley & Bhaven Sampat, *Do Applicant Patent Citations Matter? Implications for the Presumption of Validity* (Stanford Law & Econ., Working Paper No. 401, 2010), available at <http://ssrn.com/abstract=1656568> (showing unpublished data generated by study that was conducted at the patent level, not claim-by-claim).

32. For this study, we used Westlaw to broadly search for all ex parte BPAI cases decided between October 13, 2009 and October 13, 2010 that included the phrase "patentable subject-matter." Those opinions were then examined to determine whether the claims were rejected on subject matter eligibility grounds and, if so, whether the claims were also rejected on any other ground. The table of collected data is available online at <http://www.patentlyo.com/CrouchMerges.BPAIDecisions.2010.xlsx>.

to show an exceptionally high rate of doctrinal overlap and lends credence to the idea that, by initially avoiding subject-matter-eligibility questions, many of those potential issues will be avoided.

The *Bilski* case likely represents a scenario that could have been decided on other grounds—especially to the extent that the claims are interpreted broadly as an attempt to “patent . . . the concept of hedging risk and the application of that concept to energy markets.”³³ The general concept of hedging is old and *Bilski*’s particular application of that concept was likely obvious at the time of his invention.³⁴

Although patent examiners are instructed to assert all applicable reasons for rejection in each office action rejection,³⁵ the BPAI has been somewhat aggressively following a rule of lexical priority. The tribunal’s recent decision in *Ex parte Christian*³⁶ is typical of this approach. The patent examiner rejected *Christian*’s claims as anticipated by a prior publication. On appeal, the BPAI refused to evaluate the merits of the prior art rejection and instead instituted a new ground of rejection solely focusing on subject matter eligibility grounds. This procedure was sanctioned by *In re Comiskey*.³⁷ In that case, the Court of Appeals for the Federal Circuit reiterated the “first door” approach and § 101 “threshold” language of *Bergy*,³⁸ and *Diehr*,³⁹ as it focused solely on newly raised patentable subject matter issues and held the other patentability issues moot. Thus, despite the practical overlap between the various patent

33. *Bilski v. Kappos*, 130 S. Ct. 3218, 3222 (2010).

34. See Brief of Amicus Curiae Regulatory Datacorp, Inc. in Support of Neither Party, *In re Bilski*, 545 F.3d 943 (Fed. Cir. 2008) (No. 2007-1130), 2008 WL 1842273, at *27–28 (citing Ronald White, *Some Statistical Aspects of Future Trading on a Commodity Exchange*, 99 J. ROYAL STAT. SOC’Y 297, 315 (1936)); Dennis Crouch, *Bilski: Full CAFC to Reexamine the Scope of Subject Matter Patentability*, PATENTLY-O (Feb 15, 2008), <http://www.patentlyo.com/patent/2008/02/bilski-full-cafc.html> (noting “serious obviousness problems” with the *Bilski* claims); Lauren Katzenellenbogen, Bob Irvine & David Donoghue, *Debate on In re Bilski*, 7 NW. J. TECH. & INTEL. PROP. 260 (2009) (“[*Bilski* is] a bad case [for Section 101 analysis] because the *Bilski* invention would have been rejected for obviousness, if nothing else, so really the rejecting it because of subject matter was sort of an additional reason to reject it.”).

35. See MPEP § 706 (8th ed. Rev. 7, Sept. 2008) (“The goal of examination is to clearly articulate any rejection early in the prosecution process so that the applicant has the opportunity to provide evidence of patentability and otherwise reply completely at the earliest opportunity.”).

36. *Ex Parte Christian*, No. 2009-6589, 2010 WL 3389297 (B.P.A.I. Aug. 23, 2010).

37. *In re Comiskey*, 554 F.3d 967, 973 (Fed. Cir. 2009).

38. *In re Bergy*, 596 F.2d 952, 960 (C.C.P.A. 1977) (“The first door which must be opened on the difficult path to patentability is § 101.”).

39. *Diamond v. Diehr*, 450 U.S. 175, 188 (1981).

doctrines, the adjudicative bodies tend to focus on § 101 issues to the exclusion of the other patentability doctrines.

The doctrinal overlap is not surprising given that subject matter eligibility overlaps with many of the other patentability doctrines in both purpose and operation. All of the patentability doctrines seek to ensure that granted patents are not overreaching but instead are given their appropriate scope.⁴⁰ It makes sense that claims directed toward naturally occurring phenomena—unpatentable under § 101—will likely fail the newness requirements of §§ 102 and 103(a). Likewise, the breadth of a claim directed to an abstract idea increases the likelihood that (1) an embodiment covered by the claim is already known in the art and (2) that the disclosure failed to provide an enabling written description commiserate with the scope of the claims. Even today, academics argue over whether the 1853 decision in *O'Reilly v. Morse*,⁴¹ rejecting Morse's broadest telegraph claim, should be categorized as a subject matter eligibility decision, or instead as an enablement decision.⁴² Thus, although the various patentability doctrines are each distinct in some form, they still overlap in many, often complex, ways.⁴³

The complex overlapping nature of the patentability doctrines is mirrored in most patentees' claims of inventive rights. Namely, patent applicants typically protect an invention with multiple, overlapping claims, and the validity of each relevant claim must be considered before a patent issues or infringement lawsuit concludes. In the same way that the patent doctrines can all be explained by a handful of policy goals, the set of claims defines the invention around one or more central inventive advances.

In litigation, courts are tasked with judging patentability as a binary valid/invalid inquiry. However, practicalities of patent prosecution that allow for both multiple claims and multiple pre-issuance amendments potentially serve to granularize the otherwise binary question. In the circuit analogy

40. See Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839, 843 (1990); see also Mark F. Grady & Jay I. Alexander, *Patent Law and Rent Dissipation*, 78 VA. L. REV. 305 (1992). But see *SmithKline Beecham Corp. v. Apotex Corp.*, 403 F.3d 1331, 1342 (Fed. Cir. 2005) (“The scope of the claims is not relevant to subject matter eligibility.”).

41. *O'Reilly v. Morse*, 56 U.S. 62 (1853).

42. See, e.g., Dan L. Burk & Mark A. Lemley, *Inherency*, 47 WM. & MARY L. REV. 371, 404 n.161 (2005) (“[T]he decision in *O'Reilly* may more properly be read to hold that Morse failed to enable . . . his broadest claims”).

43. The parallel purposes of the patentability doctrines suggests to us that, during patent prosecution claim amendments necessitated by a rejection under a non-subject matter eligibility patentability doctrine will often incidentally correct subject matter eligibility problems.

described above, multiple claim sets add some parallel organization to the circuit previously considered linked in series. From the standpoint of doctrinal overlap, the addition of multiple claims of varying scope suggests that an increasing variety of patent doctrines would be required to test the validity of a patent. In this common scenario, one set of claims might be most quickly invalidated based on published prior art associated with the invention, while other claims may lack a proper written description under § 112 or subject matter eligibility under § 101.

Our take-away from this discussion is the clear notion that patentability doctrines frequently overlap. In turn, this overlap suggests that some amount of conscious ordering of analysis could serve to reduce the workload of decision makers and to avoid having to decide disfavored doctrines.

C. STOPPING RULES: WHEN TO TERMINATE VALIDITY TESTING

It is one thing to talk about the order in which validity testing is performed. It is another thing altogether to ask when validity testing should stop. Should all potentially relevant validity issues be determined even if an apparently fatal defect is encountered at an early stage of testing? In general, the practice in court has been to terminate proceedings when the first fatal defect is encountered.⁴⁴ Once a patent is held invalid, the other potentially useful doctrines become moot and the court may lose its jurisdictional findings.⁴⁵ We support this practice.

In contrast, in early-stage patent examination, PTO examiners often reject claims on multiple grounds. The difference in approach may well be justified by understanding some subtle differences between a patent being challenged in court and a patent application being examined at the PTO. A typical patent application is involved in multiple rounds of examination before a patent eventually issues or the application is left abandoned. During that period, applications are regularly amended in order to overcome examiner rejections or to take into account other information discovered during the examination period. However, because almost every applicant has a wide variety of potential claim amendments they can implement, the process does not necessarily follow any linear or predictable pattern. Finally, patent examiners may have less confidence in their rejections—either

44. *See, e.g.*, *Geo M. Martin Co. v. Alliance Machine Systems Intern. LLC*, 618 F.3d 1294, 1300 (Fed. Cir. 2010) (discussing district court refusing to decide moot issues); *Blackboard, Inc. v. Desire2Learn, Inc.*, 574 F.3d 1371 (Fed. Cir. 2009) (refusing to decide obviousness question after ruling that the claims were invalid as indefinite).

45. Some courts do prefer to provide alternative grounds for judgment as a way to add credibility to the decision and to bolster the decision's potential to sustain an appeal.

because of their own lack of experience or because they intentionally have made questionable rejections in order to force the patent applicant to prove his case. Based on all of these factors, the PTO has chosen to take the approach of attempting to address all potential patentability issues in the initial examination decision. The PTO's apparent hope is that this full up-front analysis will reduce the rounds of negotiation because patent applicants better understand the types of amendments necessary to obtain a patent and spurious examiner rejections can be dealt-with as a unit.⁴⁶

While we see merit in the PTO's approach, we suggest here that subject matter eligibility questions be treated as an exception and delayed until after the application passes muster with all other patentability doctrines. Our conclusion derives from our cautious hope that most subject matter eligibility questions will be corrected during the process of overcoming the examiner's obviousness, indefiniteness, and enablement rejections, and therefore, that the PTO will only rarely need to reveal its examiners' relative lack of skill in judging philosophical questions of abstract ideas and products of nature. To be clear, we are not looking to poke fun at patent examiners or BPAI judges. Rather, our point is that the subject matter eligibility test under *Bilski* is quite difficult for anyone to implement (because of the lack of guidance), and on a comparative basis, an examiner's time is better spent applying the other patentability doctrines and at least temporarily ignoring subject matter eligibility questions.

Professor Tun-Jen Chiang has suggested that it is a mistake to consider patentable subject matter doctrine as an indivisible unit. Chiang would rather divide the doctrine into at least two categories, one of which is easy to reliably judge.⁴⁷ To the extent that the low-cost, rule-based subject matter decisions can be easily categorized and segregated from the more philosophical questions highlighted in *Bilski*, it may make sense to only postpone judgment for the more difficult questions.⁴⁸

In this vein, PTO examiners may be well equipped to easily determine whether a particular invention either (1) incorporates a particular machine into the invention or (2) transforms an article from one thing or state to

46. See MPEP § 706 (8th ed. Rev. 7, Sept. 2008).

47. Tun-Jen Chiang, *The Rules and Standards of Patentable Subject-Matter*, 2010 WIS. L. REV. 1353 (distinguishing rule-driving subject matter eligibility questions that are easier to accurately adjudicate from more flexible standards-based issues that are more difficult to accurately decide).

48. We agree with Professor Chiang that some subject matter eligibility questions are easy to resolve. However, we disagree with his notion that it is a simple matter to apriori distinguish the easy cases from the difficult ones.

another. In *Bilski*, the Supreme Court approved this machine-or-transformation test as offering an important clue to patentability, but ruled that the machine-or-transformation test could not completely answer the eligibility question.⁴⁹ Although PTO examiners could use the machine-or-transformation test as a simple rough-cut eligibility test, we see problems with that approach. Namely, initial rejections following the rule will only be roughly accurate and would lead to further arguments in the prosecution process; challenge PTO decision-making legitimacy; and potentially set-up appeals that force the courts to decide subject matter eligibility questions (especially when the PTO decides subject matter eligibility and ignores other patentability questions as it did in *Bilski*).

V. CONCLUSION

Our proposal ultimately adds up to breaking through the superficial appeal of lexical ordering and imposing a more pragmatic approach to the sequence in which decision makers evaluate patent validity doctrines. As simple as our approach is, we are convinced that it would have a number of salutary effects. It would first and foremost cut down the total cost of deciding validity issues, given that § 101 is the most vague and contentious of all the validity doctrines. It would also contribute to greater respect for patent tribunals, by removing them whenever possible from the controversial business of deciding cases under § 101. And finally, by making this a rare and unusual basis for deciding patent cases, it would make the entire validity-determination process more certain and less problematic.

To be sure, *Bilski* will spawn a huge amount of commentary and an equal amount of controversy. We believe that many will look in vain for a clear and consistent set of principles to apply in future § 101 cases; however, the answers will not be found, or at least, convincing answers will not be found in the pages of the *Bilski* opinion. It is unlikely that a single § 101 case can supply what everyone seeks; the nature of the inquiry, as shown by the long history of case law in this area, militates against this sort of firm guidance. Under the circumstances, it is best not to try to map the swampy terrain of § 101 in any great detail. Whenever possible, we argue, try something else: just avoid it.

49. *Bilski v. Kappos*, 130 S. Ct. 3218 (2010).

PATENT LAW'S PARSIMONY PRINCIPLE

Alan Devlin[†]

TABLE OF CONTENTS

I.	INTRODUCTION	1694
II.	THE CASE FOR PARSIMONIOUS PATENT RIGHTS	1701
	A. THE PATENT SYSTEM'S CHARGE: SPURRING INNOVATION	1701
	B. PROMOTING A PARSIMONIOUS APPROACH TO PATENT POLICY	1709
III.	PATENT LAW'S SILENT EMBRACE OF THE PARSIMONY PRINCIPLE	1715
	A. PATENTABLE SUBJECT MATTER	1715
	B. THE NATURE OF A PATENTEE'S PROPERTY RIGHT	1722
	C. MONOPOLY LEVERAGE AND RELATED LIMITS ON PATENT VALUE	1727
	D. DOCTRINE OF EQUIVALENTS	1732
	E. EXPERIMENTAL USE—AN UNJUSTIFIED EXCEPTION?.....	1734
IV.	THE PARSIMONY PRINCIPLE AS A GUIDE TO PATENT POLICY	1736
	A. EMPLOYING THE PARSIMONY PRINCIPLE TO INFORM CONTEMPORARY DEBATE	1737
	1. <i>Exploring the Limits of Patentable Processes—Computer Software and Business Methods</i>	1737
	2. <i>Optimal Damages in the Event of Infringement</i>	1743
	B. AN IMPORTANT CAVEAT: AVOIDING EX POST DILUTION OF PROPERTY RIGHTS.....	1746
V.	CONCLUSION	1747

© 2011 Alan Devlin.

[†] Law Clerk to the Hon. Amy J. St. Eve, U.S. District Court for the Northern District of Illinois; University College Dublin, B.B.L. (Int'l), 2004; University of Chicago, LL.M., 2005; University of Chicago, J.S.D., 2006; Stanford Law School, J.D., 2007. The author would like to express his particular gratitude to Steven Horowitz.

I. INTRODUCTION

The patent system is charged with a deceptively difficult task: maximizing social welfare by spurring innovation through the provision of optimal rewards.¹ To create perfect incentives, patent law would have to tailor inventor-specific property rights so that every innovator would receive minimally sufficient exclusivity to encourage desirable investment in R&D and subsequent commercialization.² Were it able and willing to employ such a system, society would reap the greatest possible welfare by maximizing innovation while minimizing loss to allocative efficiency. This method for promoting innovation, which seeks narrowly to tailor property rights to the incentive characteristics of each inventor, is surely optimal.³

But how could one ever follow this course in practice? The “useful arts,” though subject to continuing interpretive debate, comprise products and processes in such disparate fields of innovation as pharmaceuticals, biotechnology, computer software, medical diagnostic techniques, and business methods.⁴ Like any group of individuals, inventors are characterized by divergent motivational influences. They are an idiosyncratic lot who will respond in an asymmetric fashion to any one set of incentives established by an intellectual-property regime. Those implementing the patent system would need perfect access to information about each inventor’s reservation return—the expected return below which each inventor would decline to innovate. Quite obviously, such information is elusive. Indeed, it is testament to the difficulty of aligning incentives with the traits of individual inventors that patent law has largely spurned any suggestion that it attempt to do so.⁵ For the most part, the patent system operates on a “one-size-fits-all” basis.⁶ Yet, to the extent the current patent system creates a blanket set of rules that

1. See U.S. CONST. art. I, § 8, cl. 8; *Sinclair & Carroll Co. v. Interchemical Corp.*, 325 U.S. 327, 330–31 (1945); *Biotechnology Indus. Org. v. District of Columbia*, 505 F.3d 1343, 1346 (Fed. Cir. 2007).

2. See Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839, 876–79 (1990).

3. See WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW* 297–300 (2003).

4. See ROGER E. SCHECHTER & JOHN R. THOMAS, *PRINCIPLES OF PATENT LAW* 23–60 (2004).

5. Patent law has displayed some flexibility with respect to readily apparent, industry-specific characteristics. Although the current doctrine purports to apply rules with horizontal uniformity, the manner in which rules have actually been applied occasionally differs, sometimes significantly, depending on the industrial context.

6. Natasha N. Aljalian, *The Role of Patent Scope in Biopharmaceutical Patents*, 11 B.U. J. SCI. & TECH. L. 1, 18 (2005); Rebecca S. Eisenberg, *The Role of the FDA in Innovation Policy*, 13 MICH. TELECOMM. & TECH. L. REV. 345, 364 (2007).

apply in comparable fashion across numerous different industries, the current system almost inevitably over-incentivizes innovation in some contexts and under-incentivizes it in others.⁷ Such an outcome is undesirable. As economists typically regard dynamic-efficiency gains as a greater source of long-run social welfare than incremental gains in static efficiency, the prospect of under-rewarding inventors is unsettling.⁸ This is particularly so in capital-intensive, risky fields of innovation, such as the pharmaceutical industry, where desirable levels of R&D depend critically on a sufficient expected return.⁹ In fields characterized by rapid, incremental, “follow-on” innovation, such as computer software, the prospect of excessive rewards may be of greater concern because it may operate to frustrate cumulative innovation.¹⁰ Unfortunately, because we lack sufficient information to adjudge the incentive preferences of different inventors, it is likely unavoidable that some such over- and under-compensation will continue to occur.

There is an apparent solution to this quagmire. Since the law cannot mold property rights to fit the incentive characteristics of every industry and of every inventor lying therein, the only conclusive solution is simply to maximize incentives to invent. To do this, one must appeal to the famous Blackstonian conception of property as “that sole and despotic dominion which one man claims and exercises over the external things of the world, in total exclusion of the right of any other individual in the universe.”¹¹ Only by granting patentees an unqualified right to exclude, thus enabling them to appropriate the full social value of their inventions, can society be sure that it will not mistakenly under-incentivize valuable innovation. Many have

7. See Michael W. Carroll, *One For All: The Problem of Uniformity Cost in Intellectual Property Law*, 55 AM. U. L. REV. 845 *passim* (2006); Alan O. Sykes, *TRIPs, Pharmaceuticals, Developing Countries, and the Doha “Solution,”* 3 CHI. J. INT’L L. 47, 57 (2002).

8. See Thomas O. Barnett, Remark, *Maximizing Welfare Through Technological Innovation*, 15 GEO. MASON L. REV. 1191, 1194–96 (2008); Joseph E. Brodley, *The Economic Goals of Antitrust: Efficiency, Consumer Welfare, and Technological Progress*, 62 N.Y.U. L. REV. 1020, 1026 (1987) (explaining that dynamic efficiency is by far the most important form of efficiency); Jedediah Purdy, *People As Resources: Recruitment and Reciprocity in the Freedom-Promoting Approach to Property*, 56 DUKE L.J. 1047, 1051 n.4 (2007) (noting that “[d]ynamic efficiency works perfectly only with perfect internalization of all benefits and costs attending the exercise of one’s property rights”).

9. See DAN L. BURK & MARK A. LEMLEY, *THE PATENT CRISIS AND HOW THE COURTS CAN SOLVE IT* 38–40 (2009).

10. See *id.* at 38–40, 156–58.

11. See 2 WILLIAM BLACKSTONE, *COMMENTARIES ON THE LAWS OF ENGLAND* 2 (Univ. of Chi. Press 1979) (1765).

articulated this view, which may be characterized as the maximalist perspective.¹²

Despite its ostensible attraction, the maximalist approach is fundamentally flawed. The most obvious frailty lies in an axiomatic fact: it is impossible to create property rules that enable owners to prevent externalities.¹³ This is especially so with respect to information goods, which are typically characterized as non-excludable.¹⁴ The law is therefore incapable of enabling inventors to capture the full value of their inventions, for some free-riding will always take place. Moreover, patentees would need to engage in perfect price discrimination, which is evidently impossible.¹⁵ But even if the law enjoyed the capacity to award inventors Blackstonian property rights, there are several reasons why it should decline to bestow them.

First, tracing problems would give rise to preclusive transaction costs.¹⁶ Given the cumulative nature of innovation, almost every increment in knowledge implicates the contributions of myriad inventors.¹⁷ If the law granted each such inventor a right to enjoin the use of whatever information he added, the search and negotiation costs required to advance technology would quickly become preclusive.¹⁸ Second, were a maximalist regime successfully implemented in pure form, society's members would be indifferent to the fact of innovation, for each would be charged his or her

12. See Rebecca S. Eisenberg, *Patents and the Progress of Science: Exclusive Rights and Experimental Use*, 56 U. CHI. L. REV. 1017, 1025–26 (1989) (addressing the concept of expansive patent rights); Einer Elhauge, *Do Patent Holdup and Royalty Stacking Lead to Systematically Excessive Royalties?*, 4 J. COMPETITION L. & ECON. 535, 541 (2008) (arguing that if patentees are denied the ability to extract the full social value of their inventions, they will devote suboptimal resources to the innovative process); Dennis Michaels, *Bioprospecting Agreements: Forging a Comprehensive Strategy for Managing Genetic Resources on Public Lands*, 22 ENVIRONS ENVTL. L. & POL'Y J. 3, 64 (1999); see also Nuno Pires de Carvalho, *The Primary Function of Patents*, 2001 U. ILL. J.L. TECH. & POL'Y 25, 30 (opining that “the [patent] reward should be tantamount to the invention’s worth”); Vincent P. Tassinari, *Patent Compensation Under 35 U.S.C. § 284*, 5 J. INTELL. PROP. L. 59, 154 (1997) (opining that the “American public has a right to expect that its inventors [receive] the full value of their invention[s] in the marketplace”).

13. See Mark A. Lemley, *Property, Intellectual Property, and Free Riding*, 83 TEX. L. REV. 1031, 1065 (2005).

14. See Kristen Osenga, *Information May Want to Be Free, But Information Products Do Not: Protecting and Facilitating Transactions in Information Products*, 30 CARDOZO L. REV. 2099, 2141 (2009).

15. See Norman Siebrasse, *A Property Rights Theory of the Limits of Copyright*, 51 U. TORONTO L.J. 1, 40 n.94 (2001); Oliver E. Williamson, *Economies as an Antitrust Defense Revisited*, 125 U. PA. L. REV. 699, 720 (1977).

16. See LANDES & POSNER, *supra* note 3, at 214.

17. See *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 427 (2007).

18. *Cf. id.*

reservation price for availing of it. Third, the utilitarian case for a frugal approach is compelling. Every additional dollar awarded to an innovator beyond what was necessary to spur her to invent and commercialize her invention is a dollar wasted.¹⁹ Such a system would result in windfalls that provide no larger gain and, indeed, bear a cost. Fourth, innovation requires not only ex ante incentives in terms of prospective reward, but it also depends intimately on the availability of contemporary know-how. Complete proprietization of information reduces access to technological information, thus diminishing the scientific knowledge available for follow-on innovation.²⁰

Patent law has clearly declined to grant inventors the full social value of their technological contributions.²¹ This is unsurprising, but, as a more realistic alternative, should the law err on the side of generous overcompensation? In other words, perhaps the law would rightly eschew a dogmatic interpretation of the maximalist approach, but would resolve to interpret ambiguities in favor of significantly over-rewarding inventors. This would be a more reasonable approach to decision-making under conditions of uncertainty. Decision theory, which is the branch of microeconomics that deals with choice in the presence of indeterminacy, often suggests that policymakers should favor Type II errors (false negatives) over Type I errors (false positives).²² Such an approach would find greater harm from a mistaken denial of intellectual-property protection to a deserving inventor than an erroneous patent award to an unworthy innovator. Therefore, one might expect an informed IP system to reflect this bias and maximize inventor compensation to the extent permitted by the statutory framework put in place by Congress. This Article refers to such a policy as the “qualified-maximalist” approach.

19. See David S. Abrams, *Did TRIPS Spur Innovation? An Analysis of Patent Duration and Incentives to Innovate*, 157 U. PA. L. REV. 1613, 1615 (2009).

20. See, e.g., C. Scott Hemphill, *Paying for Delay: Pharmaceutical Patent Settlement as a Regulatory Design Problem*, 81 N.Y.U. L. REV. 1553 (2006) (discussing the importance of follow-on innovation to patent policy).

21. See *infra* Part III. Time-limited monopoly is the most obvious example of this determination. See Samson Vermont, *A New Way to Determine Obviousness: Applying the Pioneer Doctrine to 35 U.S.C. § 103(a)*, 29 AIPLA Q.J. 375, 400 (2001). Limits on patentees' ability to contract with potential rivals and licensees for greater protection of their IP and denying patentees an automatic right to injunctive relief provide further examples. Thus, it is clear that a literal interpretation of the maximalist approach is not welcomed by the contemporary patent system.

22. See Melvin Aron Eisenberg, *Bad Arguments in Corporate Law*, 78 GEO. L.J. 1551, 1553 (1990) (“Wrongly considering that a theory is incorrect is known as a Type I error. Wrongly considering that a theory is correct is known as a Type II error.”).

Were the law presently to reflect this policy, we would expect to detect its presence within the current doctrine. Such a system would presumably entitle patentees to injunctive relief in all cases of proven infringement of a valid patent. There would also be no limitation on the patenting of processes beyond the traditional requirements of novelty, utility, nonobviousness, and statutory bars. Patent-holders would enjoy the benefit of an expansive doctrine of equivalents, which would serve to capture subsequent inventions that encapsulate the nature of patentees' innovation, but not their claims. Patentees would have the right to contract with licensees to prevent them from challenging the validity of licensed patents, to require royalties beyond the life of the patents, to bundle or otherwise to tie the pertinent technology to non-patented goods, and to dictate price terms by which licensees must abide in selling the goods incorporating the patented technology.

Yet we do not see any of these things.²³ Patent law has therefore discarded a qualified-maximalist path in its jurisprudence. Instead, this Article contends that a parsimony principle guides the patent system: the law attempts to cabin patentees' compensation and to prevent windfalls. It declines to achieve this goal in a formalistic manner through the adoption of categorical rules, as recent developments make clear.²⁴ Instead, the law adopts an incremental approach to resolving policy issues that emerge over time. Adopting all evidence that may be available, current patent doctrine imposes meaningful and evolving limits on patentees' exclusive rights. In doing so, the law appears to accept at least some risk of Type I errors in certain settings. The Article concludes that the parsimony principle possesses significant power in explaining the current makeup of patent law, and thus offers a helpful framework within which to resolve contemporary challenges facing the patent system.

Despite this, the qualified-maximalist approach may be detectable in at least one regard: the apparent reality of leading innovators' seemingly massive overcompensation. The phenomenon is most obvious in the realm of copyright: leading artists, both in film and music, command returns that dwarf those of similarly talented individuals who lack sufficient fortune to make it to the very top.²⁵ These two industries are prime examples of winner-

23. See *infra* Part III.

24. See, e.g., *Bilski v. Kappos*, 130 S. Ct. 3218, 3222 (2010) (declining to adopt a rigid test for determining the contours of patentable subject matter).

25. See LANDES & POSNER, *supra* note 3, at 49–50. The closest analogy to the patent field appears to be the pharmaceutical industry, where leading companies regularly make astronomic profits. However, the ostensible fact of overcompensation may mask complex, capital-intensive, and inherently risk-filled innovation structures.

takes-all markets, which grant vast rewards to the few chart-toppers and relatively meager profits to all others, including those who are close seconds.²⁶ The windfalls enjoyed by market-topping singers and actors come at considerable social cost. Driven by the prospect of dizzying financial fortune, prospective stars flock to these industries in excessive numbers.²⁷ Moreover, the windfall awarded leading artists accentuates deadweight loss, as these excessive dollars drive a greater wedge between the marginal cost of producing and disseminating art and the ultimate cost paid by the consumer.²⁸ As a result, leading artists' excessive returns price some consumers out of the market.

Overcompensation consistent with the qualified quasi-maximalist approach also appears to be present in the patent context. Commentators often single out pioneer-drug manufacturers for enjoying excessive financial rewards,²⁹ though the pharmaceutical industry is not the only possible example. It may well be the case that certain (even many) inventors of business methods and computer software receive greater returns from the patent system than would be required to spur their innovative efforts in the first place.³⁰ This is suggested by the rancorous debate as to whether innovation in these fields is properly subject to patent protection at all, given the widespread, though not uncontested, view that adequate incentives exist to develop these technologies that are independent of the patent system.³¹ It is also evidenced by the fact that innovation proceeded in fine fashion in these fields before widespread patenting of these types of inventions was made possible by the Federal Circuit's decision in *State Street Bank & Trust Co. v. Signature Financial Group, Inc.*³² in 1998.

26. *See id.*

27. This is a fundamental result of uncontroversial economic theory.

28. *See* DENNIS W. CARLTON & JEFFREY M. PERLOFF, *MODERN INDUSTRIAL ORGANIZATION* 95–96 (4th ed. 2005).

29. *See* Jessica L. Greenbaum, *TRIPS and Public Health: Solutions for Ensuring Global Access to Essential AIDS Medication in the Wake of the Paragraph 6 Waiver*, 25 J. CONTEMP. HEALTH L. & POL'Y 142, 152 (2008).

30. *See* Peter S. Menell, *A Method for Reforming the Patent System*, 13 MICH. TELECOMM. & TECH. L. REV. 487, 497, 506 (2007); Hans A. von Spakovsky et al., *The Limited Patenting of Computer Software: A Proposed Statutory Approach*, 16 CUMB. L. REV. 27, 44–45 (1986).

31. *See, e.g., In re Bilski*, 545 F.3d 943, 1005 (Fed. Cir. 2008) (en banc) (Mayer, J., dissenting), *aff'd but criticized*, *Bilski v. Kappos*, 130 S.Ct. 3218 (2010); Arti K. Rai, John R. Allison & Bhaven N. Sampat, *University Software Ownership and Litigation: A First Examination*, 87 N.C. L. REV. 1519, 1521 (2009).

32. 149 F.3d 1368, 1373 (Fed. Cir. 1998) (adopting the principle that a process need only produce a useful, concrete, and tangible result to be patentable); *see also* John R. Allison & Emerson H. Tiller, *The Business Method Patent Myth*, 18 BERKELEY TECH. L.J. 987, 991 &

Certain patentees receive far greater pecuniary rewards than would have been required for them to produce those particular inventions *ex ante*. Does this suggest that the patent system adopts a qualified-maximalist approach? The answer is no because the windfall enjoyed by certain artists is highly distinguishable from the seemingly exorbitant returns commanded by patentees in certain industries. The distinction lies in the fact of cross-subsidization.³³ Industrial innovation, most obviously in the pharmaceutical sector, is characterized by large-scale R&D over a large number of diversified projects.³⁴ Companies operating in this environment rely on large profits from successful endeavors to cover losses from unfruitful ones.³⁵ To create a solid platform for ongoing innovation, these profits thus need to be of an order of magnitude greater than the cost of developing and commercializing the successful inventions.³⁶ It has been estimated that only one out of 10,000 research endeavors into potentially viable drugs will ultimately result in a marketable product.³⁷ Thus, far from indicating the adoption of a qualified-maximalist principle, hugely profitable patents are entirely consistent with a parsimonious approach.

Of course, this is not to suggest that contemporary doctrine is optimally tailored or that the IP system does not over-reward patentees in some industries. As this Article concludes, room for improvement does exist in the current patent regime and the necessary means for attaining such gains are fully supported by patent law's parsimony principle. This holds true even if that principle has not yet achieved unqualified or flawless application.

Although patent law has generally flouted a maximalist or quasi-maximalist approach to spurring innovation, the patent system has instead followed a parsimonious path in constructing a system of incentives for spurring innovation.³⁸ Questions of patentable subject matter, which have forever vexed the courts, strongly implicate this principle, as do the doctrine of equivalents, proper remedies in the event of infringement, the misuse doctrine, and other aspects of patent policy. The parsimony principle bears

n.10 (2003) (illustrating the dramatic rise in business-method patent claims in the wake of *State Street Bank*).

33. See Brian J. Dunne, *Compatibility Requirements, Systems Competition, and Copyright*, 39 U. TOL. L. REV. 813, 822 n.52 (2008).

34. See *id.*

35. See *id.*

36. Only 30% of FDA-approved drugs recoup their R&D costs. See Christopher L. Lockwood, Comment, *Biotechnology Industry Organization v. District of Columbia: A Preemptive Strike Against State Price Restrictions on Prescription Pharmaceuticals*, 19 ALB. L.J. SCI. & TECH. 143, 148 (2009).

37. *Id.*

38. See *infra* Part II.

great explanatory power in elucidating the current makeup of the law. For instance, it explains the convoluted series of U.S. Supreme Court decisions governing the sphere of patentable subject matter.³⁹ Yet, the principle also yields normative value in resolving pressing issues that presently afflict the patent system. This Article explains that patent law's parsimony principle is a desirable lodestar, indicates how the law has largely evolved in a manner consistent with this principle, and offers some thoughts on how this approach can help guide the courts, Congress, and the PTO through the anticipated turbulent times ahead. The Article also points to a number of caveats, however, that suggest that dogmatic adherence to the parsimonious approach would be improper in all cases.

II. THE CASE FOR PARSIMONIOUS PATENT RIGHTS

A. THE PATENT SYSTEM'S CHARGE: SPURRING INNOVATION

Patent law is charged with propelling the U.S. economy forward by fostering a vigorous and robust process of innovation,⁴⁰ which is no simple task. Invention, innovation, technological advancement, and similar synonyms all capture the idea of scientific progress, but such progress takes place in wholly dissimilar fashions across a bewildering array of industries.⁴¹ The incentives that drive innovation in one context may be quite different from those relevant to others.⁴² Of course, the almighty dollar is a leading incentive across industries generally, but the relative weight to be attached to purely pecuniary rewards will differ, as will the amount required to spur optimal levels of R&D. Moreover, the task of spurring technological progress is complicated by the fact of industrial evolution itself. Scientific development continues to blaze ahead, which is no doubt a testament to the intellectual-property regime already in place.⁴³ Many of the industrial contexts within which innovation now takes place, however, could not have been

39. See *Diamond v. Chakrabarty*, 447 U.S. 303 (1980); *Parker v. Flook*, 437 U.S. 584 (1978); *Gottschalk v. Benson*, 409 U.S. 63 (1972).

40. This mandate emanates from the U.S. Constitution, which grants Congress the power to "promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." U.S. CONST. art. I, § 8, cl. 8.

41. See Dan L. Burk & Mark A. Lemley, *Policy Levers in Patent Law*, 89 VA. L. REV. 1575, 1581–89 (2003).

42. See 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, 1840 (2008).

43. See Patrick Doody, *The Patent System Is Not Broken*, 18 INTEL. PROP. & TECH. L.J. 10, 13 *passim* (2006).

within the Framers' contemplation at the time the U.S. Constitution was adopted.⁴⁴ Nor were many crucial areas of innovation foreseeable by Congress at the passage of the 1952 Patent Act.⁴⁵ That piece of legislation remains Congress's latest, all-encompassing attempt to fulfill its constitutional prerogative of advancing science and the useful arts.⁴⁶ The Act does so by creating ownership rights in the fruit of innovation.

Yet those rights were crafted with more traditional industries in mind. The specified bundle of rights associated with a patent grant may have been designed for sectors that require significant monetary outlays for innovation and that produce goods implicating only a limited number of patented products or processes.⁴⁷ It is possible that exclusive rights crafted for innovation in pharmaceutical, mechanical, or other capital-heavy industries may be well suited to those contexts, yet find incongruous application to new-economy fields founded on computer software or information technology.⁴⁸ The latter fields give rise to myriad, overlapping patent rights. It is not unusual, for instance, for a semiconductor chip to implicate thousands of patents.⁴⁹ Even a single patentee's obstinate refusal to license can enjoin commercialization of a product that implicates his technology only in a peripheral manner.⁵⁰ If a court declines to award injunctive relief, conscious infringement of a patent will give rise to treble-damages liability.⁵¹ Computer-software patents have been accused of asphyxiating follow-on research, which frustrates the enthusiastic efforts of individual programmers who seek to build on prior knowledge.⁵² Across a number of high-technology sectors, huge numbers of overlapping patent rights, which give rise to ubiquitous blocking scenarios, bear the potential to turn the patent regime on

44. See *Bilski v. Kappos*, 130 S. Ct. 3218, 3227–28 (2010).

45. See *id.*

46. 35 U.S.C. § 1 (2006).

47. See Daniel R. Cahoy & Leland Glenna, *Private Ordering and Public Energy Innovation Policy*, 36 FLA. ST. U. L. REV. 415, 447 n.167 (2009); Graeme B. Dinwoodie & Rochelle C. Dreyfus, *Diversifying Without Discriminating: Complying with the Mandates of the TRIPS Agreement*, 13 MICH. TELECOMM. & TECH. L. REV. 445, 445–46 (2007).

48. See, e.g., JAMES BESSEN & MICHAEL J. MEURER, *PATENT FAILURE: HOW JUDGES, BUREAUCRATS, AND LAWYERS PUT INNOVATORS AT RISK* 27 (2008).

49. Deborah Platt Majoras, *A Government Perspective on IP and Antitrust Law*, 38 RUTGERS L.J. 493, 497 (2007).

50. For this reason, companies in patent-laden industries typically prefer weaker patent rights. See Jay P. Kesan & Andres A. Gallo, *The Political Economy of the Patent System*, 87 N.C. L. REV. 1341, 1352–53 (2009).

51. *In re Seagate Tech., LLC*, 497 F.3d 1360, 1368 n.3 (Fed. Cir. 2007) (en banc).

52. See James Ernstmeier, Note, *Does Strict Territoriality Toll the End of Software Patents?*, 89 B.U. L. REV. 1267, 1276–77 (2009); Stephen McJohn, *Scary Patents*, 7 NW. J. TECH. & INTELL. PROP. 343, 367 (2009).

its head by braking, rather than driving, innovation.⁵³ Presumably, such a phenomenon was not within the foresight of Congress in 1952.⁵⁴ Yet many inventors in the new economy remain in need of pecuniary compensation, so some exclusive rights or alternative incentives systems (such as prizes or rewards) are surely necessary.

So what form should the exclusive rights associated with a patent take? To the uninitiated, this might appear to be an odd question. After all, property rights conferred on inventors should presumably be akin to ownership rights in more traditional contexts of realty and personalty.⁵⁵ Owners of such items largely have free reign to dictate the ends to which their property can be employed.⁵⁶ For example, the owner of a piece of land has a virtually unqualified right to control its use.⁵⁷ Anyone wishing to employ the property to his benefit must bargain for permission from the landowner, as the exclusivity inherent in a property right allows owners to extract the social value that flows from the land's use. Such rights are justified, in spite of the social costs they impose by denying third-party access, because they align private owners' incentives with the social optimum.⁵⁸ They do so by encouraging owners to devote capital to improvement and to consume resources in a responsible fashion. In the absence of property rights, third parties would appropriate the benefits of the improver's private efforts, thus eviscerating any incentive for the latter to undertake similar efforts in the future. Similarly, economics teaches that un-owned, scarce resources will be over-consumed—a phenomenon typically referred to as the tragedy of the commons.⁵⁹

The benefits of ownership in technology are similar, albeit not identical. Without a right to exclude others from availing of their novel technological insights, inventors may decline to devote the necessary capital to uncover like

53. Carl Shapiro, *Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting*, in 1 INNOVATION POLICY AND THE ECONOMY 119, 121 (Adam B. Jaffe et al. eds., 2001).

54. See David B. Conrad, *Mining the Patent Thicket: The Supreme Court's Rejection of the Automatic Injunction Rule in eBay v. MercExchange*, 26 REV. LITIG. 119, 143–44 (2007).

55. For a discussion of the possible value of treating patents as giving rise to strong property rights, see F. Scott Kieff, *Property Rights and Property Rules for Commercializing Inventions*, 85 MINN. L. REV. 697, 703 (2001); R. Polk Wagner, *Information Wants to Be Free: Intellectual Property and the Mythologies of Control*, 103 COLUM. L. REV. 995, 997 (2003).

56. See, e.g., *Allred v. Harris*, 14 Cal. App. 4th 1386, 1390 (1993).

57. See Richard R.W. Brooks, *The Relative Burden of Determining Property Rules and Liability Rules: Broken Elevators in the Cathedral*, 97 NW. U. L. REV. 267, 313–14 (2002).

58. See STEVEN SHAVELL, FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW 11–22 (2004).

59. See Garrett Hardin, *The Tragedy of the Commons*, 162 SCI. 1243, 1244–45 (1968).

insights in the future.⁶⁰ The free-rider problem is no less endemic in the intellectual realm than it is in the physical. Indeed, it is apt to be considerably worse, given the widely accepted fact that knowledge goods are more easily appropriated than tangible ones.⁶¹ Thus, at least in principle, there is no obvious reason why property rights should be any different in the technological sphere of patent law than they are in more traditional contexts. As Judge Easterbrook has explained, “[p]atents give a right to exclude, just as the law of trespass does with real property.”⁶² Yet, unlike the tragedy of the commons, information is a public good, such that it is not vulnerable to overconsumption once an inventor has garnered a sufficient return to justify the initial research and investment.

But is it really the case, as the Patent Act suggests, that patents should provide their owners similar rights as those enjoyed by owners of traditional property?⁶³ Does a powerful right to exclude that allows a qualifying inventor to extract all, or much, of his discovery’s social value best conform with the constitutional objective of promoting technology? There are certainly some who think so.⁶⁴ Viewed from the maximalist perspective, patent law should enable innovators to prevent any third party from gratuitously benefiting from their inventions.

However, there are strong reasons to reject this approach. Anyone who attempted to implement such a guiding principle would be engaging in a fool’s errand not only because the costs of such an approach would surely outweigh the gains, but also because any such efforts would be in vain. It is simply not possible to construct property rights that allow owners to internalize all beneficial uses of their discovery.⁶⁵ Moreover, traditional

60. See Ariel Porat, *Private Production of Public Goods: Liability for Unrequested Benefits*, 108 MICH. L. REV. 189, 224 n.101 (2009).

61. See Clarisa Long, *Patent Signals*, 69 U. CHI. L. REV. 625, 631–32 (2002).

62. Frank H. Easterbrook, *Intellectual Property Is Still Property*, 13 HARV. J.L. & PUB. POLY 108, 109 (1990).

63. 35 U.S.C. § 261 (2006) (“[P]atents shall have the attributes of personal property.”).

64. But see Adam Mossoff, *Exclusion and Exclusive Use in Patent Law*, 22 HARV. J.L. & TECH. 321, 323–26 (2009) (explaining that the traditional exclusion concept of the patent system is inadequate).

65. The first obstacle is an obvious one: to extract all social value from an invention, the inventor’s rights would have to be perpetual. Yet a system of ever-lasting property rights over a valuable invention would give rise to preclusive transaction and tracing costs. We might overcome this objection by observing that inventors with sufficiently high discount rates will care only about returns that they can derive in the short or medium run. Presumably, not all inventors will care about the value of IP that lasts beyond their deaths. But even in those situations it is impossible to prevent some externalities. Just as passersby benefit from the aesthetic quality of an improved house—a gain that the homeowners

principles of ownership cover realty and personalty, whose attributes can be vastly different than those of information goods.⁶⁶ This Article thus argues that bestowing inventors with *full* control over the disposition of their technologies is folly. The award of exclusive rights over knowledge carries social costs that increase in proportion with the elasticity of demand for that information.⁶⁷ Where such elasticity is high, increases in price beyond the marginal cost will result in sharp decreases in demand and hence output. This result clearly has serious social welfare losses. By granting innovators an absolute ability to exclude others, not only will monopoly prices proliferate, but downstream uses of the relevant technology for follow-on innovation will also be hindered. Real-life complaints of such effects abound in certain sectors, most notably in the field of computer software.⁶⁸ Expanding ownership rights in information can be expected to increase social welfare for a time, but it would also be anticipated that further proliferation of property rights eventually will choke access to knowledge, hinder the commercialization of products that implicate numerous technologies, over-reward initial inventors, and transfer wealth from consumers in an unjustified manner.

A maximalist approach to patent policy is therefore impractical and imprudent, given the overwhelming social costs involved in widespread overcompensation.⁶⁹ Excessively broad property rights over information may

cannot charge for—so too do inventors lack means to prevent third parties from benefiting in some indirect way from discoveries that they unearth.

66. For the author's extended thoughts on the relationship between tangible and intellectual property, see generally Alan Devlin, *Indeterminism and the Property-Patent Equation*, 28 YALE L. & POL'Y REV. 61 (2009).

67. See David W. Opperbeck, *Patent Damages Reform and the Shape of Patent Law*, 89 B.U. L. REV. 127, 131, 175–76 (2009).

68. See, e.g., Carl Shapiro, *Patent System Reform: Economic Analysis and Critique*, 19 BERKELEY TECH. L.J. 1017, 1018 (2004).

69. Contrary to what some might initially believe, there is such a thing as excessive innovation. Money devoted to R&D in future technology can, and in some instances will, generate a greater marginal gain for society by being devoted instead to alternative ends. Windfall profits for inventors—unless necessary to subsidize a broader platform of innovation—create serious distortions. The best-known example is the “gold-rush” phenomenon associated with winner-takes-all markets. The prospect of windfall gains acts as a powerful magnet, attracting more competition than is warranted from a societal viewpoint. Some innovators will eschew safer, and more socially valuable, investment opportunities to obtain a shot at the artificial reward at the end of such an innovation rainbow. Moreover, a maximalist patent system will, by definition, be of further reach than one tailored to the specific characteristics of the relevant industry or inventor. The further spread of property rights in the information field creates search and negotiation costs on the part of consumers, intermediaries, and downstream innovators that would not otherwise exist. And beyond these more nuanced concerns lies the obvious objection that consumers of patented

create windfall profits for some, but generally would leave society with wholly disproportionate losses.⁷⁰ Thus, property rights must be expanded with caution. The catch of course is that, without some reward, supply is apt to diminish or even to disappear, so it is no solution to jettison property rights in information altogether, though some radically-minded scholars have advocated this view.⁷¹

Instead, aggregate welfare can be maximized by ascertaining the optimal trade-off between, on one hand, the relative costs and benefits of free access to information for consumption and follow-on innovation, and, on the other hand, ownership rights that guarantee successful inventors a meaningful pecuniary return. The optimum level of protection occurs where the incremental social gain from a marginal increase in exclusivity equals the concomitant increase in social costs.⁷² As is often the case in economics, however, it is easier to state the solution generally than it is to propose and implement a plan with any reasonable precision. And unfortunately, policymakers are bereft of much of the information that they would need to tailor property rights in this optimal manner. It is an unremarkable fact that people are motivated by a host of idiosyncratic factors that span a vast spectrum. Spurring innovation thus presents huge conceptual and practical challenges.

It is therefore fascinating to consider how patent law has attempted to fulfill its constitutional mandate of promoting the useful arts.⁷³ Economic theory tells us that each inventor's reward should be tailor-made.⁷⁴ For instance, under an economic theory, the academic whose research yields profound insights into a new area of technological endeavors should not receive property rights in his discoveries if the prospect of tenure or prestige in the academic community offers sufficient compensation and if the university does not rely on pecuniary returns from the patented inventions of its faculty to fund the institution. Similarly, an inventor who solves a technical problem for her own benefit should be barred from claiming a

technology are required to pay more money than was necessary to reward the relevant inventors. Wealth transfers might in themselves be viewed as objectionable, but the artificially enhanced prices associated with a maximalist patent regime reduce consumption because most consumers' demand curves are downward-sloping. This translates into deadweight loss and a perpetual elimination of aggregate social wealth.

70. See Eisenberg, *supra* note 12, at 1026–28.

71. See generally Mark D. Janis, *Patent Abolitionism*, 17 BERKELEY TECH. L.J. 899 (2002) (discussing the history of patent abolitionism).

72. See LANDES & POSNER, *supra* note 3, at 66.

73. See U.S. CONST. art. I, § 8, cl. 8.

74. See LANDES & POSNER, *supra* note 3, at 300.

patent over her solution. The same should hold true for those who innovate primarily out of the joy that the process brings them. Some people love writing books; others adore painting; others still derive immense satisfaction from developing and improving software. From the technical standpoint of economics, none of these individuals should be eligible for property protection, at least insofar as they would continue innovating without pecuniary subsidization from the IP system.⁷⁵ But those inventors who face serious R&D and commercialization costs require strong exclusive rights if they are to engage in innovation. The pharmaceutical and biotechnology industries constitute obvious examples.⁷⁶

Even this very brief foray into the world of innovation reveals that no single standard or rule can spur all desirable creativity and invention. Rewards that are sufficient to some may be wholly inadequate to others, but compensation that is acceptable to all inventors will inevitably over-reward many. Those responsible for formulating patent doctrines are thus faced with a quagmire. They know that different inventors will react to a single body of incentives in disparate ways, but they do not possess sufficient information about each inventor to tailor innovator-specific property rights. Or, to be more precise, such information is not available at acceptable cost. Indeed, because it is so difficult to determine why particular inventors innovate, patent law has typically declined to incorporate inventor- or even industry-specific principles into its doctrine.⁷⁷ Instead, the patent laws generally operate on a “one-size-fits-all” basis, attempting to spur optimal levels of innovation through the provision of largely uniform reward structures.⁷⁸ As a result, it is uncontested that patent law currently operates in an imperfect

75. The only complicating factor concerns commercialization. To transform a conceptualized invention into marketable products requires considerable investment, particularly when the products cannot be disseminated digitally. Inventors who innovate for the joy of invention should be awarded patent protection when such ownership rights are necessary to spur commercialization. *See* Kieff, *supra* note 55.

76. *See* Benjamin N. Roin, *Unpatentable Drugs and the Standards of Patentability*, 87 TEX. L. REV. 503, 511 (2009).

77. *See* Michael W. Carroll, *One for All: The Problem of Uniformity Cost in Intellectual Property Law*, 55 AM. U. L. REV. 845 (2006) (exploring the one-size-fits-all nature of much of the U.S. patent system). *But see* BURK & LEMLEY, *supra* note 9 (exploring certain ways in which the courts have applied ostensibly identical patent rules in different ways depending upon the industry at issue).

78. *See, e.g.,* Rebecca S. Eisenberg, *Patents, Product Exclusivity, and Information Dissemination: How Law Directs Biopharmaceutical Research and Development*, 72 FORDHAM L. REV. 477, 486 (2003) (“Our patent laws are one-size-fits-all, applying essentially the same rules to biopharmaceutical research that apply to automotive engineering, information technology, semiconductors, and rocket science. But the needs of these fields for patent protection differ.”).

manner, at least insofar as when its efficacy is measured by its ability to spur optimal levels of innovation.⁷⁹

Despite this melancholy observation, it does not follow that patent law's one-size-fits-all approach is undesirable from a policy perspective. Tailoring rules to context-specific scenarios necessarily involves a cost because courts have to make case-by-case determinations. That cost is, of course, eliminated by the adoption of uniform rules. Such uniformity is warranted when the cost of attempting to convey accurate rewards exceeds the benefit that such specificity would bestow. Given the limited access to information enjoyed by policymakers, the cost of accurately crafting inventor-specific rewards would be high indeed.⁸⁰

There is therefore at least some basis for believing that patent law's relatively uniform approach is justified. The question thus arises of what particular set of rules will act as the optimal heuristic, and we are thus prompted to enter the domain of error analysis. When called upon to craft rules in an indeterminate environment, one must inquire into the relative costs of erring on one side rather than the other. In the criminal context, for instance, it is widely accepted that mistaken guilty verdicts are more serious than erroneous judgments of innocence.⁸¹ In the patent field, one question of

79. See Dinwoodie & Dreyfus, *supra* note 47, at 446 (questioning whether patent law's "one-size-fits-all" approach can survive).

80. Some industry-specific modifications in doctrine are both cost-justified and compelled by a parsimonious approach. Accord Eric E. Johnson, *Calibrating Patent Lifetimes*, 22 SANTA CLARA COMPUTER & HIGH TECH. L.J. 269, 269, 285–89 (2006). The fact remains, however, that sufficient information to tailor rewards precisely to individual inventors' incentive characteristics remains unavailable. Nevertheless, to the extent that more efficient heuristics can be derived from observable differences in various industries' innovation profiles, they should obviously be employed. For example, one can safely observe that significant rewards are required to cover the level of capital required for R&D in the pharmaceutical sector. Yet one can also point to the fact that computer software is copyright-protectable and that innovators in this field will receive some reward even in the absence of patent protection. Therefore, a parsimonious approach to patent doctrine would counsel broader exclusive rights in the pharmaceutical rather than the computer-software industry. This discussion is expanded upon below. See *infra* Part IV.

81. Judge Harlan explains:

In a criminal case . . . we do not view the social disutility of convicting an innocent man as equivalent to the disutility of acquitting someone who is guilty. . . .

. . . .
 . . . [T]he requirement of proof beyond a reasonable doubt in a criminal case [is] bottomed on a fundamental value determination of our society that it is far worse to convict an innocent man than to let a guilty man go free.

In re Winship, 397 U.S. 358, 372 (1970) (Harlan, J., concurring).

significance might be whether overcompensation or under-compensation is worthy of greater concern. Were policymakers forced to operate in an environment of complete uncertainty—that is, if they were denied any ability to ascribe probabilities to various outcomes—the answer to this question would be determinative. If excessive levels of innovation are preferable to insufficient levels of such activity, we should then simply maximize innovation.

This path finds its definition in the “maximalist” approach to patent policy, which this Article has already rejected for the multitude of reasons.⁸² Nevertheless, were we to provide inventors with an unqualified ability to exclude others from benefiting from their discoveries, they would be able to appropriate the full social value of their inventions. When subject to the incentives associated with such rewards, inventors would decline to innovate only if the cost of invention and commercialization exceeded the social value created by the ensuing technology. Thus, a maximalist regime ensures that all socially desirable breakthrough (as opposed to cumulative) innovation takes place.⁸³ For this reason, some commentators tout the benefit of a patent system that would bestow inventors with the greatest possible value that can be tied to their innovation.⁸⁴

B. PROMOTING A PARSIMONIOUS APPROACH TO PATENT POLICY

This Article declines to follow the maximalist ideal because policymakers are blessed with more than zero information about the innovative process. We need not eliminate all Type I errors in innovation (under-compensation) by accepting swathes of Type II errors (overcompensation). It is surely the case that allowing inventors to extract the entire social value of their inventions will exceed the optimal level of protection in most cases. Thus, patent law can achieve superior outcomes by eschewing wild biases in favor

82. See *supra* Part I.

83. See W. KIP VISCUSI ET AL., *ECONOMICS OF REGULATION AND ANTITRUST* 92, app. (1992) (“If the inventor cannot expect to appropriate all of the economic value of his invention, he will under-invest in inventive activity.”); David E. Adelman & Kirsten H. Engel, *Reorienting State Climate Change Policies to Induce Technological Change*, 50 ARIZ. L. REV. 835, 849 (2008); Gideon Parchomovsky & Peter Siegelman, *Towards an Integrated Theory of Intellectual Property*, 88 VA. L. REV. 1455, 1459 (2002); Jaideep Venkatesan, *Compulsory Licensing of Nonpracticing Patentees After eBay v. MercExchange*, 14 VA. J.L. & TECH. 26 (2009) (suggesting that it might be proper to allow patentees to derive the “full value” of their inventions in some settings). *But see* Ian Ayres & Paul Klemperer, *Limiting Patentees’ Market Power Without Reducing Innovation Incentives: The Perverse Benefits of Uncertainty and Non-Injunctive Remedies*, 97 MICH. L. REV. 985, 987 *passim* (1999) (challenging the maximalist view); Louis Kaplow, *The Patent-Antitrust Intersection: A Reappraisal*, 97 HARV. L. REV. 1813, 1828 (1984).

84. See sources cited *supra* note 83.

of under- or over-rewarding inventors and instead attempting to approximate more nuanced rules. It can do so by reference to the unique and visible characteristics of different industries—characteristics that enable patent law to mollify or magnify its reward regime depending on the context at hand.⁸⁵ Of course, patent law’s one-size-fits-all structure places significant restrictions on this ability, but it is widely accepted that the patent system has a number of “policy levers” that the courts can, and have, manipulated to society’s advantage.⁸⁶

Abandoning the pure maximalist approach, both on pragmatic and normative grounds, we should ask how strong patent rights ought to be in light of the framework adopted by Congress. Put differently, if patent law rightly rejects a maximalist lodestar, what guiding principle is left in its stead? Precisely tailoring of property rights’ breadth and duration to individual inventors’ incentive characteristics is wholly unrealistic, so some sort of guide is required. A number of possibilities emerge. One avenue is to reject the pure maximalist approach, which seeks to eliminate any Type I error, and instead to err on the side of significant overprotection in cases of uncertainty. This path, coined the “quasi-maximalist” approach by this Article, would grant patentees generous property rights. More specifically, the quasi-maximalist approach would recognize (both on normative and positive grounds) the limits on patentee compensation established by Congress, such as constraints on duration and breadth. But it would give full force to the rights espoused by the legislature, granting qualified inventors as much protection as would be consistent with the statutory language. Although some limited instances of under-compensation would occur, given that inventors would be denied property rights that would enable them to enjoin *all* valuable uses of their technologies, such powerful patents would ensure the presence of powerful ex ante incentives. In such a regime, Type II errors would surely swamp Type I. Even when debate erupts as to whether patents in a particular field are producing a “tragedy of the anticommons” by hindering innovation more than the grant of excessive rewards would spur it,⁸⁷ one adhering to this approach would strongly resist tempering patentees’ exclusive rights, for patentees would be entitled to derive as much profit as would realistically be possible through their IP rights. Injunctions would thus

85. Such context-specific determinations are not cost-prohibitive in all cases.

86. See Burk & Lemley, *supra* note 41 *passim*.

87. See, e.g., Giovanni Dosi et al., *Knowledge, Competition and Innovation: Is Strong IPR Protection Really Needed for More and Better Innovations?*, 13 MICH. TELECOMM. & TECH. L. REV. 471, 477, 480 (2007); Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCI. 698, 698–99 (1998).

be expected to routinely issue in the event of proven infringement, patentable subject matter would encompass all new and useful processes, and patentees would be free to enter into any contractual arrangements with licensees that they wish, subject only to antitrust constraints.

This approach may be fairly tied to the “property rights movement,” which seeks to import principles of property law into the intellectual realm and bestow inventors with unlimited rights to exclude.⁸⁸ Although such commentators believe that inventors of valuable technologies deserve to be richly rewarded, they nevertheless recognize that limits on patentees’ exclusive rights are warranted in certain circumstances.⁸⁹ Nor would they contest the principle of patents being time-limited, but they might however believe that patents in certain industries should be of greater duration.⁹⁰

A further alternative is the reverse of the maximalist approach, which is known as the abolitionist approach to patent law. From this perspective, the cost of the patent system so clearly exceeds the perceived benefits that the entire system should be withdrawn.⁹¹ Some people do advocate this extreme view.⁹² For them, the IP system’s restriction on individual freedom is so objectionable that patents should be eliminated.⁹³ They posit the mantra that “information just wants to be free” and believe that altruistic influences, social recognition, and the innate joy of inventing provide sufficient incentives to innovate.⁹⁴ For abolitionists, presumably reduced levels of capital investment in R&D are worthy of concern, but constitute the necessary price of a larger good in the form of individual liberty.⁹⁵

Despite the impassioned arguments some have leveled against the IP regime, it is difficult to take the abolitionist case seriously, at least insofar as it purports to apply to investment-heavy research sectors like the

88. See generally Peter S. Menell, *The Property Rights Movement’s Embrace of Intellectual Property: True Love or Doomed Relationship?*, 34 *ECOLOGY L.Q.* 713 (2007) (discussing the durability of the property rights movement’s embrace of intellectual property).

89. For instance, Professor Richard Epstein, who is regarded by some commentators as a leading proponent of the property rights movement, has acknowledged that limited circumstances exist in which patentees should be denied injunctive relief in the event of infringement. See, e.g., Richard Epstein, *The Property Rights Movement and Intellectual Property: A Response to Peter Menell*, 31 *REGULATION* 58, 59 (2008).

90. The pharmaceutical industry is the prime candidate in this regard.

91. See, e.g., MICHELE BOLDRIN & DAVID K. LEVINE, *AGAINST INTELLECTUAL MONOPOLY* (2008); N. STEPHAN KINSELLA, *AGAINST INTELLECTUAL PROPERTY* 19 (2008).

92. See Janis, *supra* note 71.

93. See *id.*

94. See Thomas O. Barnett, *Interoperability Between Antitrust and Intellectual Property*, 14 *GEO. MASON L. REV.* 859, 865 (2007).

95. For a discussion on the abolitionist perspective, see Janis, *supra* note 71.

pharmaceutical industry.⁹⁶ The pace and sheer scale of U.S. innovation remain the envy of the world.⁹⁷ The demonstrable success of the new U.S. economy—an economy founded in large part on a platform of strong IP rights—is testament to the power of patent law in inducing scientific progress. In light of this historic and ongoing success, and absent compelling empirical or theoretical evidence to the contrary, it would be massively irresponsible to jettison the system that has laid the foundation for such valuable technological innovation.⁹⁸

Although the current IP regime is of course not without its flaws, isolated failures in the patent system are nevertheless not inconsistent with that system performing an overall good. Put differently, the demonstrable net success of the patent laws does not establish that the contemporary system is optimally tailored. It could be that the patent system operates with great effectiveness in R&D-investment-heavy industries, which produce products based on a limited number of patent rights.⁹⁹ But it may also be that new areas of commercial innovation that give rise to innumerable blocking patents owned by numerous entities are less well-suited to traditional patent principles. Denying IP protection to entire swathes of such industries is worthy of skepticism, but, on the basis of current knowledge, it is still empirically unclear whether certain, limited instances of patent protection create net social losses. In lieu of a dogmatic rejection of the benefits of IP generally, the more responsible solution may lie in incremental adjustment.

96. A more reasonable variant of this account would advocate for the removal of patent rights only from certain suspect sectors, such as with business methods and computer software.

97. For a discussion on America's position as the world's most innovative economy in light of the 2008–2009 credit crisis, see *Innovation in America: A Gathering Storm?*, ECONOMIST, Nov. 22, 2008, at 84. *But see* David Goldman, *Recession's Latest Victim: U.S. Innovation*, CNNMONEY.COM, (Dec. 11, 2009), http://money.cnn.com/2009/12/11/news/economy/patent_filings/ (discussing the phenomenon of slowed U.S. innovation and patent filing in light of the 2009 recession).

98. This holds true even in contested industries, such as computer software, where high levels of innovation belie the existence of an IP system that impedes innovation to such a significant degree that it should be removed entirely. *See* Andrew Beckerman-Rodau, *The Supreme Court Engages in Judicial Activism in Interpreting the Patent Law in eBay, Inc. v. MercExchange, L.L.C.*, 10 TUL. J. TECH. & INTELL. PROP. 165, 172 (2007).

99. Of course, it may also be the case that even the pharmaceutical industry, dependent as it is on strong patent rights, may presently be garnering greater profits than necessary to spur innovation. *See Winds of Change: Merck's Woes Illuminate the Shifts Taking Place in the Drugs Industry*, ECONOMIST, Dec. 13, 2008, at 89 (referring to a study that suggested that a 20% drop in drug prices for Medicare purchases would only reduce leading companies' profits by 5% and inferring that no threat to the pharmaceutical industry would exist as long as it continues to innovate).

This brings us to the parsimonious path, which this Article finds the most compelling approach to patent law. Like the quasi-maximalist perspective on patent jurisprudence, it acknowledges that the optimal outcome would be to match incentives precisely to rewards, but similarly recognizes that pragmatic considerations foreclose such an outcome. It may instead be characterized by reference to displaying a qualified bias in favor of constrained compensation in close cases.

As defined by this Article, the parsimony principle instructs that policymakers and courts should not be entirely agnostic in their perceived abilities to judge the unique incentive characteristics of various industries, as the principle teaches that patentee compensation can and should differ depending on the relevant industrial setting. Where evidence exists that inventors in certain fields require less potent patent rights to induce them to invest in desirable levels of R&D, those rights should be diluted in an appropriate fashion. The problem, of course, is that the relevant evidence will rarely be irrefutable, as problems associated with new technologies, and hence new incentive structures, are likely to build incrementally and remain contestable for some time. But as industry concerns grow and become more vocal—as has been the case in the information-technology and business-method settings—the PTO and the courts should avail of patent law's policy levers incrementally to constrain net (as opposed to necessarily individual) compensation. Perhaps the best way to implement such policies is to adjust the nonobviousness and novelty bars to limit the proprietary scope of patented technologies in fields of concern. Such a parsimonious approach to patent policy would see subtle, evolving limitations on patentees' exclusive reach, which may be characterized by minimalist alterations to patent rights in reaction to perceived evidence. Ultimately, the principle would attempt to significantly constrain inventors' pecuniary return and, in so doing, minimize windfalls while still maintaining adequate incentives to invent. By operating in such a fashion, policymakers would naturally demonstrate more faith in their abilities to craft responsible rules of asymmetric effect than would those who adhere to a quasi-maximalist approach. This Article contends, however, that sufficient information is available to mold industry-specific patent rights that would better approximate the theoretical optimum level of protection.

Of course, the preceding categories are subjectively malleable and could conceivably overlap, particularly at the border of the quasi-maximalist and parsimonious approaches to patent policy. But these two categories can be distinguished on the basis of degree and innate bias in the event of qualified uncertainty. The parsimonious path may be defined in part by a greater proclivity for Type I errors than the quasi-maximalist approach would have.

It may also be further defined by an ongoing attempt to cabin rights in appropriate settings as further evidence becomes available.

Can we inject further specificity into the definition of the parsimony principle? As the boundaries of the principle become more clear-cut, it will lend itself increasingly well to more useful applications. In addition to the preceding exploration, it may be helpful to clarify that the parsimony principle is not an approach that seeks to err on the side of under-compensation in cases of uncertainty. The principle, defined in this manner, would be problematic for several reasons, with the most prominent problem being that the law does not demonstrate such a bias. Protection in patent law remains very generous, particularly in comparison to copyright law. Moreover, accepting a large proportion of Type I errors would be a dangerous policy, given the greater benefits of dynamic over static efficiency.¹⁰⁰ So defined, this parsimony principle would be normatively undesirable and would also provide a poor positive explanation of the current patent system. Instead, for those who believe that patent law, despite its many historical and self-evident successes, may have overstepped its bounds, the parsimonious path offers a responsible course by which to limit perceived instances of patentee overcompensation.

Of course, it is the indeterminate nature of innovation, and the elusive point of optimal protection, which denies us the benefit of perfect clarity. Were we confidently able to define an optimal principle that would cause patent doctrine to approach the optimal outcome, we would have solved the problem of indeterminism, but no heuristic devised on the basis of seriously incomplete information could reasonably aspire to such an end. The more modest ambition of this Article is to point to a broad theme of cabined exclusivity that produces lower levels of patentee compensation than would result from granting inventors the full panoply of benefits that could plausibly be derived from the statutory framework put in place by Congress in 1952. Instead of seeing a conscious bias in favor of significant overcompensation, we can observe a more nuanced system that attempts to impose meaningful, though responsible, limitations on patentees' exclusive rights. This degradation of patentees' rights takes place in an incremental fashion, as further information becomes available to policymakers. As the environment within which patent law operates becomes more information-rich, by virtue of the large-scale natural experiment of which the intellectual-property system is a part, we can expect the parsimony path to lead incrementally, yet inexorably, closer to the social optimum.

100. See Barnett, *supra* note 8, at 1194.

Before considering how the parsimony principle can be employed to inform modern debate, particularly on the role of patentable subject matter and calculating damages, this Article explains a somewhat surprising result: patent law has silently embraced the parsimony principle in its substantive doctrine.¹⁰¹ The law has subtly evolved in a manner that seeks to limit patentees' exclusive reach to a level that is unlikely to result in significant overcompensation.¹⁰² The following Part espouses a novel interpretation of the patent system. In doing so, it concludes that the parsimony principle yields great instructive power in explaining the current constitution of the law.

III. PATENT LAW'S SILENT EMBRACE OF THE PARSIMONY PRINCIPLE

A. PATENTABLE SUBJECT MATTER

It might seem odd to begin a discussion of patent law's subtle adoption of the parsimony principle by appealing to a topic that some might read to support an opposing view. Until very recently, the tale of patentable subject matter has been of near-unqualified expansion.¹⁰³ Historically, a host of restrictions on patent eligibility existed, but computer software, business methods, immoral inventions, gene sequences, and biological material have all been deemed patent-ineligible at one point or another.¹⁰⁴ Yet in a series of facilitative rulings, the Supreme Court and Federal Circuit opened up virtually all areas of innovative activity to patent protection, save those that are inescapably abstract or are discovered in lieu of being invented, with the latter restriction resulting from the perennial ban on patenting preexisting natural phenomena.¹⁰⁵ The high Court's famous remark that "anything under the sun that is made by man" is patent-eligible aptly summarizes the law's evolving reach.¹⁰⁶

101. *See infra* Part III.

102. *See infra* Part III.

103. *See* SCHECHTER & THOMAS, *supra* note 4, at 58–60.

104. *See* Gottschalk v. Benson, 409 U.S. 63, 71–72 (1972) (computer software); Hotel Sec. Checking Co. v. Lorraine Co., 160 F. 467, 469 (2d Cir. 1908) (business methods); Lowell v. Lewis, 15 F. Cas. 1018, 1019 (C.C.D. Mass. 1817) (immoral inventions); Nicholas M. Zovko, Comment, *Nanotechnology and the Experimental Use Defense to Patent Infringement*, 37 MCGEORGE L. REV. 129, 152 (2006) (biotechnology).

105. *See* Diamond v. Diehr, 450 U.S. 175, 177 (1981) (computer software); Diamond v. Chakrabarty, 447 U.S. 303, 305 (1980) (oil-eating bacteria); State St. Bank & Trust v. Signature Fin. Grp., 149 F.3d 1368, 1370 (Fed. Cir. 1998) (business methods); Utility Examination Guidelines, 66 Fed. Reg. 1092, 1093 (Jan. 5, 2001) (genes).

106. *Chakrabarty*, 447 U.S. at 309.

In light of the patent system's wide-reaching application, what does the question of patentable subject matter teach about the law's purported adoption of a parsimony principle? First, its expansive reach across fields of innovative activity is not at all inconsistent with a parsimonious approach since the parsimony principle would rather cabin the breadth and duration of exclusive rights to remedy possible overprotection than withdraw entirely from fields where such overprotection is arguably present. Innovation is an activity of huge societal import, which emanates from more than traditional contexts alone. Biotechnology, computer software, gene sequencing, medical diagnostic methods, and other fields of research bear the potential to yield vast contributions to social welfare. While theory suggests that a patent thicket can detract from scientific progress in certain of these fields,¹⁰⁷ there is no convincing evidence that indicates that innovation would continue apace in these fields in the absence of any property rights. This being the case, patent law's wide reach would generally seem to be commendable.¹⁰⁸

Second, gaps in patentable subject matter say important things about the perceived need for property rights to spur innovation.¹⁰⁹ It must be borne in mind that denying inventors in specified fields any prospect of patent protection is something of a nuclear option. A parsimonious approach would seek to curtail exclusive rights to prevent windfalls, but would generally not deny some right of ownership. Such outright denials can often, though not invariably, bear an excessive risk of eviscerating incentives to invent. Yet, as the Court has emphatically stressed, its statement that "anything under the sun that is made by man" is patentable "is not to suggest that § 101 has no limits or that it embraces every discovery. The law of nature, physical phenomena, and abstract ideas have been held not patentable."¹¹⁰ As it turns out, these foundational limitations on patent-eligible subject matter can only

107. See, e.g., BURK & LEMLEY, *supra* note 9, at 89–90 (detailing the possible patent-thicket problem in information-technology industries); Karl Bergman & Gregory Graff, *The Global Stem Cell Patent Landscape*, 25 NATURE BIOTECHNOLOGY 419, 422 (2007) ("The characteristics of the stem cell patent landscape are consistent with conditions that could give rise to a patent thicket."); Arti K. Rai & Rebecca S. Eisenberg, *Bayh-Dole Reform and the Progress of Biomedicine*, 66 LAW & CONTEMP. PROBS. 289, 297 (2003) ("Concern about an anti-commons . . . is quite pressing in contemporary biomedical research that draws upon many prior discoveries made by different scientists in universities and private firms.").

108. The question of whether computer software and business methods are appropriately subject to patent protection is one of the most divisive contemporary issues in patent law. It is also one that is fundamentally implicated by this Article's focus on the parsimony principle.

109. Alternatively, it may say something about society's interest in not promoting innovation in all possible fields of endeavor.

110. *Chakrabarty*, 447 U.S. at 309.

be satisfactorily explained by reference to patent law's parsimony principle. These fields of discovery bear unique potential for overcompensation, given their upstream nature and the concomitant proclivity for ubiquitous downstream application.

Yet these restrictions on patentable subject matter have never been explained in such terms by the judiciary. Instead, the courts have spoken in somewhat conclusory terms, focusing instead on the three atextual non-patentable categories identified above—abstract ideas, laws of nature, and natural phenomena. In 1852, the Court provided its definitive exposition that “a principle is not patentable. A principle, in the abstract, is a fundamental truth; an original cause; a motive; these cannot be patented.”¹¹¹ In 1944, the Second Circuit echoed this view, noting that “epoch-making ‘discoveries’ or ‘mere’ general scientific ‘laws,’ without more, cannot be patented.”¹¹² The exclusion of fundamental principles has largely been explained on the ground that they naturally exist in nature and therefore cannot be “invented.”¹¹³ The Supreme Court has concluded that such principles are “part of the storehouse of knowledge of all men” and therefore cannot be appropriated by any single entity.¹¹⁴

This account of why laws of nature are non-patentable is quite unconvincing.¹¹⁵ A storehouse of knowledge only exists when human toil and ingenuity reveal fundamental rules of physics that may have been extraordinary elusive for an extended period of time. The Court has reiterated time and again that Einstein's $e=mc^2$ could not have been patented.¹¹⁶ Note that this and other breakthrough scientific discoveries easily surmount the conditions of novelty, nonobviousness, and utility. Until the moment of discovery, such laws of nature have escaped human understanding for our entire history, despite enduring efforts to understand the universe within which we reside. One can hardly call the nonobviousness or novelty of such discoveries into question. Ground-breaking scientific discoveries of the kind that define the late nineteenth and twentieth centuries have given rise to unprecedented levels of technological innovation. Science continues to develop at an extraordinary rate and social welfare benefits

111. *Le Roy v. Tatham*, 55 U.S. 156, 175 (1852).

112. *Katz v. Horni Signal Mfg. Corp.*, 145 F.2d 961, 961 (2d Cir. 1944).

113. *See, e.g., Ex parte Lundgren*, No. 2003-2088, 2004 WL 3561262, at *14 (B.P.A.I. Sept. 28, 2005).

114. *Funk Bros. Seed Co. v. Kalo Inoculant Co.*, 333 U.S. 127, 130 (1948).

115. *See, e.g., Ariel Simon, Reimventing Discovery: Patent Law's Characterizations of and Interventions Upon Science*, 157 U. PA. L. REV. 2175, 2180 (2009).

116. *See, e.g., Diamond v. Chakrabarty*, 447 U.S. 303, 309 (1980).

accordingly.¹¹⁷ Much, if not all, of this progress can be tied to the discovery of fundamental principles. Therefore, no one can credibly challenge the utility of such discoveries.

If patent law is founded on the principle of incentivizing socially desirable innovation, then the preclusion of fundamental principles is a profound incongruity. The gains from their discovery can be extraordinary, and they can be confoundingly difficult to unearth. Given that vast rates of intellectual and pecuniary capital may be required to successfully discover rules of nature that bear great potential value for society, the utilitarian case for patent protection would appear to be strong. It is no answer to proclaim that such principles belong to all mankind. Presumably, the Court views abstract discoveries with consternation because of their broad field of use, unencumbered by limitations to specific applications, threatens to hinder all subsequent applications of those basic scientific discoveries. The transaction costs involved in licensing fundamental principles to myriad users may be large, but to reject patent protection on this basis is to do so solely on the ground that the excluded discoveries are unusually valuable. It is easy to forget that transaction costs are what translate into a monetary reward for the deserving inventor (or discoverer). It is thus not satisfactory to draw a distinction, as the Court has attempted to do, between “invention” and “discovery.”¹¹⁸ The U.S. Constitution explicitly envisions the patent eligibility of both.¹¹⁹

There is a strong normative justification for limitations on patent law’s reach, though it is not one that has been adopted by the Court. By denying patent protection to those who unearth fundamental principles, but granting patent eligibility to those who can apply those principles to articulated and limited ends, the patent system effectively cabins the potential for overprotection. Complete ownership of a fundamental principle will result in exclusive rights of exceptionally broad reach and, hence, extraordinarily high profitability. It is likely that the returns associated with such an expansive property right would result in a windfall for the discoverer. Here, the distinction between the quasi-maximalist and parsimonious approaches is evident. Bestowing a person who uncovers an abstract principle of great value with a property right over all of its conceivable applications would

117. See Harold T. Shapiro, *Stopping Science: Human Cloning—Should It Be Stopped?*, 9 HEALTH MATRIX 303, 307 (1999).

118. See *In re Alappat*, 33 F.3d 1526, 1582 (Fed. Cir. 1994) (Rader, J., concurring); see also Troy L. Gwartney, Note, *Harmonizing the Exclusionary Rights of Patents with Compulsory Licensing*, 50 WM. & MARY L. REV. 1395, 1398 n.5 (2009).

119. U.S. CONST. art. I, § 8, cl. 8.

result in the discoverer recouping a large degree of the principle's social value, but it would not allow her to derive greater value than is inherent in her discovery. However, the return she garnered would most likely exceed what would have been necessary to spur her research efforts. Adopting this perspective, patent law's parsimony principle limits inventors' exclusivity to a degree that is likely to provide a sufficient return, but is unlikely to result in wholly disproportionate profits.¹²⁰

In short, fundamental principles that bear the potential for ubiquitous application are perhaps uniquely poised to provide those who control them with profits far greater than necessary to induce their discovery, but allowing people to patent particular applications of those principles ensures that sufficient returns are possible for commercialization. The result—inventors enjoying a sphere of exclusivity over specific applications of principles rather than over all applications of such principles—is a balanced and sensible approach to managing incentives.

Thus, the law's traditional exclusion of fundamental principles from patentable subject matter suggests a parsimonious approach. A related feature of the patentable subject-matter doctrine concerns the Court's departure from the Patent Act's instruction that "any new and useful process" is patent-eligible.¹²¹ The judiciary has resisted giving force to this command, believing that some processes are too open-ended for patent protection.¹²² Here, the parsimonious approach is similarly evident, though the courts have struggled mightily to find a harmonious limiting principle. One need merely review the tortured reasoning of *Gottschalk, Flook, Diamond*, and most recently *Bilski* to appreciate how the Court has been unable to articulate a satisfactory curb on process patenting.¹²³

The judiciary's struggle with untethered processes is very much related to its rejection of fundamental principles as a class of patent-eligible subject matter. It is a basic aspect of patent law that a claim "cannot be construed so broadly to cover every conceivable way or means to perform" that

120. See *Lab. Corp. of Am. Holdings v. Metabolite Labs, Inc.*, 584 U.S. 124, 126–27 (2006) (Breyer, J., dissenting).

121. See, e.g., Dotan Oliar, *The (Constitutional) Convention of IP: A New Reading*, 57 UCLA L. REV. 421, 456 (2009) ("Congress could clarify that section 101, which says that 'whoever . . . discovers any new and useful . . . composition of matter . . . may obtain a patent,' means what a plain reading suggests—that is, that one may patent a newly discovered natural element, for example—rather than what the Supreme Court interpreted it to mean—namely, that the natural product doctrine bars such protection.").

122. See, e.g., *id.*

123. See John A. Burtis, Comment, *Towards a Rational Jurisprudence of Computer-Related Patentability in Light of In re Alappat*, 79 MINN. L. REV. 1129, 1137 (1995).

function.¹²⁴ Similarly, use of a means-plus-function limitation serves as an important curb against potential over-rewarding.¹²⁵ The theory that justifies these limitations is identical to that which counsels the preclusion of fundamental principles from patent protection: the means-plus-function limitation acts as an important curb on patentees' ability to gain exclusive rights over a broader range of conduct than is appropriate to their respective inventions. Were patentees entitled to claim their inventions through the means-plus-function method, without having those claims limited by the patent document's specification,¹²⁶ they would be able to preempt vast swathes of activity. Similarly, all-encompassing method claims would grant such patentees far greater rewards than would be warranted by the circumstances attendant on their R&D efforts. The law's exclusion of such claims constitutes a resounding endorsement of its embrace of parsimonious ideals.

A final word on the question of patentable subject matter relates to the nonobviousness condition.¹²⁷ Technically, this requirement is separate from the general question of what areas of innovation are eligible for patent protection, but the issues are sufficiently related to warrant discussing these two patentability hurdles together. Whether a parsimonious or quasi-maximalist bias should color determinations under § 103 is an important one. Nonobviousness, which has been aptly characterized as "the ultimate condition of patentability,"¹²⁸ serves a crucial role in siphoning undeserving inventions out of patent eligibility. Obvious increments in knowledge, even if novel and useful, are in all likelihood inevitable. Therefore, from a utilitarian

124. *Mas-Hamilton Grp. v. LaGard, Inc.*, 156 F.3d 1206, 1214 (Fed. Cir. 1998).

125. A related exception to the rule that a patent's reach is based on the substance of its claims lies in 35 U.S.C. § 112, ¶ 6, which allows patentees to engage in means-plus-function claiming, pursuant to which they outline the steps and consequences involved in a process without detailing the underlying structure in the claim itself. However, such otherwise untethered claims are construed in light of the specification, lest the claimed invention be overbroad and grant the owner an excessive return. *See Kemco Sales, Inc. v. Control Papers Co.*, 208 F.3d 1352, 1361–62 (Fed. Cir. 2000). But placing the controversial phenomenon of means-plus-function claiming aside, which tends to eviscerate the purpose of claiming and simply grants patentees what they invented instead of what they claim, claims are essentially self-contained. It is perhaps noteworthy that the incongruity between means-plus-function claiming and the traditional manner of detailing the boundaries of one's invention has caused some critics to call for the former's elimination. If this were put into effect, the U.S. patent system's procedure for establishing the legal reach of one's invention would be rendered considerably more harmonious.

126. *See Kemco Sales*, 208 F.3d at 1361–62.

127. 35 U.S.C. § 103 (2006).

128. *See* Robert P. Merges, *Commercial Success and Patent Standards: Economic Perspectives on Innovation*, 76 CALIF. L. REV. 803, 812 (1988) (quoting NONOBVIOUSNESS—THE ULTIMATE CONDITION OF PATENTABILITY: PAPERS COMPILED IN COMMEMORATION OF THE SILVER ANNIVERSARY OF 35 U.S.C. 103 (John F. Witherspoon ed., 1980)).

standpoint awarding patents for such inventions creates monopoly distortions with little countervailing benefit. Specifically, the nonobviousness condition plays a crucial role in limiting overcompensation.¹²⁹ The problem is that obviousness determinations are notoriously difficult at the margin and are vulnerable to hindsight bias,¹³⁰ as it may be that many elegant solutions to intractable problems seem straightforward after the fact. This phenomenon threatens to deprive deserving inventors of patent eligibility, which would result in under-compensation and an undesirable diminution in ex ante incentives.¹³¹

Judicial exposition of general principles for making § 103 determinations may be somewhat helpful, though they are inevitably limited in their efficacy because such determinations are necessarily context- and fact-specific. The more interesting question is whether the PTO and courts should make nonobviousness assessments with a proclivity for Type I or II errors. The answer would seem to depend on the industrial context in which the non-obviousness determination is being made. Given that some evidence exists that patents in certain areas of innovation—most notably business methods, computer software, and medical diagnostic techniques—may be over-rewarding inventors,¹³² § 103 demarcations must be made discriminately in those settings. Where the obviousness bar for patentability becomes too readily surmountable, a host of patents over inevitable inventions will cast a windfall upon their owners and choke downstream innovation through unwarranted search and negotiation costs, as well as cause potential hold-out problems.¹³³ The dubious quality of many business-method patents, remarked upon with some concern by Justice Kennedy in *eBay*, emanated from lax nonobviousness analyses.¹³⁴ The failure largely resulted from a combination of little prior art, given the lack of historical patentability of

129. *See id.* at 820–26.

130. *See* Gregory N. Mandel, *Patently Non-Obvious: Empirical Demonstration that the Hindsight Bias Renders Patent Decisions Irrational*, 67 OHIO ST. L.J. 1391 (2006).

131. It should be remembered that those incentives are already implicated by the uncertainty inherent in ex post determinations, like non-obvious assessments, which threaten to deprive inventors of rewards after investment in R&D has already taken place.

132. *See, e.g.*, sources cited *supra* note 107.

133. *See* Merges & Nelson, *supra* note 2, at 865–66 (explaining that the hold-up problem is apt to be more acute where the invention underlying the original, as opposed to the improvement, patent is less significant); Roin, *supra* note 76, at 532–36. Of course, the nonobviousness condition loses much of its normative legitimacy when applied to capital-intensive research endeavors, such as pharmaceuticals, where even obvious inventions may not be inevitable absent IP protection or an alternative reward structure. *See id.* at 532–36.

134. *See* *eBay, Inc. v. MercExchange, L.L.C.*, 547 U.S. 388, 396–97 (2006) (Kennedy, J., concurring).

business methods, and the lack of business expertise on the part of PTO examiners.¹³⁵ The Supreme Court's recent decision in *KSR*, which raised the hurdle for nonobviousness determinations, should apply with particular vigor to the fields of innovation within which the patent system's presence is being most criticized.¹³⁶ There, the parsimonious principle should find its strongest application.

In contrast, fields in which patents are the undisputable foundation of ongoing innovation, such as in the pharmaceutical industry, should receive more lax nonobviousness assessments. In such cases, the cost of development has invariably been enormous and the social gain to society from commercialization is similarly large. Denying patent protection on a technical ground in such a context is difficult to justify on normative grounds.¹³⁷ Here, the parsimonious approach should be applied in a more cautious manner.

It is telling indeed that the patent system has already begun to move in this direction.¹³⁸ The Federal Circuit has displayed a readiness to apply obviousness determinations with particular vigor in certain fields, including computer software.¹³⁹ Yet it has simultaneously alleviated inventors in other industries, notably biotechnology and pharmaceuticals, from complying with a stringent nonobviousness requirement.¹⁴⁰ This phenomenon, though hard to ground in the statute, is a strong signal of the parsimony principle's presence.

B. THE NATURE OF A PATENTEE'S PROPERTY RIGHT

The general conditions of patentability discussed above demonstrate a strong endorsement of the view that patent law has charted a parsimonious path. This Section considers an aspect of the law that explicitly implicates patentees' ability to earn rewards in excess of what was necessary to induce them to invent and commercialize their discoveries. Specifically, it addresses the nature of the property rights the patent system bestows on qualifying

135. See John R. Allison & Starling D. Hunter, *On the Feasibility of Improving Patent Quality One Technology at a Time: The Case of Business Method Patents*, 21 BERKELEY TECH. L.J. 729, 732–33 (2006); Julie E. Cohen, *Reverse Engineering and the Rise of the Electronic Vigilantism: Intellectual Property Implications of "Lock-Out" Technologies*, 68 S. CAL. L. REV. 1091, 1178 (1995).

136. *KSR Int'l Co. v. Teleflex, Inc.*, 550 U.S. 398 (2007).

137. See Benjamin Roin, *Unpatentable Drugs and the Standards of Patentability*, 87 TEX. L. REV. 503 (2009).

138. See generally Dan L. Burk & Mark A. Lemley, *Is Patent Law Technology-Specific?*, 17 BERKELEY TECH. L.J. 1155 (2002).

139. See *id.* at 1167.

140. See Burk & Lemley, *supra* note 41, at 1677.

inventors. In doing so, it explores the extent to which the characteristics of that right translate into optimal rewards.

The property rights created by the patent system are unique in an important respect. The Seventh Circuit defined property as “a mixture of rights to use and to exclude others from using.”¹⁴¹ But patents grant only the latter power—they do not convey on their owners an affirmative right to practice their inventions.¹⁴² This feature of the patent system, which gives rise to the phenomenon of “blocking patents,”¹⁴³ has both attractive and potentially troubling qualities.¹⁴⁴ The absence of an unqualified ability to use one’s invention enhances the return of prior inventors, whose patented technology implicates later innovation. By granting those earlier patentees a right to enjoin follow-on invention, the law bolsters the value of their IP.¹⁴⁵ This might appear to be an exception to the parsimonious approach, which is generally reflected in patent law and perhaps would—if applied in strong form in this instance—grant improvers a right to practice their new and useful inventions without having to negotiate for permission.

But, of course, the law may be wise not to bestow improvers with an affirmative right to practice their inventions. The parsimonious approach only works if patent law operates to cabin patentees’ exclusive rights, *without* reducing them below the necessary level to induce innovation.¹⁴⁶ Were patents to reflect traditional principles of property law and allow their owners to practice their inventions—in the same way that a property owner generally need not bargain for permission to use her own land as she sees fit—there would be a serious risk of under-compensation in at least some industries. This is most likely to be the case in certain high-tech industries where cumulative innovation is rapid and where, as a corollary, any one increment

141. *Chi. Prof'l Sports Ltd. P'ship v. Nat'l Basketball Ass'n*, 961 F.2d 667, 670 (7th Cir. 1992).

142. See Sina Muscati, *Terminator Technology: Protection of Patents or a Threat to the Patent System?*, 45 IDEA 477, 507–08 (2005).

143. Blocking patents arise when incremental improvements over a patented product or process, not sufficiently extraordinary to trigger the application of the reverse doctrine of equivalents, are themselves patented. In such cases, the improver cannot practice the relevant technology without permission from the original patentee. See Timothy R. Holbrook, *Equivalency and Patent Law's Possession Paradox*, 23 HARV. J.L. & TECH. 1, 12 (2009).

144. See Robert Merges, *Intellectual Property Rights and Bargaining Breakdown: The Case of Blocking Patents*, 62 TENN. L. REV. 75 (1994); see also Maureen A. O'Rourke, *Toward a Doctrine of Fair Use in Patent Law*, 100 COLUM. L. REV. 1177, 1204 (2000).

145. See Robert Merges, *A Brief Note on Blocking Patents and Reverse Equivalents: Biotechnology as an Example*, 73 J. PAT. & TRADEMARK OFF. SOC'Y 878, 878–79 (1991).

146. As emphasized above, the parsimony principle does not counsel a bias in favor of under-compensation.

in technology will likely be defunct in a short amount of time. The problem would be most serious in situations where a single instance of major innovation precedes a series of rapid, but modest, advancements. Those later improvements would receive patent protection¹⁴⁷ and, due to their superiority, would deprive the inventor who made the underlying breakthrough of a reward. This scenario is a real danger, given the commonly accepted fact that innovation tends to be “lumpy,” rather than smooth and continuous.¹⁴⁸ Sudden technological advancements often follow on the heels of a paradigm-altering discovery.¹⁴⁹ Society needs to incentivize such breakthrough inventions more than to reward those ensuing minor improvements that build on a rapid advance in the art.¹⁵⁰ Thus, the lack of an absolute right to use a patented product or process can be thought of as a means to ensure optimal innovation that would otherwise be compromised by free-riding.

The absence of a right to practice one’s invention, however, may be problematic in some circumstances. In particular, such an absence threatens to over-reward inventors of technologies that represent merely a modest improvement over the prior art, but which nevertheless capture commercialized products. Where a patentee can stake a credible claim to a piece of a marketable product, it can credibly threaten to enjoin sales.¹⁵¹ This can hold true even if the allegedly infringed patent implicates only a peripheral part of the relevant invention.¹⁵² Because inventors of even

147. 35 U.S.C. § 101 (2006).

148. See Herbert Hovenkamp, *Restraints on Innovation*, 29 CARDOZO L. REV. 247, 257 (2007).

149. In this respect, it is well known that innovation tends not to occur in a continuous, linear fashion, but instead occurs in waves following significant technological breakthroughs. See, e.g., J. Gregory Sidak & David J. Teece, *Dynamic Competition in Antitrust Law*, 5 J. COMPETITION L. & ECON. 581, 604–05 (2009).

150. See Christopher A. Cotropia, *Patent Law Viewed Through an Evidentiary Lens: The “Suggestion Test” as a Rule of Evidence*, 2006 BYU L. REV. 1517, 1525.

151. Such a patentee can also threaten to extract significant damages in the event of ongoing infringement. See Mark A. Lemley, *Ignoring Patents*, 2008 MICH. ST. L. REV. 19, 19–20.

152. The threat to marketers of technology is so significant, and the search and negotiation costs so immense, that patent pools now abound. See, e.g., R. Justin Koscher, *A Patent Pool’s White Knight: Individual Licensing Agreements and the Procompetitive Presumption*, 20 DEPAUL J. ART, TECH. & INTELL. PROP. L. 53 *passim* (2009) (discussing the role of patent pools in markets subject to potential patent-thicket problems). These entities seek to group all necessary patents to commercializing a technology and to license them at agreed rates. But these entities do not always incorporate every patent. Sometimes, this is deliberate, as patentees either elect not to join or strategically withhold notice of their rights until such time as an infringing product is taken to market. In both such situations, patentees can extract a far greater reward than licensees would have been willing to pay *ex ante*.

modest and subsequently improved-upon technologies can exclude commercialization efforts by improvers, the formers' property rights can result in overcompensation.

The above illustrates how asymmetric incentives to innovate complicate the construction of optimal patent rules. The nature of the property right awarded to deserving inventors lies at the very heart of the patent system. Congress, when enacting the initial patent regime, elected to eschew one traditional aspect of property: the right to use what one owns. The concept of property in patent law is therefore distinct, but whether that distinction is desirable depends significantly on context. Given the rapid, cumulative innovation in many industries subject to patent protection, such as biotechnology, computer software, and IT,¹⁵³ initial inventors of foundation technologies may be left seriously under-compensated by subsequent minor improvements that eliminate consumer demand for the original technological feat. There, patent law's right to exclude, but not to use, is entirely consistent with a parsimony principle. Yet, other scenarios may exist in which the patent system's refusal to grant affirmative rights to practice may frustrate follow-on innovation by creating myriad blocking positions that create immense transaction costs and, if negotiations are sequential in nature, give rise to hold-outs.

The patent system cannot undo these harms without also losing or offsetting crucial advantages. There may be a way to temper the nature of patentees' right to exclude, however, in situations where the lack of an affirmative right to practice one's invention threatens to ensnare cumulative innovation. The answer lies in the legal force given patentees' exclusive rights.

Literally, and by reference to traditional property, a patentee's "right to exclude" would seem to reflect a right to injunctive relief, since exclusion, by definition, refers to an act of shutting someone out. When a trespasser walks into a person's home, and when a company sells a product using another entity's patented technology, the exclusive right in both cases should presumably be the same. The owner can expel the unwanted person from her home, while the inventor can force the company to cease selling products that infringe her patent. The mechanism by which the law facilitates this outcome is the injunction. An owner's right to such a remedy in cases of

153. See Peter Lee, *The Evolution of Intellectual Infrastructure*, 83 WASH. L. REV. 39, 93 (2008).

trespass is virtually unqualified.¹⁵⁴ One would perhaps expect that a patent-owner's entitlement to similar relief should be similarly unfettered.

A patentee's ability to obtain injunctive relief is generally important, for only that legal right entitles him to negotiate a favorable license that rewards him in a manner that definitively exceeds his reservation reward.¹⁵⁵ Absent such ability, a patentee will face a serious threat of under-compensation, at least when there is no a priori basis for believing that courts accurately compute damages or assess attorneys' fees in the patentee's favor.¹⁵⁶ Given that the judiciary often lacks reliable means by which to observe patentees' reservation licensing rates, it may indeed be the case that damages run a significant risk of under-compensating inventors.¹⁵⁷ This is most likely to be the case where there is no licensing history for the assessing court to observe.¹⁵⁸ Consistent with the importance of injunctive relief to ensuring adequate compensation, it was long the rule that courts would apply injunctive relief in the event of proven infringement.¹⁵⁹

Yet the ready availability of injunctive relief is by no means necessary to ensure adequate compensation in all cases. Indeed, as explored *supra*, in certain situations a powerful right to exclude may result in a windfall return to inventors whose contributions were paltry. The latter is most aptly represented by the phenomenon of "patent trolls," which in their purest form provide no social value of any kind.¹⁶⁰ Instead of inventing new and

154. Inadvertent encroachment is an exception in some cases. See Olivia Leigh Weeks, Comment, *Much Ado About Mighty Little—North Carolina and the Application of the Relative Hardship Doctrine to Encroachments of Permanent Structures on the Property of Another*, 12 CAMPBELL L. REV. 71 (1989).

155. See, e.g., Ian Ayres & Kristin Madison, *Threatening Inefficient Performance of Injunctions and Contracts*, 148 U. PA. L. REV. 45, 56 (1999).

156. See Tim Carlton, Note, *The Ongoing Royalty: What Remedy Should a Patent-Holder Receive When a Permanent Injunction Is Denied?*, 43 GA. L. REV. 543, 548 (2009) (arguing that the award of reasonable royalty rates by courts under-compensates patentees).

157. See RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* 69 (5th ed. 1998); F. Scott Kieff, *Coordination, Property, and Intellectual Property: An Unconventional Approach to Anticompetitive Effects and Downstream Access*, 56 EMORY L.J. 327, 346–48 (2006); Stephanie Pall, Note, *Willful Patent Infringement: Theoretically Sound? A Proposal to Restore Willful Infringement to its Proper Place Within Patent Law*, 2006 U. ILL. L. REV. 659, 666 (observing that "the original basis for treble damages was not a desire to punish an infringer for increased culpability, but rather a desire to prevent jury under-compensation to the patentee").

158. See Leslie T. Grab, *Equitable Concerns of eBay v. MercExchange: Did the Supreme Court Successfully Balance Patent Protection Against Patent Trolls?*, 8 N.C. J.L. & TECH. 81, 104 (2006).

159. See *MercExchange, L.L.C. v. eBay, Inc.*, 401 F.3d 1323, 1338 (Fed. Cir. 2005), *rev'd*, 547 U.S. 388 (2006).

160. See Robin M. Davis, Note, *Failed Attempts to Dwarf the Patent Trolls: Permanent Injunctions in Patent Infringement Cases Under the Proposed Patent Reform Act of 2005 and eBay v. MercExchange*, 17 CORNELL J.L. & PUB. POL'Y 431, 438–39 (2008).

useful technologies, these non-practicing entities purchase copious amounts of patents (sometimes from bankrupt companies), do not seek to commercialize them, and then sue for infringement when oblivious companies market infringing goods.¹⁶¹ In such situations, there may be good ground for qualifying patentees' general right to injunctive relief.

In perhaps the most convincing indication that the parsimony principle influences patent law, the Supreme Court in 2006 did just that. In the landmark decision *eBay*—a case involving a non-practicing entity—the Court refused to bestow patentees with a presumptive right to injunctive relief in the event of established infringement.¹⁶² Instead, in order to prevail, the party seeking an injunction would have to satisfy the traditional factors (irreparable injury, inadequacy of remedies at law, balance of hardships favoring party seeking injunction, and public interest) that inform courts' decisions to grant equitable relief.¹⁶³ Despite commentators vilifying the decision for dealing a fatal blow to inventors' incentives to invent by jettisoning a property rule,¹⁶⁴ the case is better viewed through the lens of patent law's parsimony principle. In cases where infringement took place in full notice of the patentee's rights, where the patent constitutes a major, rather than a peripheral, aspect of the infringing product, where the patentee was actively marketing its technology, or in other compelling cases, injunctive relief is near-certain to be forthcoming. But in those cases where such relief is apt to over-reward patentees, such as where they are both non-practicing and non-licensing, where strategic hold-out is demonstrably present, or where the patent thicket is dense and an innocent infringer, despite its demonstrable best efforts, courts are now likely to be more cautious before awarding injunctive relief.¹⁶⁵ This recent development in the law constitutes a resounding endorsement of a parsimonious approach to patent policy.

C. MONOPOLY LEVERAGE AND RELATED LIMITS ON PATENT VALUE

A patent's worth lies not only in the technological merit of the invention it claims, but also in the legal means by which a patentee can realize the value of the invention. There is a fundamental distinction between patents and

161. See Opderbeck, *supra* note 67, at 130 n.12.

162. *eBay, Inc. v. MercExchange, L.L.C.*, 547 U.S. 388 (2006).

163. *Id.* at 391.

164. See Richard A. Epstein, *Richard A. Epstein: Court Soft on Property Rights*, FIN. TIMES (May 16, 2006 6:14 PM), <http://www.ft.com/cms/s/2/6a9ade66-e4ef-11da-80de-0000779e2340.html> (last visited Nov. 30, 2010); Brief of Various Law & Economics Professors as Amici Curiae in Support of Respondent, *eBay, Inc. v. MercExchange, L.L.C.*, 547 U.S. 388 (2006) (No. 05-130), 2006 WL 639164.

165. See, e.g., Venkatesan, *supra* note 83.

other forms of ownership rights—the former are uniquely insecure. Courts invalidate patents at high rates, in approximately 50% of cases according to some empirical studies.¹⁶⁶ This afflicts the incentive structure weighing on companies' decisions to invest capital in R&D. Probabilistic rights to exclude reduce the expected return from innovation, complicate the commercialization process by undermining post-invention investment, and have an especially negative effect on risk-averse inventors.

Observing the reality that many contemporary patents are little more than lottery tickets with better odds, one might correctly conclude the “bad patent” problem requires remedial action.¹⁶⁷ A crucial issue concerns the source from which enhanced patent certainty should appropriately emanate. If it results from heightened prosecution standards, which reduce the incidence of erroneous grant rates, then more secure patent rights are an unquestionable boon for society. While efforts are being made to achieve higher levels of performance at the PTO, ultimate success remains elusive. Nevertheless, increased patent certainty need not emanate from the PTO alone. Private contract between patentees and certain third parties can also increase patent certainty. The question whether patentees should be able to appropriate additional benefits for themselves is controversial.

In an unconstrained legal environment, patentees may be able significantly to enhance the certainty of their property rights.¹⁶⁸ Patents could purchase such security in a number of ways. First, many patentees derive much of their income from licensing fees.¹⁶⁹ By negotiating a suitable contractual provision, they could prevent licensees from challenging the validity or reach of their licensed patents. Similarly, patentees could negotiate a provision in such contracts that any subsequent judicial determination to the licensed patent's invalidity or non-infringement would not affect the royalties due. Third, patent holders could pay potential infringers who might challenge their patents in court not to do so. Fourth, patentees could regulate risk by structuring royalty payments over different timelines, including after

166. John R. Allison & Mark A. Lemley, *Empirical Evidence on the Validity of Litigated Patents*, 26 AIPLA Q.J. 185, 205 (1998) (46%); Jean O. Lanjouw & Mark Schankerman, *Protecting Intellectual Property Rights: Are Small Firms Handicapped?*, 47 J.L. & ECON. 45, 59 (2004) (finding trial win rates of “close to 50 percent”); Kimberly A. Moore, *Judges, Juries, and Patent Cases—An Empirical Peek Inside the Black Box*, 99 MICH. L. REV. 365, 385 (2000) (58%).

167. The bad-patent problem refers to the perceived tendency of the PTO to award many patents that would not survive serious scrutiny in court or upon reexamination. See Mark A. Lemley, *Is the Patent Office a Rubber Stamp?*, 58 EMORY L.J. 181, 181–82 (2008).

168. See Alan Devlin, *The Stochastic Relationship Between Patents and Antitrust*, 5 J. COMPETITION L. & ECON. 75, 77 (2009).

169. See Ian Ayres & Gideon Parchomovsky, *Tradable Patent Rights*, 60 STAN. L. REV. 863, 875 (2007) (observing that patentees' return on innovation depends on licensing fees).

expiration. All these concessions would not be free; rather, a patentee would have to give up something of value to induce another to bear the cost of potential invalidity. The amount paid, assuming both parties can reach agreement, will depend on the parties' relative appetite for risk.¹⁷⁰ If a patentee wants to avoid the dire consequences of courts' finding his patent invalid, he can protect himself by paying a licensee or potential infringer a premium to take on that risk instead. Finally, in addition to boosting the reliability of their patent rights through contract, patentees could also enhance their profits through product tying.¹⁷¹

Patents protected by such arrangements would be more valuable than they would be in isolation. Contractual agreements will enhance patentee compensation, without elevating it to the full social value of the invention (for part of that value would be spent on compensating licensees and potential infringers for not challenging the relevant patents). With these contractual agreements, a quasi-maximalist regime would grant patentees considerable freedom of contract. Instead of encountering this reality, however, one instead finds a host of restrictions on contractual freedom, all of which are designed to limit patents' value. Magnifying the effect of these limitations is the patent-misuse doctrine.¹⁷² Once more, the parsimonious approach is evident.

First, courts have abandoned the doctrine of licensee estoppel, which prevented patent licensees from bringing suit to challenge the validity of the patents upon which they were contractually bound to pay royalties.¹⁷³ Were the doctrine still in force, subject of course to contractual agreement, licensees would assume the risk of invalidity. Obviously the licensing patentee must compensate its licensees for taking on such risk. But, if it is willing to pay the requisite price, this conclusively establishes that the patentee derives greater utility from reduced, though secure, royalties than

170. Contracts operate as paradigmatic risk-shifting devices.

171. For the author's discussion of the economics of product tying (which involves conditioning the sale of one product on the purchase of another), see Alan Devlin, *A Neo-Chicago Perspective on the Law of Product Tying*, 44 AM. BUS. L.J. 521 (2007). It should be noted that none of the preceding courses of action necessarily results in patentees deriving greater value than is inherent in their respective technologies. After all, the technological value of any particular patent is set. A patentee cannot realize a greater return than that which is inherent in this ceiling, unless he uses his IP to mask a cartel agreement with owners of rival technology. Assuming that there is no substitute for the patented invention, no agreement with a licensee or potential infringer will affect the innate value of that technology. Instead, the various agreed-to measures merely operate to shift risk according to the parties' relative appetite for it.

172. See *Mercoid Corp. v. Mid-Continent Inv. Co.*, 320 U.S. 661, 666–68 (1944).

173. *Lear v. Adkins*, 395 U.S. 653, 671 (1969).

greater royalties whose recurrence depend on the validity of the underlying patent. In the famous case of *Lear v. Adkins*, the Supreme Court eliminated the doctrine, thus curtailing patentee compensation (which is tied to utility, rather than just pecuniary income).¹⁷⁴

The loss of licensee estoppel limits patentees' freedom to purchase security, but de facto limitations remained on licensees' ability to avail of the rule in *Lear*. For until very recently, licensees generally lacked standing to bring a suit for declaratory judgment of the relevant patent's invalidity.¹⁷⁵ Such entities would have to manufacture a case or controversy by deliberately withholding royalty payments, thereby committing breach of contract. Many licensees were understandably hesitant to take the chance of losing a validity challenge in court and then being liable under contract.¹⁷⁶ This changed with the Court's recent decision in *MedImmune, Inc. v. Genentech, Inc.*,¹⁷⁷ which seriously undermined patentees' ability to derive reliable royalties.¹⁷⁸ In that decision, the Court paved the way for licensees to challenge the validity of their licensed patents without having to withhold royalties.¹⁷⁹

Further restrictions on patentees' ability to derive heightened levels of utility from their inventions exist. "Reverse payments" constitute a particularly potent example of inventors favorably regulating the security of their patents.¹⁸⁰ Pursuant to this phenomenon, pioneer drug manufacturers pay generic entrants not to enter their markets until the expiration of their patents.¹⁸¹ While the Federal Circuit found these arrangements generally to be within the exclusive power of a patent grant, opposition from the antitrust agencies is mounting.¹⁸² It seems likely that Congress will amend the Patent

174. *Id.*

175. *See* *MedImmune, Inc. v. Genentech, Inc.*, 549 U.S. 118, 126–37 (2007) (reversing the Federal Circuit's determination that a patent licensee lacked standing to bring suit unless it first breached its licensing contract).

176. *See* Majoras, *supra* note 49, at 503.

177. 549 U.S. 118 (2007).

178. *See also* Rochelle Cooper Dreyfuss & Lawrence S. Pope, *Detroning Lear? Incentives to Innovate After MedImmune*, 24 BERKELEY TECH. L.J. 971 (2009).

179. *MedImmune*, 549 U.S. at 126–37.

180. *See* C. Scott Hemphill, *Paying for Delay: Pharmaceutical Patent Settlement as a Regulatory Design System*, 81 N.Y.U. L. REV. 1553 (2006) (discussing the phenomenon of reverse payments).

181. Preserve Access to Affordable Generics Act, S. 369, 111th Cong. (2009).

182. *Compare In re Ciprofloxacin Hydrochloride Antitrust Litig.*, 544 F.3d 1323 (Fed. Cir. 2008), *with* Jon Leibowitz, Chairman, Fed. Trade Comm'n, Speech Before the Center for American Progress: Pay-for-Delay Settlements in the Pharmaceutical Industry: How Congress Can Stop Anticompetitive Conduct, Protect Consumers' Wallets, and Help Pay for Health Care Reform (June 23, 2009) (transcript available at <http://www.ftc.gov/speeches/leibowitz/090623payfordelayspeech.pdf>).

Act to outlaw the practice.¹⁸³ Again, such a limitation will necessarily limit inventors' expected return from obtaining a patent, which dilutes the incentive to invent. A maximalist or quasi-maximalist approach to patent doctrine would allow reverse exclusionary payments, which enable patentees to garner compensation that maximizes ex ante incentives. After all, a right to pay a third party not to infringe can be reconciled with the "right to exclude" envisioned by the Patent Act. A quasi-maximalist regime, which would fall heavily on the side of avoiding Type I errors, would give full force to a patentee's exclusive rights. The fact that the law seems likely to outlaw reverse payments sends yet another strong signal that the parsimony principle is guiding the development of patent law.

Finally, the misuse doctrine purports to prohibit patentees from extending their lawful monopolies beyond their inherent intellectual property confines.¹⁸⁴ Obviously, the doctrine is designed to cabin patentee compensation.¹⁸⁵ Courts often apply the misuse doctrine to condemn product tying, where a seller insists that a consumer purchase a non-patented product as a condition of obtaining the patented one.¹⁸⁶ However, this account of bundling and requirements contracts as nefarious devices for leveraging monopoly power is unjustifiable. Product tying rarely operates to extend monopoly power from a tying to a tied market. Rather, its primary function is one of price discrimination.¹⁸⁷ The welfare effects of such a practice are ambiguous under current economic theory, though it is certainly the case that price discrimination enhances a seller's profits.¹⁸⁸ This being the case, the patent misuse doctrine's prohibition of tying serves to limit patentees' rewards, even though the patentees' means for maximizing their

183. See S. 369.

184. See *Mercoird Corp. v. Mid-Continent Inv. Co.*, 320 U.S. 661, 668 (1944).

185. See *id.*

186. The patent misuse doctrine is also employed to prohibit any attempt by patentees to garner royalties beyond the expiration of their patents. However, a patent's value is fixed, so a patentee cannot demand more than that amount in licensing fees. If a patentee insists that a licensee pay post-expiration royalties, it will either have to forego royalties in the early life of the license or receive a smaller royalty per-payment. Neither results in the patentee receiving more value than is inherent in her IP. See *Scheiber v. Dolby Labs., Inc.*, 293 F.3d 1014, 1020 (7th Cir. 2002).

187. See Richard A. Posner, *The Chicago School of Antitrust Analysis*, 127 U. PA. L. REV. 925, 926 (1979).

188. See Daniel A. Crane, *Intellectual Liability*, 88 TEX. L. REV. 253, 284–85 (2009). For the most recent account of the economics of product tying, see Einer Elhauge, *Tying, Bundled Discounts, and the Death of the Single Monopoly Profit Theory*, 123 HARV. L. REV. 397 (2009). For the author's exploration of this area, see Devlin, *supra* note 171, at 523.

return cannot reliably be shown to impose social costs.¹⁸⁹ Again, the parsimony principle would appear to be present.¹⁹⁰

D. DOCTRINE OF EQUIVALENTS

Aptly characterized as the most controversial canon in patent law, the doctrine of equivalents is notable for its inconsistency with the larger patent framework.¹⁹¹ It is a foundational principle of patent law that a patentee is entitled not to what she actually invented, but to the precise contours of what she actually claimed.¹⁹² Courts construe claims with reference to the specification contained in the patent document. Where the claims are clear, however, courts determine the reach of a patent document in a *Markman* hearing, primarily by reference to those claims alone.¹⁹³

The doctrine of equivalents is contentious because it acts as an exception to the rule that a patent's reach is coterminous with its claim limits.¹⁹⁴ According to the courts, “[t]he scope of a patent is not limited to its literal

189. See John P. Conley & Christopher S. Yoo, *Nonrivalry and Price Discrimination in Copyright Economics*, 157 U. PA. L. REV. 1801, 1810 (2009) (observing that banning price discrimination can reduce the production of desirable public goods).

190. As explored above, the parsimony principle does not incorporate a bias in favor of under-rewarding inventors. Rather, it seeks to impose meaningful limitations on patentees' abilities to reap windfall profits, while simultaneously displaying sensitivity to the dangers of depriving inventors of a sufficient return. On this basis, 35 U.S.C. § 271(d)(4)'s instruction that a patentee's refusal to license cannot constitute patent misuse makes eminent sense. As explained *infra* Section IV.B, patentees' profits should generally be cabined by restrictions on patent scope and duration, rather than by attempts to eviscerate their right to exclude. Protecting patent rights through liability rules is generally apt to cause greater harm than good.

191. See Michael J. Meurer & Craig Allen Nard, *Invention, Refinement, and Patent Claim Scope: A New Perspective on the Doctrine of Equivalents*, 93 GEO. L.J. 1947, 1948–49 (2005).

192. Although the system of claiming one's invention might strike the casual observer as somewhat odd, it is in fact a desirable mechanism for demarcating the boundaries of one's exclusive rights. Potential infringers need merely consult the claims to adjudge the propriety of their future courses of action. In short, the tenet of the law that an inventor establishes the boundaries of his invention through his or her patent claims injects some modicum of certainty into a system that is in dire need of it.

193. See *Silicon Graphics, Inc. v. ATI Techs., Inc.*, 607 F.3d 784, 792 (Fed. Cir. 2010) (“A construing court's reliance on the specification must not go so far as to import limitations into claims from examples or embodiments appearing only in a patent's written description . . . unless the specification makes clear that the patentee . . . intends for the claims and the embodiments in the specification to be strictly coextensive.” (quoting *JVW Enters. v. Interact Accessories, Inc.*, 424 F.3d 1324, 1335 (Fed.Cir.2005) (internal quotation marks omitted))).

194. Even in the case of means-plus-function claiming, the “metes and bounds” of the invention arguably emanates from the language of the claims, even if it does so using those claims as a conduit to the specification. See, e.g., *Messerschmidt v. United States*, 29 Fed. Cl. 1, 57 (1993).

terms but instead embraces all equivalents to the claims described.”¹⁹⁵ The doctrine of equivalents operates to magnify a patent’s reach beyond the literal scope of the claims by capturing activity that is substantially equivalent to the patented product or process. Under its most common iteration, the test asks whether the accused product performs substantially the same function in substantially the same way to achieve the same result (the “function-way-result” test).¹⁹⁶

The existence of a generous doctrine of equivalents would pose a significant problem to this Article’s contention that the patent system evolved along a parsimonious path. Yet, the absence of such a doctrine would similarly be in tension with the parsimony principle. A patent’s inability to capture a substantially identical product or process purely on account of linguistic limitations would enable widespread circumvention of patent documents. Importantly, such circumvention would be distinguishable from an “invent around.”¹⁹⁷ An invent around potentially leads to improvement and novel technologies, because it does not carry inherent social gain beyond eliminating monopoly rents.¹⁹⁸ While that effect may seem desirable, and for consumers in the short run surely it would be, the easy bypass of patentees’ reach would result in dangerous levels of under-compensation.

The reach of the doctrine is therefore all-important. An overly expansive doctrine would allow inventors to enjoin whole swathes of activity that, while related to the patented technology, is not coterminous with it. This reach would of course yield a windfall for the patentee, with all the negative results explored in Part II, *supra*. However, eliminating the doctrine altogether would not be prudent. The optimal reach of the doctrine is elusive as a practical matter, but the parsimonious approach would aim to impose meaningful limits on compensation.

In *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, in 2000, the Federal Circuit expanded the doctrine of prosecution history estoppel to prohibit any application of the doctrine of equivalents to an element in a patent claim that

195. *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 535 U.S. 722, 732 (2002).

196. *Graver Tank & Mfg. Co. v. Linde Air Prods. Co.*, 339 U.S. 605, 608 (1950).

197. *See, e.g., WMS Gaming, Inc. v. Int’l Game Tech.*, 184 F.3d 1339, 1355 (Fed. Cir. 1999) (noting that “the patent law encourages competitors to design or invent around existing patents”); *State Indus., Inc. v. A.O. Smith Corp.*, 751 F.2d 1226, 1236 (Fed. Cir. 1985) (“One of the benefits of a patent system is its so-called ‘negative incentive’ to ‘design around’ a competitor’s products, even when they are patented, thus bringing a steady flow of innovations to the marketplace.”).

198. *See State Indus.*, 751 F.2d at 1236.

was previously narrowed in the prosecution process.¹⁹⁹ This was a draconian ruling that ignited a firestorm of protest amongst the patent bar.²⁰⁰ The Supreme Court subsequently vacated the decision; the Court adopted a rebuttable presumption that narrowing a claim before the PTO triggers an estoppel.²⁰¹ But *Festo's* legacy significantly limited the doctrine of equivalent's ability to enhance patentee compensation. According to recent scholarship, the doctrine experienced a largely unnoticed demise.²⁰² This phenomenon surely reflects the parsimonious evolution of patent law.

E. EXPERIMENTAL USE—AN UNJUSTIFIED EXCEPTION?

Patent law's silent embrace of the parsimony principle would appear to be vulnerable to a notable exception. As explored in the preceding pages, patent doctrine supports patentees' right to obtain a lucrative return on valuable inventions, but nevertheless imposes limits on these returns such that inventors cannot obtain rewards equal to the social value of their technological discoveries. These limitations are far from edentulous. Restrictions on the kind of patented innovation, the increasing denial of injunctive relief in the event of infringement, wide-ranging limitations on contractual freedom, and reduced patent scope through diminished use of the doctrine of equivalents—in addition to claiming limitations, combine to deny patentees much of the social gain that accompanies their inventions.²⁰³ We would not associate such constraints with a maximalist or quasi-maximalist regime. The law governing third-party experimental use of a patented invention constitutes a potentially glaring exception.

An experimental use defense against claims of infringement would appear to serve an important function in constraining patentees' exclusive rights. Allowing rivals to experiment freely on a patentee's technology may facilitate invent-around diluting the patent-holder's market power.²⁰⁴ Moreover, a patent regime can foster cumulative innovation by allowing interested entities to unmask the inner workings of patented technologies through experimentation. Yet, the patent system adopted a decidedly non-parsimonious approach to the issue of experimental use. With highly restricted exceptions, there is no right to practice another's patented

199. 234 F.3d 558 (Fed. Cir. 2000) (en banc), *vacated*, 535 U.S. 722 (2002).

200. See John R. Allison & Mark A. Lemley, *The (Unnoticed) Demise of the Doctrine of Equivalents*, 59 STAN. L. REV. 955, 956–57 (2007).

201. *Festo*, 535 U.S. 722.

202. See Allison & Lemley, *supra* note 200.

203. See *supra* Sections II.A–D.

204. See Ted Hagelin, *The Experimental Use Exemption to Patent Infringement: Information on Ice, Competition on Hold*, 58 FLA. L. REV. 483, 533–34 (2006).

invention for experimental purposes. Courts allow experimentation if it is “not solely for amusement, to satisfy idle curiosity, or for strictly philosophical inquiry.”²⁰⁵ In *Madey v. Duke University*, the Federal Circuit went so far as to find that a university’s academic research constituted patent infringement and was not protected by the doctrine.²⁰⁶ The obvious question thus arises: can the current law on experimental use be squared with patent law’s parsimony principle? Perhaps counter-intuitively, the answer is very much in the affirmative.²⁰⁷

To understand why, it is important to appreciate some nuanced aspects of the law on experimentation. First, where the doctrine of experimental use applies, no infringement takes place and the patentee is entitled to zero compensation.²⁰⁸ As this Article suggests, judges should apply the parsimony principle in a minimalist fashion, so that downward adjustments in patentee compensation should occur in incremental steps. Granting rivals free reign to practice their competitor’s patented invention for the purpose of designing around it deprives the patentee of the exclusive rights to which she is literally entitled under the Patent Act.²⁰⁹ Economists recognize that access prices set at zero are rarely desirable, especially in the presence of fixed costs,²¹⁰ Such is the case here. Zero compensation is tantamount to under-compensation. It is not surprising then, that a patent system built on a parsimony principle would decline to adopt an expansive experimental use exemption.

The law’s refusal to countenance a broader experimental use doctrine is further justified on an additional ground. There is an unappreciated de facto experimental use right in the patent system. The cost of litigation coupled with the immense difficulty of detecting unauthorized experimentation results in considerable under-enforcement. Thus, as a practical matter, the ban on experimental use for commercial purposes is relevant for large-scale experimentation by a known rival of a patentee. In such circumstances, the

205. 307 F.3d 1351, 1362 (Fed. Cir. 2002).

206. *Id.*

207. For the author’s larger defense of the current law on experimental use, see Alan Devlin, *Restricting Experimental Use*, 32 HARV. J.L. & PUB. POL’Y 599 (2009).

208. See Sonya J. Bible, *Does the Experimental-Use Defense to Patent Infringement Still Exist?*, 13 SMU SCI. & TECH. L. REV. 17, 30–33 (2009).

209. 35 U.S.C. § 154(a) (2006).

210. See, e.g., William J. Baumol & Daniel G. Swanson, *The New Economy and Ubiquitous Competitive Price Discrimination: Identifying Defensible Criteria of Market Power*, 70 ANTITRUST L.J. 661, 668 (2003).

patentee's normative case for compensation is compelling and a parsimonious interpretation of the law would reasonably deny free access.²¹¹

As noted, the parsimony principle may advocate the cautious cabining of patentee's exclusive rights, lest those rights confer a windfall at the expense of cumulative innovation, consumers, and society generally. The minimalist manner in which judges should impose such limitations suggests that the current law on experimental use is sensible. Were patentees' enforcement efforts both pervasive and effective, such that they could enjoin all experimentation on their inventions, this could result in overcompensation. The maximalist and quasi-maximalist approaches might approve of the current law even in this case, given the zero-access-price issue, but a parsimonious approach would suggest otherwise. In such a setting, a compulsory-access regime at an access price specified by the government would balance patentee compensation with the broader social benefits of enhanced knowledge. Nevertheless, current reality, which is characterized by patentees' limited ability to detect—and to afford to challenge—experimentation, makes the fear of patentee overcompensation on this ground attenuated.

IV. THE PARSIMONY PRINCIPLE AS A GUIDE TO PATENT POLICY

Having both explored contemporary patent doctrine and unearthed a unifying principle that can guide the law, this Article now demonstrates the effects of applying the parsimony principle to current debates in patent jurisprudence. This Part focuses on two areas of particular contemporary importance: the proper limitations to be imposed on the patentability of processes and the optimal manner in which to quantify damages in the event of infringement.²¹² As the following discussion makes clear, the parsimony principle provides a useful framework for analyzing contemporary challenges.

211. The counter-argument would be that this is where an experimental-use exception is most needed, as major rivals are most unlikely to license one another. Such an objection would not be entirely convincing. First, rivals routinely engage in expansive reciprocal cross-licensing of their patent portfolios and enter into patent pools. *See* Davis, *supra* note 160, at 438–39. Second, and more fundamentally, it is precisely in the context of experimentation by a known rival that some pecuniary compensation is most important in providing patentees a sufficient reward. For the author's larger discussion of this point, see Devlin, *supra* note 207.

212. The ensuing discussion is of course normative and for that reason it is important to specify the substance and contours of the parsimony principle. For without the necessary specificity, the Article could offer only general observations as opposed to specific prescriptions. *See supra* Section II.B.

A. EMPLOYING THE PARSIMONY PRINCIPLE TO INFORM
CONTEMPORARY DEBATE

1. *Exploring the Limits of Patentable Processes—Computer Software and
Business Methods*

As patent law transcended its original fields of relevance to encapsulate the new economy, the parsimony principle took on new levels of importance. Today, patent law sweeps broadly. The one-size-fits-all doctrine employed across a wide variety of disparate industries creates asymmetric incentive effects. Patentee overcompensation is a very real threat in some new-economy industries, most notably the computer-software and business-method sectors.²¹³ An ever-increasing number of voices continue to contribute to the already voracious cries for intervention in this field. Foremost among these aired concerns is that earlier acquirers of patent protection are preempting vast swathes of subsequent innovation.²¹⁴ Whether patents are presently impeding innovation in such controversial settings as computer software, business methods, and medical diagnostic techniques is an empirical question. Despite raucous debate, this question remains unanswered. But the ongoing controversy is notable in particular for the nuclear option espoused by critics of the patent system.

Some of these commentators advocate the wholesale elimination of patent protection over computer software, business methods, and medical diagnostic techniques, amongst other fields of innovation.²¹⁵ Absent convincing empirical evidence that the patent system is indeed inflicting serious social costs, a parsimonious approach to innovation policy would react to such calls in an agnostic fashion. While it may indeed be the case that a vast web of preemptive patent rights encumbers follow-on invention, it does not follow that those ownership rights are not spurring crucially important innovation that lays a foundation for future R&D. Were the patent system to withdraw from these fields entirely, it would likely solve one problem only to create another.

The worst story painted by the evidence so far is that individual and small-firm innovation is being hindered by the proliferation of patents owned by a limited number of large companies.²¹⁶ Thus, if innovation is being

213. See *supra* sources cited and text accompanying note 107.

214. See Charles Babcock, *Stanford Law Professor Raps Patents as Barrier to Innovation*, INFORMATIONWEEK (Apr. 7, 2005), <http://www.informationweek.com/news/software/showArticle.jhtml?articleID=160502321>.

215. See Janis, *supra* note 71.

216. Large companies with expansive patent portfolios are more likely to have formal cross-licensing arrangements in place with one another. Such arrangements eliminate

stifled, it would seem to be with respect to a limited population of inventors. Of course, this is not to downplay the crucial role in innovation played by individuals and modest-scale enterprises. But if the overall rate of innovation is of concern, rather than the distribution of innovative activity, then the anecdotal evidence of patent law's nefarious effect on overall technological progress in the IT sector is not yet compelling.²¹⁷ This being so, a broad withdrawal of patent protection for computer software and methods of conducting business may be of questionable prudence.

Instead, if there is reasonable basis for believing that patent rights in the IT sector are currently too broad or numerous,²¹⁸ the law might benefit by looking to the subtle role played by the parsimony principle elsewhere. As a general matter, this principle suggests further limitations to the breadth and duration of exclusionary rights so as to contain the danger of overreaching, and hence overcompensation. The principle also operates at the margin by shaving degrees of exclusivity, and hence profitability, from patent rights rather than resorting to drastic options. Modest increments in policy-setting are desirable, for they limit the potential harm of erroneous decision-making. If, as at least some evidence suggests, computer-software and business-method patents currently provide original inventors too much control over the subsequent development of technology, then the answer might lie in modest adjustments in patent law's policy levers.²¹⁹

In particular, the courts might display some willingness to employ the reverse doctrine of equivalents with greater regularity in the IT sphere.²²⁰ At present, the doctrine has very limited practical significance, which is defensible in industries where patents are clearly crucial for ongoing innovation and where patent thickets are less problematic.²²¹ But in the field

blocking positions. Yet, individuals and companies with more modest portfolios may not be privy to such agreements, thus further encumbering their research and commercialization efforts.

217. See Ronald J. Mann, *Do Patents Facilitate Financing in the Software Industry?*, 83 TEX. L. REV. 961, 1004–09 (2005).

218. See, e.g., Daralyn J. Durie & Mark A. Lemley, *A Structured Approach to Calculating Reasonable Royalties*, 14 LEWIS & CLARK L. REV. 627, 640 (2010).

219. See John F. Duffy, *Rules and Standards on the Forefront of Patentability*, 51 WM. & MARY L. REV. 609, 622–23 (2009).

220. The reverse doctrine of equivalents serves to free a sufficiently large leap forward in the prior art from the claims of an earlier patent that would otherwise block use of the new technology. See *Graver Tank & Mfg. Co. v. Linde Air Prods. Co.*, 339 U.S. 605, 608–09 (1950).

221. See *Roche Palo Alto L.L.C. v. Apotex, Inc.*, 531 F.3d 1372, 1378 (Fed. Cir. 2008) (“The reverse doctrine of equivalents is rarely applied, and this court has never affirmed a finding of non-infringement under the reverse doctrine of equivalents.”).

of computer software and business methods, perhaps the judiciary should see significant improvements on patented prior art as being sufficiently creditable to warrant being viewed as a break from prior technology. This would require the courts to distinguish between mundane or limited advances in technology from truly significant leaps that advance the art to a new level. Perhaps an appropriate, though very rough, metric for applying the reverse doctrine of equivalents in fields of potential overcompensation would be where a technological advancement renders the asserted patent's technology commercially defunct. Of course, this would only be a partial solution. Courts would need to implement this doctrine cautiously. This is because an overextension of the reverse doctrine of equivalents would risk under-rewarding initial inventors by denying them exclusive rights over subsequent discoveries.

Having noted such concerns, this Article proceeds to address the computer-software and business-method debates in greater detail. In adopting innovation policy, it is important to tie the incentive characteristics of inventors in particular industries to the nature of the reward that society may grant them. This Article explains the controversial nature of patenting innovation in these two particular fields on several grounds.

First, originalists object to patent protection in these spheres because they are not reminiscent of traditional spheres of innovation.²²² For such commentators, Congress designed the patent system to reward inventors in industrial settings involving scientific progress, not mere methods of conducting one's business or algorithms that instruct computers to perform a particular task.²²³ These objections are not especially compelling. The patent system should be concerned with promoting the invention of new, useful, and non-obvious products and processes. That such desirable output might emanate from contexts outside the contemplation of Congress in 1952 is not surprising, for it is the inevitable consequence of a highly innovative economy.

Second, unusual incentives to invent might exist independent of the patent system with respect to these two contentious areas of innovation. These separate incentives are highly relevant to any policymaker attempting to formulate patent policy on the basis of a parsimony principle. One wishing to cabin patentee compensation to a level not grossly exceeding his reservation return must know what return an inventor would enjoy were all

222. See John R. Thomas, *The Patenting of the Liberal Professions*, 40 B.C. L. REV. 1139, 1141 (1999).

223. See Francisc Marius Keeley-Domokos, Note, *State Street Bank & Trust Co. v. Signature Financial Group, Inc.*, 14 BERKELEY TECH. L.J. 153, 167–69 (1999).

patent protection denied to him. In some settings, the lack of such a reward would eliminate innovative activity. No company will devote billions of dollars' research to potential drug research for altruistic purposes.²²⁴ But computer software and business methods are highly dissimilar. The former is already subject to copyright protection, given that it is an original work of authorship that is fixed in a tangible form of expression.²²⁵ It is likely the case that this form of IP protection will suffice for some inventors. As for business methods, depending on the nature of the particular innovation, it might be the case that the cost-reduction or consumer-demand-enhancing virtue of a novel method of conducting business will in itself spur their creation. In both these cases, further protection under the patent system will by definition result in windfall profits.²²⁶

If this were the full story of innovation in these two fields, then we could safely advocate for the elimination of patent protection within them. But like so many things in life and law, this issue is far from straightforward. With respect to software, there are rather dramatic differences between the protections afforded an inventor by the patent and copyright regimes, respectively. There is no fair-use exception in the former field of IP, nor is there a "clean room" exemption for inventors who independently discover a previously patented technology.²²⁷ Although copyright law grants protection of far greater duration, the preventive powers afforded qualifying inventors is considerably weaker.²²⁸ Given these vast differences, copyright protection is surely appropriate for some inventors of software, yet insufficient for others.²²⁹ In short, some inventors of new and useful software will receive

224. It has been estimated that the average cost of developing a successful drug is approximately \$800 million. *See* INT'L FED'N OF PHARM. MFRS. & ASS'NS, A REVIEW OF EXISTING DATA EXCLUSIVITY LEGISLATION IN SELECTED COUNTRIES (4th rev. ed. 2005), available at http://www.ifpma.org/documents/NR2306/DataExclusivity_JAN05_revised.doc.

225. *See* *Computer Assocs. Int'l v. Altai*, 982 F.2d 693, 703–04 (2d Cir. 1992).

226. A further contributing factor to the widespread criticism of these processes is their notoriety for being erroneously granted. In other words, the technical requirements of patentability were being applied in a subpar fashion by the PTO, due in large part to the lack of prior art by which to judge obviousness and novelty. *See* Burk & Lemley, *supra* note 138, at 1169 n.59.

227. *See* Lemley, *supra* note 151, at 33.

228. The lack of an independent invention defense may be particularly relevant in the computer-software industry, where instances of clean-room invention are likely to abound. Copyright protection alone may deny original inventors an adequate return. Of course, it bears noting that this same phenomenon warrants less-strong patent rights in the computer-software industry, for otherwise downstream cumulative research may be unduly hindered.

229. *See* Michael Chapin, Note, *Sharing the Interoperability Ball on the Software Patent Playground*, 14 B.U. J. SCI. & TECH. L. 220, 228 (2008).

inadequate returns absent patent protection. The parsimonious approach to patent jurisprudence would not therefore support a large-scale withdrawal of patent protection in the field of software.

Similarly, with respect to business methods, some such forms of innovation may be forthcoming absent patent protection, but not all. In a prior work, the author explored various forms of innovation and concluded that a particular subset of business methods—namely those that are self-consuming—should not be patent-eligible.²³⁰ This is because free-market pressures to reduce cost and increase consumer demand act as independently sufficient incentives to invent in-house processes. Commercialization costs for such inventions are limited and the social costs of patenting them are apt to be significant, given their prime candidacy to act as mechanisms for raising rivals' costs.²³¹ Yet, it is unlikely that all forms of conducting business create such innate utility to their inventors that they would be invented absent patent eligibility. Some inventors may devote resources to formulating valuable business methods with an eye toward licensing the relevant processes to interested third parties. Thus, the parsimony principle would not seem to counsel an absence of patent protection in this field.

That the parsimony principle does not command such an outcome in the computer-software and business-method contexts does not mean that these fields' unique proclivity for self-creation is irrelevant. Quite to the contrary, it is highly probable that patent law needs less potent exclusionary rights to trigger socially desirable levels of innovation in these contexts than in other, less controversial areas of patentable discovery. A parsimonious approach would therefore counsel more demanding patentability prerequisites and reduced levels of exclusivity in these fields.

Controversy around computer-software and business-method patents has raged for some time, most particularly since the Federal Circuit's 1998 decision in *State Street Bank*.²³² That decision paved the way for patentability of processes that bore "a useful, concrete, and tangible result."²³³ *State Street* opened the door for the patentability of computer programs and business

230. See Alan Devlin & Neel Sukhatme, *Self-Realizing Inventions and the Utilitarian Foundation of Patent Law*, 51 WM. & MARY L. REV. 897 (2009).

231. The infamous Amazon "one-click" patent is an obvious example. See Method and System for Placing a Purchase Order via a Communications Network, U.S. Patent No. 5,960,411 (filed Sept. 12, 1997).

232. *State St. Bank v. Signature Fin. Grp.*, 149 F.3d 1368 (Fed. Cir. 1998), *abrogated by In re Bilski*, 545 F.3d 943 (Fed. Cir. 2008), *rev'd*, *Bilski v. Kappos*, 130 S. Ct. 3218 (2010).

233. *Id.* at 1373.

methods on an unprecedented scale.²³⁴ The proper treatment of business methods and computer software was once again in the spotlight due to the Federal Circuit and Supreme Court's respective decisions in *Bilski*.²³⁵ The Federal Circuit jettisoned the test it enunciated in *State Street* and instead adopted a "machine-or-transformation" test for the patentability of methods.²³⁶ The relevance of the Federal Circuit's espoused machine-or-transformation test for this Article's purposes lies in the negative impact it would have had on the patentability of business methods and computer software. Even putting aside crucial questions about the application of the machine-or-transformation test—and in particular what degree of connection to a computer would suffice to say that a business method or algorithm is "tied to a particular machine"—it became apparent that the Federal Circuit's decision seriously restricted the reach of patentable subject matter.²³⁷ With the exception of internally consumed business methods, which courts ought not to regard as being within the sphere of patentability, the Federal Circuit's decision in *Bilski* can be criticized for sweeping too broadly. As the preceding discussion indicates, the parsimony principle would rather cabin excess through incremental adjustment in doctrine rather than entirely withdraw the benefits created by the patent system's allotment of private ownership in innovation.

The Supreme Court's decision to grant *certiorari* in *Bilski* created a maelstrom of attention. The Court was inundated with an excess of sixty amicus curiae briefs, which took a range of different positions.²³⁸ Many saw the case as the perfect opportunity for the Court to signal the death knell of business methods and computer software as proper subjects of patent law.²³⁹ They were to be disappointed. Instead of formally excluding particular forms of innovation, such as business methods, from patent protection, the Court saw fit to first return to principles.²⁴⁰ Justice Kennedy, writing for the

234. See, e.g., Matthew E. Fink, Note, *Patenting Business Methods in Europe: What Lies Ahead?*, 79 IND. L.J. 299, 301 (2004).

235. *In re Bilski*, 545 F.3d 943 (Fed. Cir. 2008), *aff'd*, *Bilski v. Kappos*, 130 S. Ct. 3218 (2010).

236. *Id.* at 959.

237. See Christopher A. Brown, *Developments in Intellectual Property Law*, 43 IND. L. REV. 837, 837 (2010).

238. See *Merit Briefs for November Supreme Court Cases, Term 2009–2010 (ABA Division for Public Education)*, A.B.A., <http://www.abanet.org/publiced/preview/briefs/nov09.shtml> (last visited Dec. 13, 2009).

239. One can verify this by a casual look at some of the amicus briefs filed with the Supreme Court in the case. See also Christopher A. Cotropia & James Gibson, *The Upside of Intellectual Property's Downside*, 57 UCLA L. REV. 921, 943 (2010).

240. *Bilski v. Kappos*, 130 S. Ct. 3218, 3226–27 (2010).

majority, noted the value of the machine-or-transformation test, but rejected the contention that that particular test is the exclusive means to identify the contours of patentable subject matter.²⁴¹ Instead, it reoriented the question of unpatentable subject matter to its traditional foundation, notably “the exceptions for laws of nature, physical phenomena, and abstract ideas.”²⁴² Applying those principles, the Court had little difficulty finding the relevant invention to be nothing more than an impermissible patenting of an abstract idea.²⁴³

The Court’s approach is broadly in line with the parsimony principle. The only limitation that was not imposed involved self-realizing inventions in the form of self-consumed business methods. This, of course, is a field of invention that bears an unacceptable propensity for wide-ranging exclusive reach. As explained in Section III.A, *supra*, such processes will almost inevitably over-reward inventors and violate the parsimony principle.

Nevertheless, the Court’s decision to eschew categorical line-drawing need not be problematic if courts vigorously enforce the novelty, non-obviousness, and utility conditions of patentability to ensure that any given invention goes beyond “the results of ordinary innovation [which] are not the subject of exclusive rights under the patent laws.”²⁴⁴ Thus, judges should apply the non-obviousness requirement with unusual vigor and the reverse doctrine of equivalents should find new teeth in the computer-software and business-method contexts. These changes would grant deserving inventors their due reward. It would fulfill the constitutional mandate of promoting the useful arts, whilst denying patent protection to inventors of mundane or trivial technological improvements, who would otherwise create social cost by contributing to a potential patent thicket or by frustrating cumulative research.

2. *Optimal Damages in the Event of Infringement*

Contemporary efforts at patent reform focus on the question of damages. Even beyond the threshold question of whether patentees should be entitled to an injunction or mere monetary relief, there is considerable debate as to whether courts are awarding excessive or inadequate pecuniary awards in the event of established infringement.²⁴⁵ The current methodology

241. *Id.* at 3227.

242. *Id.* at 3226.

243. *Id.* at 3231.

244. *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 427 (2007).

245. *See, e.g.,* Lemley, *supra* note 151, at 20; Opderbeck, *supra* note 67, at 137–50.

for calculating damages lies in the “entire market value rule.”²⁴⁶ That approach essentially provides that, if a patented technology lies at the heart of an invention, then courts will award royalties based on sales of the entire product. The entire market rule has been quite controversial.²⁴⁷ Infringers accuse the approach of giving rise to excessive rewards, but patentees further malign the doctrine for short-changing them in the event of infringement. The issue of damages may well be the single most divisive issue weighing on patent law today.

One’s view on reforming damages calculations depends in large part on one’s industrial background. Those in communication and computer industries decry the cost of litigation to which they are routinely subject.²⁴⁸ As observed *supra*, products in such sectors often implicate hundreds, even thousands, of patents, so awarding any one patentee an inordinate reward for infringement threatens to impose significant burdens on innocent infringers.²⁴⁹ To companies operating in these industries, the solution lies in “apportionment,” which would tie the amount of damages awarded to the relative contribution of the infringed patent’s technology.²⁵⁰ They favor such an approach because they believe that courts and juries misapprehend the entire-market rule, and that misapprehension results in infringing entities of a product’s small component being nevertheless treated as if the relevant patent lies at the heart of the commercial enterprise.²⁵¹ The Patent Reform Act of 2009 contemplated such an approach to calculating damages.²⁵²

Not surprisingly, apportionment is fiercely opposed by industries in which commercialized products do not implicate a wide variety of patented technologies.²⁵³ In many such sectors, including the biotechnology industry, a

246. See *Imonex Servs., Inc. v. W.H. Munzprufer Dietmar Trenner GmbH*, 408 F.3d 1374, 1379 (Fed. Cir. 2005).

247. See, e.g., Mark A. Lemley, *Distinguishing Lost Profits from Reasonable Royalties*, 51 WM. & MARY L. REV. 655, 656 (2009). See generally Brian J. Love, Note, *Patentee Overcompensation and the Entire Market Value Rule*, 60 STAN. L. REV. 263 (2007).

248. See Jerome H. Reichman, *Intellectual Property in the Twenty-First Century: Will the Developing Countries Lead or Follow?*, 46 HOUS. L. REV. 1115, 1121–22 (2009).

249. See Lemley, *supra* note 151, at 20 (noting that current rules “permit excessive recoveries” in patent cases); Michael Meehan, *Increasing Certainty and Harnessing Private Information in the U.S. Patent System: A Proposal for Reform*, 2010 STAN. TECH. L. REV. 1, 9 (observing that, “in the computer and software industries, there can be thousands of patents on a single product”).

250. See Lemley, *supra* note 247, at 661–69.

251. *Id.*

252. Patent Reform Act of 2009, S. 515, 111th Cong. (2009).

253. See Opderbeck, *supra* note 67, at 136.

company may establish a product on merely a single patent, or a few more.²⁵⁴ Here, the prospect for patentees' windfall is significantly reduced, for the presence of merely a limited number of patents presumably increases the expected technological contribution of each one. In such cases, the entire market value rule would seem to make sense.

The parsimony principle, far from dictating the innate superiority of one unwavering view on apportionment over another, suggests that both camps have it partially right. There is little question that juries are more likely to overcompensate, through a damages award derived from overall sales of a commercialized product, patent owners of software or IT technology than companies that own rights in pharmaceutical formulae or biotechnology. In the former setting, courts obviously should not require infringers of peripheral components to pay royalties based on entire sales, unless the courts discount the royalty rate to a level that reflects the fringe nature of the infringed technology. In the latter context, patentees ought to be entitled to a far larger proportional return.

In this regard, there is much to commend in the Federal Circuit's 2009 decision in *Lucent Technologies, Inc. v. Gateway, Inc.*, which rejected a 358 million dollar award against Microsoft.²⁵⁵ There, the jury awarded Lucent an 8% royalty on the entire market value of Microsoft's Outlook software, even though the patented method was but one of "hundreds, if not thousands" of features in the challenge product.²⁵⁶ The court properly concluded that it was "inconceivable" that "the use of one small feature, the date-picker, constitutes a substantial portion of the value of Outlook."²⁵⁷ If district courts make discerning judgments about the centrality of the patented technology to the infringing product, then the prospect of over-reward should not be of great concern. This is particularly so if courts endeavor to employ an appropriately reduced royalty rate, the dilution of which should reflect the extent to which the commercialized product includes technology other than the patented one at issue.

Ultimately, cost-justified asymmetric treatment lies at the heart of sound policy, for parsimonious treatment is difficult to achieve otherwise. There is no problem with a universal standard, as long as that standard is subject to

254. See Kevin W. McCabe, *The January 1999 Review of Article 27 of the TRIPS Agreement: Diverging Views of Developed and Developing Countries Toward the Patentability of Biotechnology*, 6 J. INTELL. PROP. L. 41, 55 (1999) ("[B]iotechnology is highly patent sensitive, in that a single patent can dominate a marketed product.").

255. 580 F.3d 1301 (Fed. Cir. 2009).

256. *Id.* at 1332.

257. *Id.*

flexible application, with sensitivity to the unique characteristics of the situation at hand. In that regard, one should not be overly concerned with the choice between the entire market rule and an apportionment alternative, as long as courts can be trusted to recognize the idiosyncrasies attendant upon each individual case.²⁵⁸

B. AN IMPORTANT CAVEAT: AVOIDING EX POST DILUTION OF PROPERTY RIGHTS

This Article contends that a pervasive theme runs through the patent system. Specifically, the patent doctrine evolved in a manner that demonstrably limits inventors' rewards, thereby diminishing the prospect of gross overcompensation.²⁵⁹ The influence of the parsimony principle on patent law is far-reaching, spanning from denying patent protection for abstract inventions, to limited rights for injunctive relief, and to limits on contractual freedom. Such constraints on windfall returns are normatively justified for a variety of compelling reasons, which are recounted at length in Section II.A.

Yet it is appropriate to end on a cautionary note. It would be easy to read the preceding pages as advocating for the widespread dilution of patentee's property rights. However, such a broad reading would be mistaken. An important distinction exists between limiting inventors' compensation through an appropriate tailoring of patent eligibility, patent duration, and patent breadth, on the one hand, and ex post dilution of property rights, on the other. The latter fear is realized most directly by limitations on patentees' exclusive rights. Although the Article advocates the withholding of injunctive relief in certain cases—most notably those involving patent trolls and in certain peripheral infringement scenarios—courts should generally regard patentees' right to exclude as sacrosanct, especially where infringers had notice of the relevant patent. If courts begin denying injunctions on a more regular basis—that is, in situations not countenanced by the exceptions discussed above—patentees will increasingly be subject to liability rules.

Such an outcome would be problematic for a variety of reasons. Strong property rights lend themselves well to efficient contractual agreement, save in cases where transaction costs become prohibitive.²⁶⁰ They do so by forcing

258. As the following discussion makes clear, however, the relative virtues of apportionment and of the entire market rule do not alter the fact that injunctive relief should generally be the preferred option.

259. See *supra* Part III.

260. See D. Bruce Johnsen, *Myths About Mutual Fund Fees: Economic Insights on Jones v. Harris*, 35 J. CORP. L. 561, 610 (2010).

those wishing to avail of an owned resource to bargain for permission up front, before they devote possibly forgone capital if the owner later denies access. Liability rules, in contrast, can encourage prospective licensees to infringe today and worry about paying tomorrow.²⁶¹ In addition to threatening to under-compensate inventors, they create costly ex post judicial disputes over “reasonable” damages.²⁶² Ultimately, the optimal choice between property and liability rules is the subject of extended scholarly analysis and is beyond the scope of this Article.²⁶³ Suffice it to say that significant gains accompany treating patents as giving rise to property rules, with exceptions to this approach limited to instances where the parsimony principle and related policy deem such exceptions warranted.

V. CONCLUSION

The study of patent law is fascinating for many reasons. Of these reasons, the most engrossing is how the law formulates policy in the impenetrably complex world of innovation. Inducing inventors to devote their human and pecuniary capital to the task of innovation is an exercise optimally conducted with sensitivity to the vastly asymmetric incentive characteristics of different inventors. Unfortunately, there is much that we do not know. We remain significantly ignorant about innovation, the macro-forces that weigh upon it, and the idiosyncratic influences that propel it at the individual level.²⁶⁴ Most of all, we lack crucial empirical data on the relationship between enhanced patent rights and industry-specific rates of innovation. While we certainly know that patents constitute the sine qua non for certain industries, notably pharmaceuticals,²⁶⁵ we still do not know whether contemporary patent rights are penurious or excessively generous.

261. Perhaps counter-intuitively, this feature of liability rules may have some virtue in the case of unused patents. See Julie S. Turner, Comment, *The Nonmanufacturing Patent Owner: Toward a Theory of Efficient Infringement*, 86 CALIF. L. REV. 179, 204 (1998).

262. See Robert P. Merges, *Contracting into Liability Rules: Intellectual Property Rights and Collective Rights Organizations*, 84 CALIF. L. REV. 1293, 1317 (1996).

263. For the classic treatment of this subject, see generally Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1111–15 (1972).

264. Cf. Richard S. Whitt & Stephen J. Schultze, *The New Emergent Economics of Innovation and Growth, and What It Means for Communications Policy*, 7 J. TELECOMM. & HIGH TECH. L. 217, 267 (2009) (“Innovation is a much-admired concept, yet in many ways still rather mysterious and elusive.”).

265. Individual pharmaceutical companies are hugely dependent on the patents they possess, as the looming expiration of many valuable patents in 2011 evidences. See Vijay Vaitheeswaran, *Generically Challenged*, THE ECONOMIST – THE WORLD IN 2010, Nov. 21,

Although the optimal contours of patent law remain elusive, it does not follow that policymakers are condemned to ignorance. In particular, the tools of economics provide a rich source of theoretical guidance. By virtue of that discipline's teachings, we know that the public goods nature of information threatens its underproduction absent a reward structure. But the award of property rights creates monopoly distortions, the negative impact of which rises in parallel with the value of the owned resource. In addition, exclusive rights of greater breadth and duration tighten access to information and increase the expense of commercialization. Combining the theoretical predictions that stronger IP rights create more poignant incentives to invent, but simultaneously result in greater social costs, there exists an optimal balance. Specifically, rewards should be tailored in such a way that the marginal gain in innovation from enhanced exclusivity precisely equals the cost to society from the increased right to exclude. Economics can define this point theoretically, but cannot identify it in practice.

This concept might be left awash in a sea of doubt, but for some further knowledge. Specifically, it is almost certainly the case that complete propertization of information, encapsulated by the maximalist ideal, would widely miss the mark. Inventors of valuable technologies will rarely require the full social value of their inventions to induce them to invent, and a uniform oligarchical regime would radically overcompensate the rest. Nor is the abolitionist approach likely to be an acceptable enablement to the optimal trade-off between exclusion and free access. We can reject this perspective because its adoption is not supported by the currently ascertained data. We are not entirely bereft of empirical evidence, for the U.S. knowledge economy stands as a towering edifice with the power of the patent regime at its foundation. Abandoning the framework for the basis of such success would be policy of the most reckless kind.

We are thus left with some important conclusions about the general nature of innovation policy, but with limited information about the precise rules that should be formulated between the extreme alternatives of maximalism and abolitionism. Given the pervasive level of uncertainty currently residing at the margins of U.S. patent law, policymakers must choose between erring on the side of significant overcompensation and imposing meaningful limitations on patentees' returns. Although both approaches would display sensitivity to the social cost of departing from the optimal level of patent protection, the quasi-maximalist path would rather

2009, at 130 (discussing the impending patent cliff facing the pharmaceutical industry in 2011).

countenance a partial windfall situation. It would accept tales of industry-specific overcompensation as being preferable to complaints of inadequate returns. Under a quasi-maximalist system, we would see doctrine crafted in such a way as to give patentees broad control over the disposition of their inventions. One would expect to seldom encounter restrictions on patentable subject matter, contractual freedom on the part of patentees to secure their rights, a generous doctrine of equivalents and, most of all, an automatic entitlement to injunctive relief in the event of established infringement. Yet, one observes none of these things.

Instead, this Article finds a near-unqualified theme in patent doctrine toward a parsimonious solution to the information quandary faced by policymakers. Of the major tenets of patent law that weigh on the prospect of patentee overcompensation, virtually all display sensitivity to this threat. Discoveries that courts and the PTO view as excessively “abstract” are ineligible for patent protection—a result explained by the fact that open-ended processes, untethered to specific applications, will foreclose vast areas of activity and will likely yield a windfall on the inventor. Similarly justifiable restrictions exist on an inventor’s ability to claim her invention. The Federal Circuit cabined the reach of the doctrine of equivalents. Serious restrictions exist on patentees’ contractual rights, which if exercised freely, would serve dramatically to enhance the value of the relevant IP rights. And in light of *eBay*, injunctive relief is by no means readily forthcoming in all cases of infringement. Without ever saying so explicitly, the patent doctrine evidently evolved along a parsimonious path.

This is not to say that the current patent system under-compensates inventors. Indeed, many respected commentators continue to worry that patents are overly generous and that the proliferation of patent rights threaten to stifle future innovation.²⁶⁶ Nevertheless, the patent system has undergone a significant effort to cabin the prospect of serious overcompensation. To the extent that the limitations imposed on the patent doctrine reduced the incidence and severity of windfall returns on inventors below what would prevail under maximalist and quasi-maximalist regimes, the gains thereby attained are commendable. This holds true even if current rewards are excessive in light of the theoretical, though empirically unidentifiable, optimum. Whether policymakers should attempt to weaken patentee compensation further, either by restricting the field of patentable subject matter or reducing the breadth or duration of awarded patent rights,

266. See Richard A. Posner, *Pharmaceutical Patents*, BECKER-POSNER BLOG (Dec. 12, 2004), <http://www.becker-posner-blog.com/2004/12/pharmaceutical-patents--posner.html>.

is a difficult question to answer. This Article argues that any further downward adjustments should be made in incremental steps rather than through more extreme measures. Absent further evidence mandating such a course of action, the complete withdrawal of patent protection from the computer software, business methods, and other controversial fields seems excessive. A better step would be to find ways to limit the force of patents awarded in such sectors and to heighten the standard required for obtaining patent protection in those industries. In particular, the reverse doctrine of equivalents should enjoy new-found force and courts should apply the non-obviousness condition with particular vigor. Should evolution in the law proceed along such lines, it would be fully consistent with the teachings of patent law's parsimony principle.

THE ILLUSION OF THE COMMONS

Jonathan M. Barnett[†]

TABLE OF CONTENTS

I.	INTRODUCTION	1752
II.	COMMON THOUGHTS ON THE COMMONS	1757
III.	THEORY: THE FRAGILITY OF THE COMMONS	1760
	A. REGIME STRUCTURES.....	1762
	B. THE COOPERATION GAMBLE	1767
	C. MAKING COOPERATION LAST.....	1769
	1. <i>Innovator Options and Types</i>	1770
	2. <i>Conditions for Cooperation</i>	1772
	a) Group Size.....	1773
	b) Capital Intensity.....	1773
	c) Asset Values.....	1774
	d) Endowment Heterogeneity.....	1774
	3. <i>From Individual to Collective Defection</i>	1776
	4. <i>Predicting Cooperation</i>	1777
IV.	EMPIRICS: THE COMPLEXITY OF THE COMMONS	1779
	A. REGIME TAXONOMY.....	1781
	B. OPEN SHARING REGIMES	1784
	C. CLOSED SHARING REGIMES.....	1788
	D. SUMMARY.....	1792
V.	CASE STUDIES: THREE ILLUSIONS OF THE COMMONS	1793
	A. CRAFT GUILDS.....	1794
	B. ACADEMIC RESEARCH	1799

© Jonathan M. Barnett.

[†] Associate Professor, University of Southern California, Gould School of Law. I am grateful for comments on earlier versions from Brett Frischmann, Gillian Hadfield, Dan Klerman, Shmuel Leshem, Bob Rasmussen, Jennifer Urban, and participants at the MIT Innovation Workshop in May 2009, the Intellectual Property Scholars Conference in August 2008, and a USC Law School Faculty Workshop in September 2007. Appreciation to Christin Chang and the library staff at the USC Law School for research and other assistance. This project has been generously supported by the financial support of the Provost's Office of the University of Southern California. All errors are mine.

C. OPEN-SOURCE SOFTWARE	1804
V. CONCLUSION	1813

I. INTRODUCTION

Conventional wisdom states that intellectual property or some other barrier to imitation is a prerequisite for intellectual production: without it, innovators¹ will decline to place capital at stake. But this proposition appears to be incompatible with markets where innovation proceeds with weak or no intellectual property rights and imitation is widespread. For example, copyrights over music are routinely violated, yet music production does not appreciably slow; software is widely pirated or voluntarily released with minimal or even no protections against copying, yet product releases continue apace; property rights over scientific theories and findings are nonexistent, yet research proceeds forward. These observations form the basis for an intellectual family of related positions that I loosely group under the rubric of the “commons thesis.” Based on the observation that innovation apparently thrives in certain contemporary or historical environments without constraints on imitation, scholarly and popular commentators increasingly conclude that intellectual property can be substantially retracted or abolished in some, most, almost all, or even all other settings at a net social gain.² Casually formulated, typical expressions in this vein run along the following lines: “Pythagoras, Galileo, and Shakespeare didn’t have intellectual property but were really creative, so . . .” or “magicians, tattoo artists, and hair stylists don’t use intellectual property but are really innovative, so . . .” or “scientists and artists aren’t in it for the money, so . . .” Novel technologies have lowered the costs of copying, modifying, and distributing informational goods, and intellectual property or other exclusionary barriers can appear to be outdated roadblocks to the production and exchange of intellectual goods. The various slogans that advance the commons model (such as “information wants to be free,” “free culture,” “free software,” and “free science”) are inherently attractive, pervade contemporary discourse, and have made substantial inroads in legal and even economic scholarship on intellectual property. Indeed, it is difficult to defend any other position: what publicly-interested argument would remain to support access restrictions if we “know” these are unnecessary to

1. By “innovator,” I refer to individuals, firms or other entities that are engaged in the invention, creation and/or commercialization of intangible goods. “Innovation markets” encompasses markets or market segments principally or substantially consisting of intangible goods in the form of creative or technological products or processes.

2. See *infra* notes 8–10 and accompanying text.

support creation and invention?³ The commons thesis has arguably become the new conventional wisdom.

But complex facts deserve complex explanations. Both the old and the new orthodoxy fall far short of adequately capturing the nuanced strategies used by individuals, firms, and other entities to capture returns on innovation. In previous publications, I have emphasized that a monolithic property-rights view that identifies intellectual property as a strict precondition for intellectual production is incompatible with the widespread use of alternative instruments that supplement or sometimes even substitute for intellectual property.³ But it would be equally myopic to adopt the polar view (as the commons thesis in its various incarnations argues with various degrees of force) that intellectual production can typically proceed at vigorous levels without any meaningful legal or other barriers to unauthorized use. If that were the case, then much, if not all, of the existing intellectual property infrastructure would be a policy error of substantial proportions, and an exceptionally successful and persistent case of privately interested rent-seeking. (A priori that possibility, of course, cannot be excluded and often may appear to be quite plausible!)

This Article provides a set of theoretical and empirical arguments that together cast serious doubt on that possibility. The conditions under which environments bereft of property rights or other exclusionary instruments have supported, are likely to support, and actually do support, capital-intensive forms of innovation appear to be profoundly limited. As innovation environments “scale up” along certain key dimensions, a basic economic logic governs: access must be—and is—regulated at some point in order to deliver the revenue streams that support the costly activities required to generate and commercialize technological and creative innovation.

Hence what I call “the illusion of the commons.” Environments that apparently sustain substantial levels of innovation investment in the absence of intellectual property are typically supported by some combination of legal and/or extra-legal instruments that restrain access to some portion of the total “consumption bundle” of products and services. Across a broad set of otherwise disparate markets and periods, there is surprisingly little support for the widely-expressed view that the “commons”—by which I mean open-access or substantially open-access environments—provides a reliable model

3. See Jonathan M. Barnett, *Property as Process: How Innovation Markets Select Innovation Regimes*, 119 YALE L.J. 384 (2009) [hereinafter Barnett, *Property as Process*]; Jonathan M. Barnett, *Is Intellectual Property Trivial?*, 157 U. PA. L. REV. 1691 (2009) [hereinafter Barnett, *Trivial*]; Jonathan M. Barnett, *Private Protection of Patentable Goods*, 25 CARDOZO L. REV. 1251 (2004) [hereinafter Barnett, *Private Protection*].

for understanding the incentive structures that elicit economically intensive levels of innovation. That in turn casts doubt on the widely-expressed normative assertion that intellectual property rights are an artificial distortion in cultural and technological markets that otherwise thrive and have thrived without access limitations.⁴ Virtually the contrary is the case: economically significant levels of innovation investment almost never appear without some form of property rights or other access limitations that operate with an approximately equivalent effect. While this critical observation challenges the commons model as a reliable basis for positive analysis of innovation environments or normative design of innovation systems, it does not simply restore the old orthodoxy that unqualified property rights are a categorical precondition for innovation. That too would be a gross oversimplification. Analytical rigor demands an intermediate approach that can account for the complexity and diversity of funding and appropriation models across the rich variety of historical and contemporary innovation settings. That approach must account for the nuanced manner in which innovator populations both implement property rights and/or other access limitations in order to capture returns on investment, while maintaining cooperative relationships that reduce the transaction-cost burdens inherent to property rights and other access limitations.

This project starts by taking the commons model extremely seriously. To do so, I use a simple rational-choice framework to identify the circumstances under which the commons model could provide a feasible environment for intellectual production. The result is a hypothetical “sharing regime”⁵ that relies on social norms to support innovation in the absence of legal or other barriers to imitation. In this construct, reputation-driven norms support an approximately reciprocal exchange of knowledge assets over time and avoid the under-provision outcome that normally results in public goods settings without exclusionary instruments. Relative to a law-based regime that relies on formal property rights to sustain contribution incentives, this norm-based regime has a great advantage: it avoids the transaction costs that burden the creation, exchange, and transmission of intellectual assets under a formal property-rights regime. But the social savings from reduced transaction costs must be paid for with the social losses from limited regulatory power. Reputation-driven norms exert no force against one-shot or other participants that have no rational interest in accumulating reputational capital and, consistent with the core findings of the economic literature on the

4. *See infra* notes 8–10 and accompanying text.

5. My specific understanding of this term, as distinguished from some related terms in the literature, is described subsequently. *See infra* note 18.

private provision of public goods, can be expected to exhibit declining force in general as any market exhibits increased group size, economic values, capital-intensity requirements and variation in innovative capacity.⁶ In short, the sharing model works in settings that are small in size, scale, value, and diversity. As markets mature and grow along those same dimensions, however, the sharing model tends to become an obsolete technology for supporting innovation.

Contrary to burgeoning “IP-skeptical” and “IP-rejectionist” currents in recent commentary,⁷ a rational-choice framework anticipates as a theoretical matter few markets that would be able to sustain economically intensive levels of innovative output without property rights or other exclusionary instruments. Note that this is a positive, not a normative, proposition. It simply is the case that certain specified factors will drive innovators to migrate toward more securely shielded environments in order to recoup innovation investments. To assess the empirical strength of these expectations, I broadly review existing evidence concerning legal and extra-legal appropriation instruments that operate in markets where innovation proceeds subject to weak or substantially incomplete intellectual property rights. This survey exercise—which provides a novel taxonomy of sharing practices in intellectual goods markets—demonstrates virtually the converse to the commons thesis. Each market that sustains economically significant levels of innovation without active adoption and enforcement of intellectual property is always allied to some other legal or extra-legal instrument that limits unauthorized usage. Put differently: the commons to which the commons thesis aspires hardly ever seems to exist! Even sharing regimes that apparently make little or no use of exclusionary instruments to secure innovation returns support this thesis. Lacking obstacles to imitation, these markets tend to be confined to technologically primitive markets where innovators have relatively insubstantial investments at risk. Consistent with theoretical expectations, reputation-driven norms substitute for intellectual property in order to sustain innovation incentives in “little-IP” settings characterized by low numbers (or large numbers organized into collective groups), low endowment heterogeneity, low capital-intensity levels, and low asset values. But these reputation-driven incentive structures tend to be replaced by property-based arrangements in “big-IP” settings characterized by large numbers, high endowment heterogeneity, high capital-intensity levels, and high asset values. In short: whenever intellectual production scales up, it adopts some form of intellectual property or other access limitations.

6. Each of these terms is defined in greater detail *infra* Part II.

7. See *infra* notes 8–10 and accompanying text.

Both theory and empirics recommend that we virtually flip the commons thesis on its head. In economically intensive settings, intellectual production does require intellectual property or some other exclusionary instrument to secure returns and thereby induce innovation. Strikingly, this proposition is made most evident in case studies of three markets that should be—and are often referenced as—among the best cases for the view that intellectual production does not require intellectual property: pre-modern craft production, academic research, and open-source software. Closer analysis shows that these “best cases” provide some of the most compelling illustrations against the commons thesis. Contrary to the “free appropriation” environment envisioned by the commons model, these markets rely on a complex mix of legal, extralegal and/or technological barriers, as complemented by reputational norms, to generate revenue streams in a manner that is ultimately consistent with the standard incentive framework. Critically, unmasking the illusion of the commons does not simply reiterate, but substantially re-orient, the standard incentive-based view of intellectual property. If we discard value-driven aspirations that innovation can typically proceed without any robust imitation barriers (or that innovation can typically proceed only with the most robust imitation barriers), then we can usefully reallocate scholarly resources to an alternative and strictly positive line of inquiry. Namely, the manner in which innovation markets use a mix of both property-based and sharing-based arrangements to support innovation incentives while mitigating the transaction costs and other social losses imposed by property rights and other exclusionary instruments. This approach offers a useful tool by which to capture the mixed incentive structures of real-world markets. It therefore improves upon both the new conventional wisdom, which privileges and consequently over-predicts the use of sharing arrangements, and the old conventional wisdom, which privileges and consequently over-predicts the use of property-based arrangements.

This Article’s core proposition is as follows: sharing arrangements are ubiquitous in innovation markets but, at least in economically significant settings, consistently operate against the background of property rights or some other combination of exclusionary instruments. The case studies vividly illustrate this nuanced view. Separated by great differences in time, location, and industry, each market operates subject to a mixed-form innovation regime that embeds sharing arrangements within a surrounding property infrastructure. This structure consists of: (1) a “sharing core,” where similarly-endowed innovators exchange knowledge assets subject to certain norm-based constraints, which is then shielded by (2) a “property perimeter” constituted by legal or extra-legal access restrictions, which in turn support a

bundled set of excludable products and services that generate revenue streams consistent with a conventional property regime. Under this hybrid regime, sharing practices do not substitute for intellectual property, but instead supply a transactional lubricant that facilitates the creation and improvement of intellectual goods while leaving intact an underlying infrastructure constituted by property rights and/or other exclusionary instruments. Again, this is a positive and not a normative proposition. It simply is the case that certain specified factors will drive firms or individuals to adopt sharing arrangements in order to eliminate the transaction-cost burdens inherent to a property regime and realize other gains from pooling knowledge assets. For this purpose, *property is a tonic, not an antidote*: sharing arrangements can scale up to “high stakes” environments by using property, contract, technology, and other exclusionary instruments to regulate access, thereby precluding the unraveling threats that threaten stand-alone sharing regimes.

Part II reviews claims to the effect that intellectual property or other exclusionary protections are not a prerequisite for intellectual production. Part III presents a stylized construct of a sharing regime that sustains innovation by recourse to social norms in lieu of property rights. Part IV presents a survey of empirical evidence on sharing regimes in cultural and technology markets. Part V presents case studies of pre-modern craft production, academic research, and open-source software.

II. COMMON THOUGHTS ON THE COMMONS

A substantial body of scholarly, popular, and policy discourse, as well as partisan positions taken by user communities and advocacy organizations, contest with varying degrees of intensity the conventional assumption that property rights or other imitation barriers are a typical precondition for innovation.⁸ Typical formulations anticipate a “world-to-come,” where

8. Any list of references will inherently be selective. *See, e.g.*, LAWRENCE LESSIG, *FREE CULTURE: HOW BIG MEDIA USES TECHNOLOGY AND THE LAW TO LOCK DOWN CULTURE AND CONTROL CREATIVITY* (2004) [hereinafter LESSIG, *FREE CULTURE*], at 29 (giving examples of non-consented use of original material by scientists, Hollywood studios, and Shakespeare), 53–61 (giving examples of film, TV, radio, and cable TV industries that were originally founded through various forms of intellectual piracy), 305–06 (arguing that an intellectual property regime that requires obtaining consent to use proprietary content stifles novel opportunities for creative expression facilitated by digital and online technologies); LAWRENCE LESSIG, *THE FUTURE OF IDEAS: THE FATE OF THE COMMONS IN A CONNECTED WORLD* 12–14 (2001) [hereinafter LESSIG, *FUTURE OF IDEAS*] (providing examples of musical creation, scientific research, and software development where innovators build on previous contributions without consent as characteristic of intellectual production); Michael A. Carrier, *Cabining Intellectual Property Through a Property Paradigm*, 54

informational goods are disseminated costlessly by intrinsically and/or altruistically motivated individuals assembled into spontaneously-ordered communities. Other formulations recall a “world-that-once-was,” where property rights were mostly absent and intrinsically and/or altruistically-motivated individuals exchanged knowledge in a collegial pursuit of intellectual inquiry and creative expression. For the commons literature, an open-access environment constitutes the natural policy baseline: that is, intellectual production often or typically proceeds well without intellectual property, which therefore largely reduces to a rent-seeking enterprise at the expense of the larger public.⁹

That of course begs the operational question: how do property-free environments induce innovators to spend resources on generating freely appropriable inventions? Or, as any investor would ask an entrepreneur: what is your revenue model? The commons literature often tends to bypass this question as immaterial: that is, it takes the view that innovators are motivated substantially or primarily by intrinsic motivations (e.g., “love of creation”) or altruistic motivations (e.g., “dedication to science”), in which case no remunerative mechanism need be identified.¹⁰ Relaxation or removal of the

DUKE L.J. 1, 36–37 (2004) (questioning the need for copyright, given that “many forms of creative expression—such as fashion, new words and slogans, jokes, and magic tricks, and the food industry—have flourished in the absence of protection”); Madhavi Sunder, *IP*³, 59 STAN. L. REV. 257, 260–61 (2006) (arguing that “rapid-fire technological advances and new forms of creative output . . . undermine utilitarian intellectual property law’s very premise: that intellectual property rights are necessary to incentivize creation”). For an example from the advocacy literature, see NANCY KRANICH, INFORMATION COMMONS: A PUBLIC POLICY REPORT 10 (2004) (noting that “throughout history” commons regimes have characterized pre-modern literary and agricultural production, with the suggestion that therefore creative production can proceed vigorously without intellectual property rights; however, author notes subsequently that “commons” research identifies restrictive conditions under which commons regimes are a sustainable regime for intellectual production). For an example from the popular literature, see John P. Barlow, *The Economy of Ideas*, WIRED, March 1994, at 8–9 (noting that storytelling, jazz improvisation, stand-up comedy routines, and other cultural forms proceed by incremental practices of free circulation and improvement, for which copyright law makes no accommodation, and arguing generally that intellectual property is of doubtful value given that innovators can usually accrue substantial returns as a result of first-mover advantage).

9. See, e.g., LESSIG, *FUTURE OF IDEAS*, *supra* note 8, at 12 (“[F]ree resources have always been central to innovation . . .”).

10. See, e.g., James Boyle, *The Second Enclosure Movement and the Construction of the Public Domain*, 66 LAW & CONTEMP. PROBS. 33, 45–46 (2003) (referring to the “innate human love of creation that continually drives us to create new things even when *homo economicus* would be at home in bed, mumbling about public goods problems”). The role of non-instrumental motivations forms the basis for a “social” model of intellectual production presented in YOCHAI BENKLER, *THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM* Chs. 3–4 (2006), which elaborates on arguments in Yochai

“rationality constraint”—that is, expected revenues must equal or exceed expected costs—may be plausibly compatible with certain “low stakes” fields of creative endeavor. It is not plausibly compatible, however, with the high stakes that private entities invest in developing and commercializing new technologies and ideas: for example, more than \$800 million on average in the case of a pharmaceutical drug¹¹ or more than \$3 billion in plant construction costs in the case of a semiconductor chip.¹² Attenuation or outright rejection of an instrumentalist model of innovative behavior distinguishes commons-styled claims from economic arguments that identify limited circumstances where imitation promotes innovation by profit-seeking firms. Invariably these arguments specify a remunerative mechanism that directly or indirectly rewards innovators based on some exclusionary protection. These models contemplate some form of property rights or other exclusionary instruments, which apply to some users or product attributes,¹³ re-appear at some other point in the aggregate bundle of products and services,¹⁴ or are waived by entitlement holders until some later time.¹⁵

Benkler, *Sharing Nicely: On Shareable Goods and the Emergence of Sharing as a Modality of Economic Production*, 114 YALE L.J. 273 (2005) [hereinafter Benkler, *Sharing Nicely*]; and Yochai Benkler, *Coase's Penguin, or, Linux and The Nature of the Firm*, 112 YALE L.J. 369 (2002) [hereinafter Benkler, *Coase's Penguin*].

11. See Joseph A. DiMasi et al., *The price of innovation: new estimates of drug development costs*, 22 J. HEALTH ECON. 151, 166 (2003) (estimating average capitalized costs of \$802 million for molecule identification and testing for drugs that underwent the FDA approval process in the 1990s (as calculated on a fully capitalized basis in 2000 dollars)). This figure is an underestimate insofar as it does not include production, distribution or marketing costs.

12. See Iikka Tuomi, *The Future of Semiconductor Intellectual Property Architectural Blocks in Europe*, JRC SCIENTIFIC AND TECHNICAL REPORTS 74 (2009).

13. See, e.g., Stan J. Liebowitz, *Copying and Indirect Appropriability: Photocopying of Journals*, 93 J. POL. ECON. 945 (1985) (arguing that original producers can price-discriminate so as to appropriate the value attributed by initial consumers to the ability to make subsequent copies); Lisa N. Takeyama, *The Intertemporal Consequences of Unauthorized Reproduction of Intellectual Property*, 40 J.L. & ECON. 511, 512–13 (1997) (arguing that imitators who saturate the low-end market allow high-end producers to credibly commit to higher-valuation “first-period” consumers that they will not subsequently sell to lower-valuation consumers at a lower price, thereby resolving the time-contingency obstacle to supracompetitive pricing).

14. See, e.g., Jonathan M. Barnett, *The Host's Dilemma: Strategic Forfeiture in Platform Markets for Informational Goods*, HARV. L. REV. (forthcoming 2011) [hereinafter Barnett, *Host's Dilemma*] (showing that platform holders in technology markets give away access or otherwise forfeit control in order to induce user investments that enhance the value of the platform, which in turn promotes the sale of complementary private goods).

15. See, e.g., Lisa N. Takeyama, *The Welfare Implications of Unauthorized Reproduction of Intellectual Property in the Presence of Demand Network Externalities*, 62 J. IND. ECON. 155, 165 (1994) (implying that network externalities motivate producers to give away samples in order to build an initial platform that increases demand for the product in the long-term). For similar arguments with respect to software publishers in particular, see Kathleen Reavis Conner & Richard P. Rumelt, *Software Piracy: An Analysis of Protection Strategies*, 37 MGMT. SCI.

Surprisingly, this recourse to property is not inconsistent with some of the most ardent expressions of the commons thesis. These lines of argument often refer to some reduced level of property-rights protection in order “to achieve balance,” but then fail to integrate this concession with the commons-styled claims that form the bulk of the remaining argument.¹⁶ Much of this Article is devoted to making explicit what is almost always implicit—or explicit but dismissively acknowledged—even in some of the strongest critiques of intellectual property. Namely, this Article identifies and describes the staying power of property rights or other exclusionary instruments in innovation markets that demand economically significant levels of investment. Property—understood broadly to refer to legal and any other form of exclusionary protection—is ubiquitous in markets for intangible goods. Recognizing and addressing directly this fact allows for construction of an integrated framework that accounts for both the staying power of “property” in markets characterized by widespread imitation and the staying power of “sharing” in markets characterized by robust innovation. Surprisingly, the same argument that establishes the inherent weakness of stand-alone sharing environments anticipates that sharing practices will persist and thrive within the secure perimeter established by property rights or other exclusionary instruments.

III. THEORY: THE FRAGILITY OF THE COMMONS

In this Part, I use a rational-choice framework to anticipate the conditions under which innovation incentives could be sustained in an environment largely bereft of intellectual property or other exclusionary

125, 126 (1991), and Ariel Katz, *A Network Effects Perspective on Software Piracy*, 55 U. TORONTO L.J. 155, 156–57 (2005).

16. For example, Professor Lessig, a leading skeptic of copyright, is careful to state that he seeks a “balance between free and controlled resources” in intellectual property law. See LESSIG, *FUTURE OF IDEAS*, *supra* note 8, at 72. Other commentators have observed, however, that this statement often jars with the relatively unqualified tenor of his critique against copyright in general. See, e.g., Sonia Katyal, *Ending the Revolution*, 80 TEX. L. REV. 1465, 1471–72 (2002) (reviewing Lessig’s *Future of Ideas*, noting that Lessig states that he maintains strong belief in private ownership but observing that this qualification is “slightly disingenuous” insofar as it is not reconciled with the general argument that copyright is unnecessary to support creativity); Julia D. Mahoney, *Lawrence Lessig’s Dystopian Vision*, 90 VA. L. REV. 2305, 2324 (2004) (reviewing Lessig’s *Free Culture*, noting that Lessig states that he is committed to “balance” in intellectual property but observing that he takes the view that peer-to-peer cooperative technologies should flatly trump intellectual property protections). In a recent book, Professor Lessig appears to adopt a more nuanced position, arguing that “sharing economies” and “commercial economies” can coexist in cultural markets. See LAWRENCE LESSIG, *REMIX: MAKING ART AND COMMERCE THRIVE IN THE HYBRID ECONOMY* 177–78 (2009).

barriers. Consistent with an economic approach, this framework assumes that participants in the innovation and commercialization process have rationally self-interested motivations: that is, effort will not be forthcoming unless it results in positive expected net benefits.¹⁷ To reconcile rational innovation with the absence of property rights or other access barriers, I design a hypothetical “sharing regime”¹⁸ that sustains innovation incentives through social norms that encourage original contributions and discourage excessive imitation. This norm-based mechanism is neither unique nor comprehensive; that is, it is neither the only model that could be formulated to sustain innovation without exclusionary barriers consistent with rational choice constraints, nor a model that encompasses all relevant variables.¹⁹ However, it may be viewed as a barebones heuristic to anticipate at a general level the conditions under which innovation can be feasibly maintained without barriers against imitation. Following the economics and political-

17. To be clear, I do not deny that altruistic or intrinsic motivations play some role in driving innovation, although this is immaterial where the firm, rather than an individual, is the operative decision maker. This is the almost universal case in technology markets and in cultural markets that require large capital investments to fund innovation or, more importantly, distribution. Moreover, by removing this factor from the analysis, we can assess how much “work” non-instrumentalist motivations would have to do to sustain innovative output. A more complex model of innovator behavior would incorporate both instrumentalist and non-instrumentalist motivations in the limited settings where that is likely to make a practical difference.

18. Alternative terms used in the literature are “semicommons,” a term recently gaining currency in the intellectual property literature, or “limited-access commons,” a more established term with a well-known valence in the political-science and economics literature on common-pool resource governance. Both terms denote fields of activity where there is open access to the relevant asset subject to (1) in the case of a “limited-access commons,” constraints imposed by community norms or other informal understandings, and (2) in the case of a “semicommons,” constraints imposed by limited applications of property rights. A “sharing regime” as used in this Article encompasses both terms insofar as it denotes both (1) open environments that operate subject to constraints imposed by norms and (2) closed environments that operate subject to constraints imposed by contract and intellectual property. For applications of the semicommons concept in the intellectual property context, see Michael J. Madison et al., *Constructing Commons in the Cultural Environment*, 95 CORNELL L. REV. 657 (2010); Henry E. Smith, *Governing the Tele-Semicommons*, 21 YALE J. ON REG. 289 (2005); Robert A. Heverly, *The Information Semicommons*, 18 BERKELEY TECH. L.J. 1127 (2003); and, for the first use of the term, Henry E. Smith, *Semicommon Property Rights and Scattering in the Open Fields*, 29 J. LEGAL STUD. 131 (2000). For a helpful survey of the various usages of these terms, see Lee Anne Fennell, *Commons, Anticommons, Semicommons*, in RESEARCH HANDBOOK ON THE ECONOMICS OF PROPERTY LAW 25 (Kenneth Ayotte & Henry E. Smith eds., forthcoming 2011).

19. Note that an obvious non-exclusionary alternative regime is a grant-based or other cash subsidy system funded by taxpayer contributions, which suffers from the informational inefficiencies inherent to any non-price-based allocation system, but avoids the deadweight losses inherent to imitations on access to non-rival goods. For sake of brevity, this option is excluded from consideration.

science literature on informal governance of “common pool” resources (as pioneered by 2009 Nobel Prize Winner Elinor Ostrom)²⁰ and the law-and-economics literature on “law and norms” (as pioneered by Professor Robert Ellickson, among others),²¹ the proposed regime replaces formal law, which coercively deters imitation, with informal norms that induce an approximately equivalent outcome. This exercise identifies circumstances where social norms can plausibly achieve collectively beneficial outcomes roughly equivalent to those that are achieved through legal sanctions. However, as is sometimes overlooked in the literature that references or applies the common-pool and law-and-norms approaches, it is important to keep in mind the limited circumstances under which social norms may replicate the outcome that would certainly be achieved by robust property rights. A norm-governed innovation regime is a locally effective—but low-cost—apparatus under a narrowly defined set of conditions. That offers a feasible but substantially imperfect substitute for its legal equivalent, which is a universally effective—but high-cost—apparatus under a broadly defined set of conditions.²²

A. REGIME STRUCTURES

Historians of science and technology generally agree that innovation is usually a cumulative process initiated by an original innovator, who contributes the initial major innovation, and then continued by subsequent innovators,²³ who contribute incremental improvements to the original innovation.²⁴ Collectively, these contributions constitute what I call the “innovation pool,” which may be construed as a stock of technological and/or creative inputs from which innovators can draw subject to any legal,

20. See ELINOR OSTROM, GOVERNING THE COMMONS: THE EVOLUTION OF INSTITUTIONS FOR COLLECTIVE ACTION (1990).

21. See ROBERT C. ELICKSON, ORDER WITHOUT LAW: HOW NEIGHBORS SETTLE DISPUTES (1991).

22. For some readers, the review of basic game-theoretic concepts (in particular, *infra* Section III.B) will be familiar; however, potential redundancy is sacrificed in order to describe precisely the formidable conditions that a sharing regime must satisfy in order to provide any plausible substitute for legal or other barriers against imitation.

23. The distinction between “original” and “subsequent” innovators is equivalent to other distinctions in the literature between “pioneers” and “improvers” or “first movers” and “second movers.”

24. See JOEL MOKYR, THE LEVER OF RICHES: TECHNOLOGICAL CREATIVITY AND ECONOMIC PROGRESS 12–13 (1990) (arguing that innovation often proceeds by introduction of a “macroinnovation,” which is then refined and developed by a long series of “microinnovations”).

technological, or other access limitations.²⁵ An innovation regime that maximizes cumulative innovative output over time must navigate the inherent tradeoff between supporting the incentives of original and subsequent innovators. First, it must sustain original innovators' incentives to make contributions to the innovation pool, which demands restrictions on access that increase transaction costs and input costs for subsequent innovators. Second, it must sustain subsequent innovators' incentives to generate derivative applications by making withdrawals from the innovation pool, which demands relaxations of access that reduce transaction costs and input costs for subsequent innovators.

Three broadly defined regimes can be instituted to govern contributions to and withdrawals from the innovation pool. Each regime achieves a different tradeoff between original and subsequent innovation incentives. These options (summarized below in Table 1) are as follows: (1) an open-access regime (commons), which imposes no withdrawal limitations and no contribution requirements; (2) a closed-access regime (property),²⁶ which imposes complete withdrawal limitations through legal or technological constraints but does not impose any contribution requirements; and (3) a limited-access regime (sharing), which uses norm-based instruments to impose contribution requirements and substantially incomplete withdrawal limitations. At one extreme, a commons environment eliminates all access restrictions and the attendant cost burden on subsequent innovation, but does not protect any portion of the innovation pool. This eliminates all incentives for original innovation, so that it can be set aside as a feasible solution to the under-innovation problem. At the other extreme, a property regime contemplates no unprotected portions of the innovation pool. This supports first-mover innovation but does so by imposing the cost burdens

25. This concept is inspired by the empirical literature on informal governance of common pool resources, which describes informal governance structures for renewable resource pools that are otherwise subject to individually rational overuse leading to a collective loss in the form of resource depletion. For the leading source, see OSTROM, *supra* note 20. These governance structures seek to avoid resource depletion by regulating individual usage over time so as to ensure that the average "withdrawal rate" does not exceed the average "replenishment rate," but without setting overly strict limitations that fail to maximize the pool's economic yield. While the analogy to a renewable resource pool is imperfect given the inexhaustibility of intellectual assets, it is applicable to the extent that, absent any limitations on the surplus of withdrawals over contributions from the collective innovation pool, innovators will be unable to accrue reputational (and collateral financial) returns, thereby precipitating an under-innovation outcome.

26. Note that a more exact term for "property regime" would be "proprietary regime" as I mean to include any regime that relies on legal *or* extralegal barriers to restrain imitation. However, the "property/commons" dichotomy is well-established in the literature so I avoid multiplying terms.

that attend a formal property-rights system, which discourages subsequent innovation.²⁷ Between these two polar alternatives lies a rich variety of intermediate sharing regimes, each of which protects some but not all portions of the innovation pool, thereby enhancing original innovation incentives relative to an open-access regime but without fully incurring the transaction and other costs that can burden subsequent innovation under a closed-access regime.

Table 1: Regime Comparison

Regime	Regulatory Instrument	Contribution Requirements	Withdrawal Limitations
Commons	None	No	No
Sharing	Norms	Yes	Yes, but incomplete
Property	Law	No	Yes

To induce innovative output without recourse to the costly apparatus of property rights or other exclusionary protections, a sharing regime must implement two social norms: (1) a contribution norm, which mandates that innovators make a certain minimum level of original contributions to the innovation pool, which is then freely accessible, and (2) a withdrawal norm, which sets a maximum limit to withdrawals by subsequent innovators from the innovation pool (or, in its weaker form, an attribution norm that allows unconstrained withdrawals but requires that subsequent innovators give credit to original contributors.)²⁸ Assuming sufficient compliance among the innovator population (the conditions for which are elaborated in the next Section), these contribution and withdrawal norms together implement a modified reciprocity principle that sustains innovation even in the absence of legal or other exclusionary barriers. Each innovator rationally makes original contributions to the common pool with two expectations. First, given general compliance with the contribution norm, it will withdraw from the

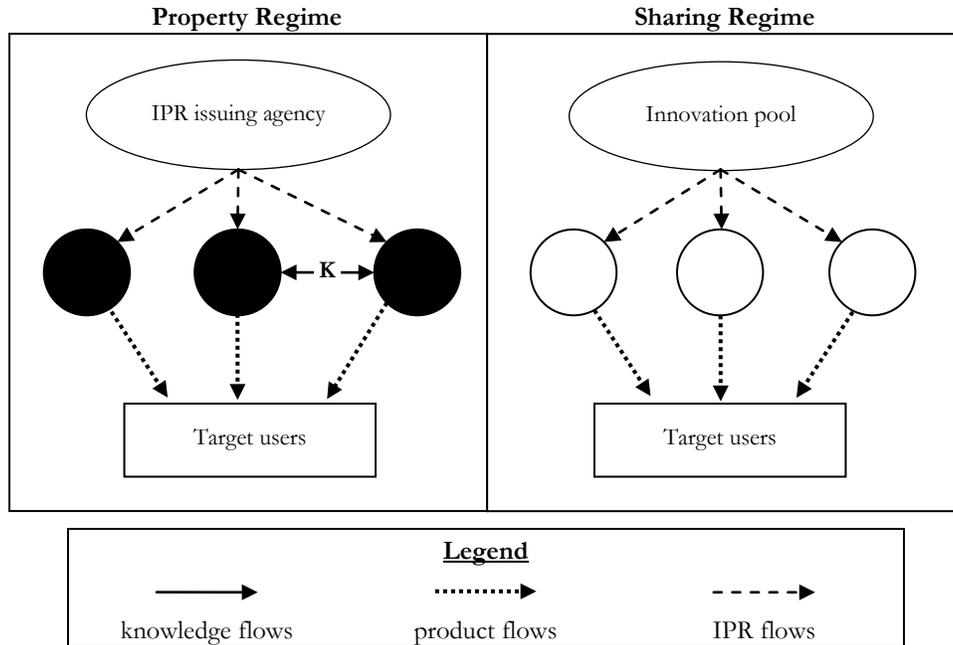
27. It may be argued that this result would not work because the first-mover innovator will rationally make its intellectual goods available to lower-cost *n*-mover innovators in order to generate derivative applications for mutual profit. This type of claim (which is equivalent to the “prospect” theory of patent rights proposed by Edmund W. Kitch, *The Nature and Function of the Patent System*, 20 J.L. & ECON. 265 (1977)) is subject to the objection that efficient contracting may be frustrated by strategic behavior, transaction costs in identifying and negotiating with follow-on innovators, and other inefficiencies associated with protected market positions. For further analysis of these difficulties, see Suzanne Scotchmer, *Standing on the Shoulders of Giants: Cumulative Research and the Patent Law*, 5 J. ECON. PERSP. 29 (1991) (discussing how to divide joint profit among innovators when one innovator’s technology builds on another’s).

28. This is the prevailing norm in academic research, *see infra* Section V.B.

pool over time roughly the same value as it contributes to it. Second, given general compliance with the withdrawal norm, there will exist some positive distance between original contributions and derivative applications, thereby precluding perfect substitution that would otherwise prevent the former from earning any premium over the latter²⁹ (in which case the standard under-innovation result would prevail). Together these two norms provide innovators with substantially unimpeded access to the innovation pool, thereby reducing total innovation costs, while imposing some limitations on withdrawals from the pool, thereby generating remunerative streams that sustain incentives to make further contributions to the pool.

Law-based property regimes and norm-based sharing regimes generate starkly different transaction structures for the generation, transmission and exchange of intellectual goods, as rendered graphically in Figure 1 below. Under a property regime, unauthorized uses of intellectual goods are punished at a high cost by legal sanctions enforced through formal dispute-resolution processes based on intellectual property rights issued by a state agency. This formal infrastructure generates a closed exchange pattern consisting of an atomized sequence of high transaction-cost negotiated transfers of intellectual goods (as denoted by K in Figure 1 below), each of which is held on an exclusive basis by an entitlement holder. Under a sharing regime, excessive withdrawals from the innovation pool and failure to make original contributions to the innovation pool are punished at low cost through business, reputational, and other social sanctions assessed by the market. This informal infrastructure generates an open exchange pattern consisting of a continuous flow of low transaction-cost non-negotiated transfers of intellectual goods, none of which is held on an exclusive basis (at least in its entirety) by any party.

29. That is: if the imitation is a perfect economic substitute for the original, then the original innovator will be unable to demand any price above marginal cost, which in turn will prevent recoupment of the fixed costs of research and development, all of which are borne by the original innovator.

Figure 1: Alternative Regimes (Pure Form)³⁰

Assuming the existence of an enforcement technology to elicit contributions to, and restrain withdrawals from, the innovation pool, a sharing regime would appear to constitute a collectively preferred arrangement that outperforms both commons and property regimes. First, the sharing regime outperforms a commons regime by securing substantial returns for first-mover innovators. Second, it outperforms a property regime by reducing the transaction costs and input costs borne by subsequent innovators. This abstract characterization translates into economic terms the strong attraction the commons model exerts over popular discourse and a good deal of scholarly commentary. High appropriation capacities combined with low transaction costs are obviously preferable to the high appropriation capacities combined with high transaction costs of a formal property regime. It now remains to identify the conditions under which this is a feasible alternative. As we shall see, those conditions are not easily satisfied.

30. As used in Figure 1 (and subsequent Figures), a blank circle refers to an “open” innovator that participates in a nominal-cost exchange of intellectual assets (i.e., a sharing arrangement); a darkened circle refers to a “closed” innovator who does not. “K” refers to any type of contractual arrangement. “IPR” refers to intellectual property right.

B. THE COOPERATION GAMBLE

From the skeptical perspective of a rational-choice observer, a sharing regime is an academic exercise in naive utopianism. In the absence of any credible detection and enforcement mechanism, no innovator has any incentive to incur the costs required to comply with the underlying reciprocity norms, in which case property rights are restored as the unique solution to under-innovation. This can be elaborated by applying the well-known logic of the “prisoner’s dilemma” game. Suppose a market occupied by two innovators, each of whom must release an innovation for a new product season. Each innovator can elect between two actions: *cooperate* (i.e., comply with sharing norms, resulting in innovation), which results in an original contribution being developed at great cost, or *defect* (i.e., not comply, resulting in imitation), which immediately replicates at little cost any original contribution made by the other innovator. The pathological result is predictable. While it may be collectively rational over the course of multiple seasons if both innovators elected *cooperate*, thereby resulting in a rich stock of technological and creative inputs to support further innovation, it is individually rational in any individual season for each innovator to elect *defect*, thereby capturing the gains from the innovation without incurring the associated development costs. If these innovators are unable to make a credible commitment to elect *cooperate*, then each innovator elects *defect*, resulting in a “waiting game” that yields zero innovation.

But it is well known that this dilemma is not without a possible solution. So long as innovators are repeat players with sufficiently low discount rates and interact over an indefinitely repeated sequence, each may rationally cooperate (i.e., innovate). Any innovator will cooperate if it expects that discounted future gains in the event of mutual cooperation exceed “one-shot” gains, less discounted future losses from a single defection (and so long as the anticipated losses from “incorrectly” electing *cooperate* in any single round are not too great).³¹ Hence, even without the coercive force of law, a repeat-player innovator will sometimes “gamble” by electing *cooperate*—without perfect assurance that the other innovator will do the same. But this

31. The game theory literature has developed multiple equilibrium strategies in the repeated prisoner’s dilemma game. See JEAN TIROLE, *THE THEORY OF INDUSTRIAL ORGANIZATION* 245–47, 258–59 (1988). Perhaps the most well-known strategy is “Tit for Tat,” which requires that a player elect *cooperate* in the initial round of an iterated sequence, and in each round thereafter, but then to revert to *defect* if the other player elects *defect*. This “cooperative” equilibrium has the technical shortcoming such that (unlike the mutual defection outcome in a one-shot prisoner’s dilemma) it cannot be identified as the unique equilibrium; however, it does describe a possible equilibrium under certain reasonable assumptions.

solution has an important limitation. While long-term cooperation gains can induce forfeiture of short-term defection gains in an indefinitely repeated sequence of two-player interactions, this does not easily follow in multi-player settings where no individual election to forego short-term defection gains can independently determine whether a cooperative equilibrium will result. The result: every innovator rationally elects *defect* (i.e., waits to imitate),³² thereby restoring the collectively undesirable outcome of zero innovation.

This problem too is not without a tenable solution. So long as there exists an external instrument that sufficiently adjusts upward and downward, respectively, the relative expected payoffs of cooperation (innovation) and defection (imitation), then the “cooperation gamble” is restored as a rational choice. The anticipated breakdown of cooperative behavior in large-number settings has been addressed extensively through supplemental material benefits—what the public goods literature calls “selective incentives.”³³ These include reputational rewards and penalties, which fill the incentive gap that would otherwise result in individually rational defection. This can be illustrated by two well-studied “good” and “bad” solutions to collective-action failure. First, in industrial cartels, cheating on collectively beneficial output constraints is chronic, compelling cartels to invest in monitoring and punishment mechanisms to achieve sufficient levels of compliance to maintain collusive pricing.³⁴ Second, in informal governance of common-pool resources, monitoring and punishment mechanisms are almost universally used to support norm-based restraints on excessive withdrawals

32. There is a technical exception to this statement: even under the assumptions stated above, cooperation may still be individually rational where an individual's marginal contribution independently determines the total amount of the collective good that is provided (the so-called “weakest-link” scenario). This may have practical importance in some contexts. For further discussion, see Elinor Ostrom, *How Types of Goods and Property Rights Jointly Affect Collective Action*, 15 J. THEORETICAL POLITICS 239, 247–48 (2003) [hereinafter Ostrom, *Types of Goods*]; RICHARD CORNES & TODD SANDLER, *THEORY OF EXTERNALITIES, PUBLIC GOODS AND CLUB GOODS* ch. 2 (1986) (applying concepts from public finance).

33. See MANCUR OLSON, JR., *THE LOGIC OF COLLECTIVE ACTION: PUBLIC GOODS AND THE THEORY OF GROUPS* 60–65 (1965). For more extensive and updated discussions of Olson's thesis and the vast theoretical and empirical literature that it has spawned, see CORNES & SANDLER, *supra* note 32; TODD SANDLER, *COLLECTIVE ACTION: THEORY AND APPLICATIONS* (1992).

34. See 1 Alexis Jacquemin & Margaret E. Slade, *Cartels, Collusion, and Horizontal Merger*, in *HANDBOOK OF INDUSTRIAL ORGANIZATION* ch. 7 (Richard Schmalensee & Robert D. Willig eds., 1989).

from the relevant resource pool.³⁵ The same mechanism can operate in the innovation context: a sharing regime can allocate reputational rewards and penalties in order to shift relative cooperation and defection payoffs and fill the incentive gap that would otherwise result in individually rational defection from the governing contribution and withdrawal norms. Reputational rewards for original contributions and reputational penalties for excessive withdrawals, plus any collateral monetary or other material rewards and penalties, eliminate the individually rational “temptation to defect” and drive a repeat-player innovator to conclude that electing *cooperate* will maximize long-term payoffs (even if there is some positive likelihood that some other players will elect *defect*). The result: the innovator rationally complies with norm-based constraints on imitative behavior, and norm-based requirements to make original contributions, even in the absence of any legal obligation to do so.³⁶

C. MAKING COOPERATION LAST

So far I have identified two minimal conditions for a viable sharing regime in any large-number setting: (1) innovators must be repeat players with sufficiently low discount rates, and (2) a reputation-based enforcement technology must exist that sufficiently rewards compliance with, and penalizes violations of, the governing reciprocity norms. Those conditions may not appear especially strenuous and, as will be seen in the subsequent empirical discussion, roughly characterize a number of cultural, research, and design markets that rely substantially on reputational norms in order to induce innovative effort. But it is important to distinguish between viability and stability. Even if the minimal viability conditions are met, a norm-based sharing regime may still be vulnerable to individually rational defections, which in turn can spawn generalized defection among the innovator population that ultimately undermines the cooperative outcome. This Section

35. In the common-pool setting, monitoring and punishment mechanisms are almost universally used to support norm-based restraints on excessive withdrawals from the relevant resource pool. *See* OSTROM, *supra* note 20.

36. For completeness, it is necessary to address another potential difficulty. Even if reputational instruments could sufficiently correct any first-order incentive problem, this enforcement technology falls prey to a second-order incentive problem insofar as it too requires individually irrational expenditures to monitor norm-compliance and allocate reputational sanctions and rewards. As a practical matter, this problem may be mitigated in markets where the reputational infrastructure is administered (1) at relatively little cost to any individual, (2) by the immediate victim of any norm-violation (e.g., failure to attribute) or third-party participants with an independent profit-based incentive to do so, and/or (3) by collective organizations that spread the costs of norm-enforcement over a wide pool of individual beneficiaries, each of whom must then incur no more than a small contribution cost.

identifies a set of “stability conditions” that can enhance or decrease the likelihood that individually rational participants will voluntarily adhere to the reciprocity principle without which a sharing regime necessarily fails. Broadly consistent with the core findings of economic research on the private provision of public goods, this exercise suggests that a sharing regime is most likely to persist in “intimate” environments that exhibit the following features: (1) group size is small, (2) required capital investment is low, (3) innovative output has low economic values, and, what will be argued is a factor of special importance, and (4) innovative endowments (that is, innovators’ capacities and talents) are roughly equivalent in value.³⁷ To the extent one or more of these conditions are not substantially satisfied, a sharing regime becomes unstable. Innovators are likely to take defection actions that precipitate either a commons regime, resulting in the standard under-innovation outcome, or a property regime, which preserves some innovation under a higher transaction-cost burden.

1. *Innovator Options and Types*

Formerly I had assumed that an innovator could elect only among two possible actions: (1) *cooperate*, in the form of making contributions to, and constraining withdrawals from, the pool; and (2) *defect*, in the form of ceasing contributions to, and making unconstrained withdrawals from, the pool. To facilitate a more complex analysis, I will now expand the innovator’s choice set so that it includes two defection options: (1) *defect(copy)* (equivalent to the *defect* option set forth above), and (2) *defect(property)*, which refers to lobbying for, adopting, and enforcing state-provided property entitlements or, more typically, enforcing formally available but dormant property entitlements that have generally been unused. That is, an innovator can incur some cost, L , in order to “activate” a property entitlement through some combination of the aforementioned actions.³⁸ That in turn bars unauthorized access to the relevant intellectual good, which permits innovation to proceed at some positive level but under the higher transaction-cost burden associated with the maintenance and enforcement of formal property rights. To give a concrete example: star scientists can (and do) defect out of the sharing norms

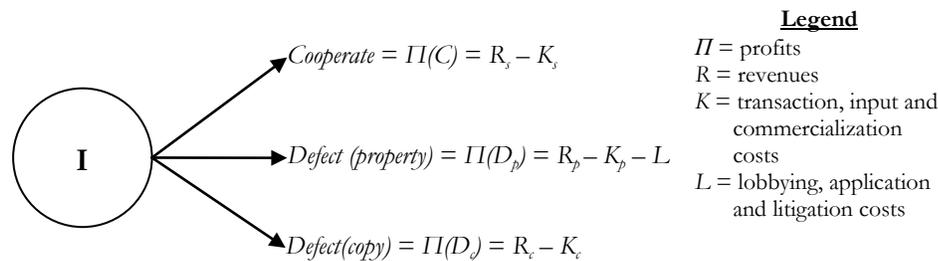
37. I emphasize the qualifier, “broadly consistent.” The vast theoretical and empirical literature on private provision of public goods is complex and not easily subject to generalization. However, the analysis below relies on the core findings of that literature with respect to each of the enumerated characteristics, noting any important open or disputed points where relevant.

38. More generally, *defect(property)* would include implementing technological measures that limit access. For simplicity of presentation, I focus on law-based access limitations in the form of property rights.

that have historically characterized academic research by accessing the state-provided patent system, which in turn can support financing and operating a for-profit business based on the legally protected technology.³⁹

The Figure below sets out the corresponding actions and payoffs facing any innovator (denoted below as “I”). Note that each defection action is individually rational for different reasons. *Defect(copy)* enables the innovator to avoid contribution costs and capture revenues from any party who elects *cooperate*. *Defect(property)* enables the innovator to block capture of its revenues by any party who elects *defect(copy)* (or, to a lesser extent, any party that elects *cooperate*). Using the notation below, we can state a simple condition for rational cooperation: it must be the case that $\Pi(D_p) < \Pi(C) > \Pi(D_c)$. If either of these inequalities is not satisfied, then the innovator will elect either *defect(copy)* ($\Pi(C) < \Pi(D_c)$) or *defect(property)* ($\Pi(D_p) > \Pi(C)$), respectively.⁴⁰

Figure 2: Innovator Actions and Payoffs



If we add one more feature to this setting, we can anticipate these defection actions based on innovator type. Suppose some distribution of “innovation endowments”—that is, the value of innovation talents and capacities—over a general innovator population. Now distinguish between two innovator types: (1) “strong” innovators who have an innovation endowment that is superior to the average endowment of the innovator population, and (2) “weak” innovators who have an innovation endowment that is inferior to the average endowment. An innovator’s endowment determines the costs it must incur in order to generate a given unit of

39. For a review of the empirical literature on this phenomenon and additional results, see Pierre Azoulay, Waverly Ding & Toby Stuart, *The Determinants of Faculty Patenting Behavior: Demographics or Opportunities?*, 63 J. ECON. BEHAV. & ORG. 559 (2007) (finding that patenting events are preceded by a flurry of publications).

40. An innovator may also elect *defect(withdraw)*, in the form of re-allocating investment resources to another use entirely, resulting in a payoff equal to $\Pi(w)$, where $\Pi(w) = R_w - K_w$. For ease of exposition, this option is generally not addressed above; for further additional discussion of this option, see *infra* note 44. Note that the various subscripts, “s,” “p,” and “c” refer, respectively, to the payoffs corresponding to an innovator’s election to *cooperate*, *defect(property)*, and *defect(copy)*.

innovative output. The stronger an innovator's endowment, the lower the costs it must incur to generate any given unit of innovative output. Conversely, the lower an innovator's endowment, the higher the costs it must incur to generate that same unit of innovative output. The relative strength or weakness of an innovator's endowment then implies its defection choice. For a weak innovator, any sufficient decrease in the reputational penalty for excessive withdrawals from the innovation pool (equivalent to an increase in the defection payoff) will induce it to elect *defect(copy)*: it incurs higher innovation costs relative to all other innovators and therefore can best compete as a copyist who avoids those costs almost entirely. That is, $\Pi(D_c) > \Pi(C)$ (and $\Pi(D_c) > \Pi(D_p)$). For a strong innovator, any sufficient decrease in the cooperation payoff (plus any reputational or other side payments) will induce it to elect *defect(property)*: it incurs lower innovation costs relative to all other innovators and therefore can best compete as an original innovator who can supply a higher quantity of innovative output at the lowest cost. That is, $\Pi(D_p) > \Pi(C)$ (and $\Pi(D_p) > \Pi(D_c)$).

2. Conditions for Cooperation

We must now identify the conditions under which a sharing regime is most likely to induce innovators to elect *cooperate* over the alternative options: *defect(copy)* or *defect(property)*. Precisely, what are the conditions under which an innovator will or will not expect that $\Pi(D_p) < \Pi(C) > \Pi(D_c)$? Existing research on the private provision of public goods, and the related literature on cooperative behavior in common-pool resource settings, identifies a number of factors that can influence individually rational incentives to comply with the norm-based constraints of a sharing regime.⁴¹ Some of the leading factors include: group size, capital intensity, economic value, and, of special interest in the ensuing discussion, endowment heterogeneity.⁴² Generally speaking, as explained in detail below, we can anticipate an inverse relationship between the value of each of these variables and innovators' propensity to elect cooperate over either of the two defection options. As

41. For a general overview of the factors that influence private provision of public goods, see CORNES & SANDLER, *supra* note 32. For an application of that literature to the common-pool resource context, see OSTROM, *supra* note 20.

42. The public goods literature and the common-pool governance literature, as well as the related literature on cartel stability, consider the effects of heterogeneity along several dimensions: resources, endowments, interests, preferences and costs, among others. I refer solely to heterogeneity in the value of innovators' talents and capacities. Note further that endowment heterogeneity is a function of the comparative value of participants' innovation assets or capacities; it is not a function of the identity of those assets or capacities. Hence, there will be a high level of endowment homogeneity where participants have different innovation assets or capacities that complement each other but have roughly equal values.

any given innovation environment exhibits increasing group size, capital intensity, economic values, and/or endowment heterogeneity, the defection payoff rises relative to the cooperation payoff. As one or more of those variables declines in value, the cooperation payoff rises relative to the defection payoff. In each case, we can then anticipate an innovator's defection choice based on its innovator type: weak innovators tend to elect *defect(copy)*; strong innovators tend to elect *defect(property)*.

The proposed relationships are summarized in Table 2 below and then explained in detail in the discussion that follows.

Table 2: Defection Actions as a Function of Innovator Type

Variable (Increasing)	Weak Innovator	Strong Innovator
Group Size	<i>Defect(copy)</i>	<i>Defect(property)</i>
Capital Investment	<i>Defect(copy)</i>	<i>Defect(property)</i>
Asset Value	<i>Defect(copy)</i>	<i>Defect(property)</i>
Endowment Heterogeneity	<i>Defect(copy)</i>	<i>Defect(property)</i>

a) Group Size

Any increase in the size of the innovator population erodes cooperation incentives for two reasons. First, it increases monitoring costs, thereby reducing the ability to punish defection with reputational sanctions and credit cooperation with reputational rewards, which effectively lowers the cooperation payoff. Second, it dilutes the individual share of collective benefits (assuming those are held fixed) that would be accrued under a sharing regime, which effectively lowers the cooperation payoff. A weak innovator will then elect *defect(copy)* in order to capture gains from stronger innovators, while a strong innovator will elect *defect(property)* in order to protect gains against weaker innovators. For both innovator types, the cooperation gamble becomes imprudent as the number of innovators sufficiently increases.

b) Capital Intensity

Suppose there is a capital-intensive innovation market that necessitates development and other "bringing to market" costs that are large relative to imitation costs borne by third parties. That means that both (1) the gains that would accrue to a party who elects *defect(copy)* and (2) the losses that would be incurred by an innovator who "incorrectly" elects *cooperate* are substantial. This both increases the payoff under *defect(copy)* and reduces the cooperation payoff. Without property rights or some other instrument by which to block imitation, large disparities between innovation costs incurred by firms that elect *cooperate* and imitation costs incurred by firms that elect *defect(copy)* imply that few if any firms will rationally make the former election. But this does

not mean that all firms will elect *defect(copy)*. While a weak innovator will elect *defect(copy)* in order to save on its disproportionately high innovation costs, a strong innovator will elect *defect(property)* in order to exploit its disproportionately low innovation costs and earn a net positive return. For both innovator types, the cooperation gamble becomes imprudent when the required capital investment is highest—or, most precisely, when the required innovation investment is substantially higher than the required imitation investment.

c) Asset Values

Strong innovators will have enhanced incentives to elect *defect(property)* where the market value of the relevant innovation in any given iteration is unusually high.⁴³ Compliance with sharing norms mandates forfeiture of a portion of that market value in any given iteration to other innovators in order to accrue long-term cooperation gains. While that long-term calculus may usually drive a repeat-player innovator to comply with the sharing norms, that may not be the case with respect to a “blockbuster” innovation for which an especially high one-time payoff could be earned if the innovator retained it exclusively by electing *defect(property)*. For the same reason, weak innovators have powerful incentives to elect *defect(copy)* when the economic value of the innovations generated by stronger innovators is especially high. For both innovator types, the cooperation gamble becomes imprudent when innovations are most valuable.

d) Endowment Heterogeneity

Relative differences in innovation endowments can predict the “direction” of an innovator’s defection in response to changes in group size, capital intensity, and economic values. But relative differences in innovation endowments can directly trigger individual defections from the sharing norms. Recall that a sharing regime (1) imposes a contribution requirement that mandates that each member contribute a certain minimum value to the innovation pool (which constitute “cooperation costs”) and (2) allows all members to withdraw up to a certain maximum value from the innovation pool (which constitute “cooperation gains”). If cooperation gains are not calibrated to reflect idiosyncratically higher or lower cooperation costs, then the cooperation gamble becomes imprudent for both strong and weak

43. Weak innovators may strategically elect *defect(property)* so as to obtain dubious property rights over valuable but unclaimed technologies and then extract nuisance settlements from stronger innovators. This corresponds to the “patent troll” phenomenon where non-operating patent holders allegedly sue technology companies opportunistically in order to extract cash settlements. For simplicity, I ignore this contingency.

innovators. A weak innovator incurs higher cooperation costs in light of its inferior innovation capacities: relative to all other innovators, it must expend greater resources to meet the contribution norm. As a result, a weak innovator is likely to anticipate that the payoff under *defect(copy)* exceeds the payoff under *cooperate*, resulting in imitation. A strong innovator incurs lower cooperation costs in light of its superior innovation capacities: relative to all other innovators, it must expend fewer resources to meet the contribution norm. However, that same fact implies that a strong innovator incurs higher indirect cooperation costs in light of the foregone profits that it could earn on the “open market” by exploiting its superior innovative capacities. Absent reputational side payments to reflect a strong innovator’s exceptional contributions, it is likely to anticipate that the payoff under *defect(property)* exceeds the cooperation payoff.⁴⁴ This positive relationship between endowment homogeneity and contribution incentives is consistent with evidence in a striking variety of contexts, including (1) cartels,⁴⁵ (2) common-pool resource settings,⁴⁶ (3) information sharing within organizations,⁴⁷ and (4) experimental simulations of public goods scenarios.⁴⁸

44. This assumes a full choice set. Even if the *defect(property)* option is not available (due to the absence of any legal or technological instrument by which to establish exclusivity), the strong innovator will elect *defect(withdraw)* in partial form. That is: it will constrain its innovative effort such that it meets the minimum contribution requirement but cease making further contributions to the innovation pool given the inability to earn returns that reflect its higher-value contribution. The result: a sharing regime preserves access at the cost of suppressing the highest-value forms of innovative output.

45. See Jacquemin & Slade, *supra* note 34, at 417–30 (reviewing studies showing that cost and product homogeneity promote cartel stability).

46. See generally Thrainn Eggertsson, *Open Access versus Common Property*, in PROPERTY RIGHTS: COOPERATION, CONFLICT AND LAW (Terry L. Anderson & Fred S. McChesney eds., 2002) (reviewing studies showing that endowment homogeneity tends to facilitate cooperative solutions to common-pool resource depletion); Jean-Philippe Platteau, *Solidarity Norms and Institutions in Village Societies: Static and Dynamic Considerations*, in THE HANDBOOK OF THE ECONOMICS OF GIVING, ALTRUISM AND RECIPROCITY (Serge-Christopher Kolm & Jean Mercier Ythier eds., 2006). For further discussion, see Ostrom, *Types of Goods*, *supra* note 32, at 257–58 (describing some limited diversity of results).

47. See generally Brian K. Thorn & Terry Connolly, *Discretionary Data Bases: A Theory and Some Experimental Findings*, 14 COMM. RES. 512 (1987) (finding that sharing of information among organization’s employees tends to decline as asymmetries in information values and benefits increase across participants).

48. These experiments tend to find that private contributions decrease as endowment homogeneity decreases, and increase as endowment homogeneity increases. See John O. Ledyard, *Public Goods: A Survey of Experimental Research*, in HANDBOOK ON EXPERIMENTAL ECONOMICS 111, 158–60 (John Kagel & Alvin Roth eds., 1995); see also Steven Hackett et al., *The Role of Communication in Resolving Commons Dilemmas: Experimental Evidence with Heterogeneous Appropriators*, 27 J. ENVTL. ECON. & MGMT. 99 (1994) (finding that in *n*-person commons dilemmas, endowment heterogeneity reduces earnings and is associated with a reduced ability to agree on allocation rules). The theoretical public-goods literature observes that the

3. *From Individual to Collective Defection*

So far I have described the conditions under which a sharing regime is most likely to induce individual innovators to elect *cooperate* over *defect(copy)* or *defect(property)* and, by implication, the conditions under which they are least likely to do so. But individual defections are only of interest to the extent that they translate into collective defections from the sharing norm among substantial portions of the innovator population. That unraveling effect follows logically from the prisoner's dilemma described at the beginning of this Part. Even an isolated defection by a single innovator may threaten the stability of a sharing regime by triggering a sequence of defections that ultimately results in widespread or universal defection from the sharing regime. As illustrated in Figure 3 below, individual defections by strong and weak innovator types can give rise to two collective defection scenarios, which in turn yields corresponding shifts in the governing innovation regime. First, if a weak innovator elects *defect(copy)*, that erodes the premium accruing to original contributions and may compel a stronger innovator to select *defect(copy)*, which in turn induces a further reduction in the innovation premium, further contraction of the innovation pool, and further elections of *defect(copy)*, ultimately resulting in the standard under-innovation result. Second, if a strong innovator elects *defect(property)*, that may compel other strong innovators to elect *defect(property)* in order to protect against actual or expected increases in litigation risk, transaction costs, and input costs. That decision in turn induces further elections of *defect(property)*,⁴⁹ resulting in widespread implementation of a formal property regime, even if a sharing regime constitutes the collectively preferred outcome.⁵⁰ Interestingly, both

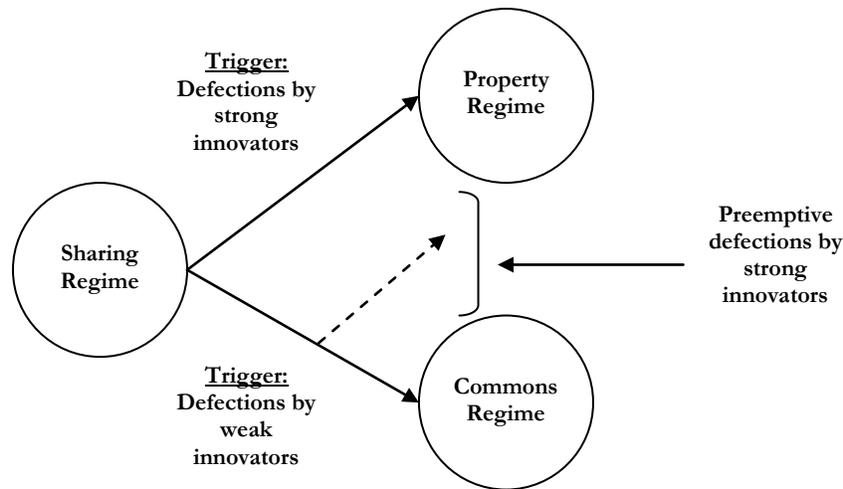
effect of heterogeneity (and group size, to the extent that heterogeneity is a positive function of group size) on private provision of public goods can be ambiguous. Specifically, under certain conditions, endowment heterogeneity can increase contribution rates where there is an increased probability that there exist extreme types who have sufficient interest and resources to unilaterally contribute to the public good independently of whether or not other contributors are doing so. See CORNES & SANDLER, *supra* note 32, at 325; Pamela E. Oliver & Gerald Marwell, *The Paradox of Group Size in Collective Action: A Theory of the Critical Mass*, 53 AM. SOC. REV. 1 (1988); Ostrom, *Types of Goods*, *supra* note 32, at 257–58. Note that this argument assumes that contributors cannot take actions to exclude non-contributing third parties from enjoying the relevant public good (i.e., cannot “convert” the public good into a private good). By contrast, the analysis above envisions that innovators can do so at some positive likelihood and some non-exorbitant cost by “activating” state-provided property entitlements (i.e., by electing *defect(property)* using the terminology introduced above), in which case substantial endowment heterogeneity can never be conducive to a high-endowment firm's incentives to contribute without making recourse to property rights.

49. For further discussion of this scenario, see Barnett, *Property as Process*, *supra* note 3.

50. Both results are sensitive to the proportions of weak, strong and other innovators in the industry. Even a substantial segment of weak innovators who elect *defect(copy)* does not

scenarios will yield the same outcome where innovators have access to a full choice set that includes *defect(property)* at some reasonable cost. Strong innovators can then preempt any underinnovation outcome, which would be triggered by *defect(copy)* elections by weak innovators. This can be accomplished by lobbying for property rights, adopting and enforcing dormant property rights, and/or adopting some other exclusionary technology. This effectively removes the *defect(copy)* option from the available choice set. Hence, as a practical matter, the three-way choice between a sharing regime, property regime, and commons regime may sometimes reduce to a two-way choice between a sharing regime and a property regime.

Figure 3: Collective Defections and Regime Shifts



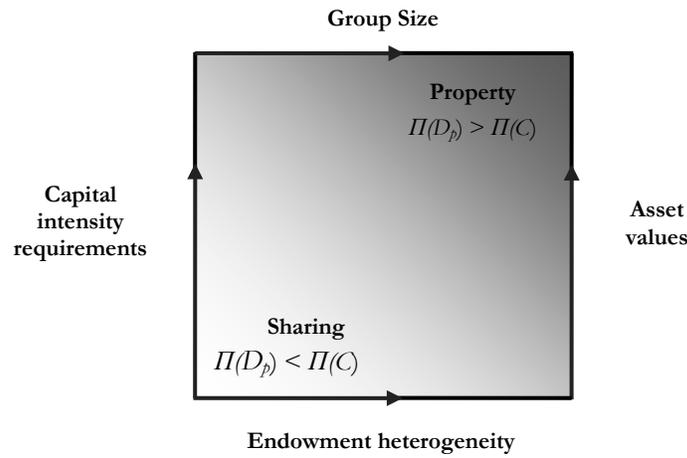
4. Predicting Cooperation

We can now consolidate this discussion into a single framework that roughly anticipates the conditions under which a sharing regime is likely to provide a stable alternative to either a commons regime or a property regime. If we assume that strong innovators can block formation of a commons regime by electing *defect(property)*, this task reduces to assessing the stability of a sharing regime relative to a property regime. Generally speaking, we can

necessarily unravel a sharing regime since the sub-population of strong innovators may rationally maximize gains by incurring the costs of complying with the governing cooperation norms, even in the face of individually rational defections by some other innovators. In the more formal language of the public-goods literature, the strong innovator population is a “privileged” or “viable” coalition playing an n -person repeated prisoner’s dilemma game where the harm caused by a single violation by any individual player or sufficiently small group of players is diffused over a broad population and may therefore be insufficient to cause the *defect(property)* payoff to exceed the cooperation payoff.

expect that the stability of a sharing regime will decline as one or more of the identified variables—group size, capital intensity, economic values, and endowment heterogeneity—increases in value. Conversely, where one or more of these same variables declines in value, a sharing regime tends to become a feasible alternative. Figure 4 below summarizes graphically the proposed impact of these variables on the stability of a sharing regime.

Figure 4: Regime Determinants⁵¹



The box diagram reflects the following pair of hypotheses, which will drive discussion in the remainder of this Article.

Hypothesis 1. The most highly developed sharing arrangements, and the least developed property-rights arrangements, should exist in environments characterized by low capital investment, low economic values, and a concentrated group of relatively few (or multiple but well-organized), and substantially similar firms or other participants (denoted by the “southwest” region in Figure 4).

Hypothesis 2. The least developed sharing arrangements, and the most developed property-rights arrangements, should exist in environments characterized by high capital investment, high economic values, and a dispersed group of multiple heterogeneous firms or other participants (denoted by the “northeast” region in Figure 4).

51. For ease of exposition, this diagram assumes linear relationships between these variables; however, there is no inherent reason to believe this would be the case in any particular instance. Endowment heterogeneity may have a much stronger effect than number of innovators on the cooperation payoff relative to the defection payoff, or vice versa, in which case the “box” would be replaced by a figure drawn with substantially different proportions. Increasing coloration denotes increasing use of practices indicative of a property regime, and vice versa.

But observe the “wide open” middle of Figure 4⁵²: this designates a broad intermediate region where the market will not support undiluted property and sharing regimes. That implies two conclusions. First, a “pure” sharing regime bereft of exclusionary protections is an unexpected occurrence outside of non-capital-intensive markets that meet certain parameter conditions. Second, a “pure” property regime bereft of sharing practices is an unexpected occurrence outside of capital-intensive markets that meet certain parameter conditions. Together those two points reduce to a *third* hypothesis that this Article will pursue as it moves from hypothetical to actual sharing regimes.

Hypothesis 3. Innovation markets will typically operate subject to a mixed-form sharing regime where low-cost knowledge-exchange practices operate with property rights or other exclusionary instruments to secure innovation returns while minimizing the associated transaction-cost burdens on innovation.

IV. EMPIRICS: THE COMPLEXITY OF THE COMMONS

A theory is only as good as its ability to account for the facts it sets out to explain. I have not proposed the hypothetical sharing regime in order to identify a universally valid set of conditions under which innovation can be sustained without exclusionary barriers. That would be a fool’s errand. As a vast experimental and empirical literature can attest, there is no one-size-fits-all solution to the collective-action failure that threatens adequate provision of public goods in a wide variety of settings.⁵³ Hence, the hypothetical sharing regime is only a useful construct if it provides a tool by which to anticipate and account for actual conditions under which innovation is likely (and not likely) in typical circumstances to proceed without robust barriers against third-party imitation. This Part reviews available information on actual sharing regimes or reasonably close variants thereof,⁵⁴ which yields a

52. Note that the northwest and southeast corners of the box yield ambiguous stability expectations: in the former case, group size and endowment heterogeneity are low, favoring sharing, but asset values and capital-intensity requirements are high, favoring property; in the latter case, group size and endowment heterogeneity are high, favoring property, but asset values and capital-intensity requirements are low, favoring sharing.

53. For views to this effect, see Ostrom, *Types of Goods*, *supra* note 32 (reviewing theoretical literature on public goods problems) and Ledyard, *supra* note 48 (reviewing experimental literature on public goods problems).

54. For purposes of this exercise, a sharing regime is understood to mean any innovation market (or market segment) where a substantial portion of the relevant pool of intellectual goods is unprotected by intellectual property protections or other access limitations, whether as a formal or effective matter.

systematic (if still preliminary) taxonomy of appropriation mechanisms in innovation markets that thrive without reliance on intellectual property protections.⁵⁵

The resulting landscape of sharing regimes and related arrangements exhibits two general tendencies that largely conform to the core theoretical expectations set forth above. First, the hypothetical model of a norm-based sharing regime, which relies substantially on reputational rewards and sanctions, is largely implemented in markets that support innovative output with little reliance on formal intellectual property rights or other access barriers. Second, these substantially pure-form sharing regimes tend to be confined to markets where innovators place little capital at risk (and, even in these settings, usually make some meaningful recourse to intellectual property or other exclusionary instruments). Beyond these small-scale environments, the anticipated result is realized: the enforcement technology behind a norm-based sharing regime can no longer easily support innovation incentives and participants increasingly deploy property rights in order to block unauthorized imitation.

But there is a third observation of vital importance. Typically the emergence of a formal property regime does not entirely displace existing knowledge-sharing arrangements, which persist even in higher capital-intensity settings involving large numbers of differentially endowed participants. Following commons-styled reasoning, this fact could be interpreted to advance the proposition that intellectual production sometimes does not require access barriers. Properly construed, however, this observation substantially embellishes the standard incentive-based understanding of intellectual property and easily integrates into a long-term payoff-maximization framework. Even under a property-rights regime, repeat-player innovators seek to preserve nominal-cost mechanisms for knowledge exchange that preserve the low transaction-cost structure of a sharing regime. Remarkably, the contractual design of these embedded sharing arrangements is driven by the same reciprocity principle that drives the norm-based design of stand-alone sharing regimes, which operate without recourse to formal property rights. Through the use of property and contract to regulate access, these finely tuned sharing arrangements can scale at economically intensive settings by regulating group size and composition so as to promote satisfaction of the reciprocity principle. Regulating access in turn precludes individually rational defections that typically threaten stand-

55. For another attempt at organizing the landscape of knowledge-sharing arrangements, see Julien Penin, *Open Knowledge Disclosure: An Overview of the Evidence and Economic Motivations*, 21 J. ECON. SURVEYS 326 (2007).

alone sharing regimes unsupported by property rights. In short, sharing is most stable with property, not without it.

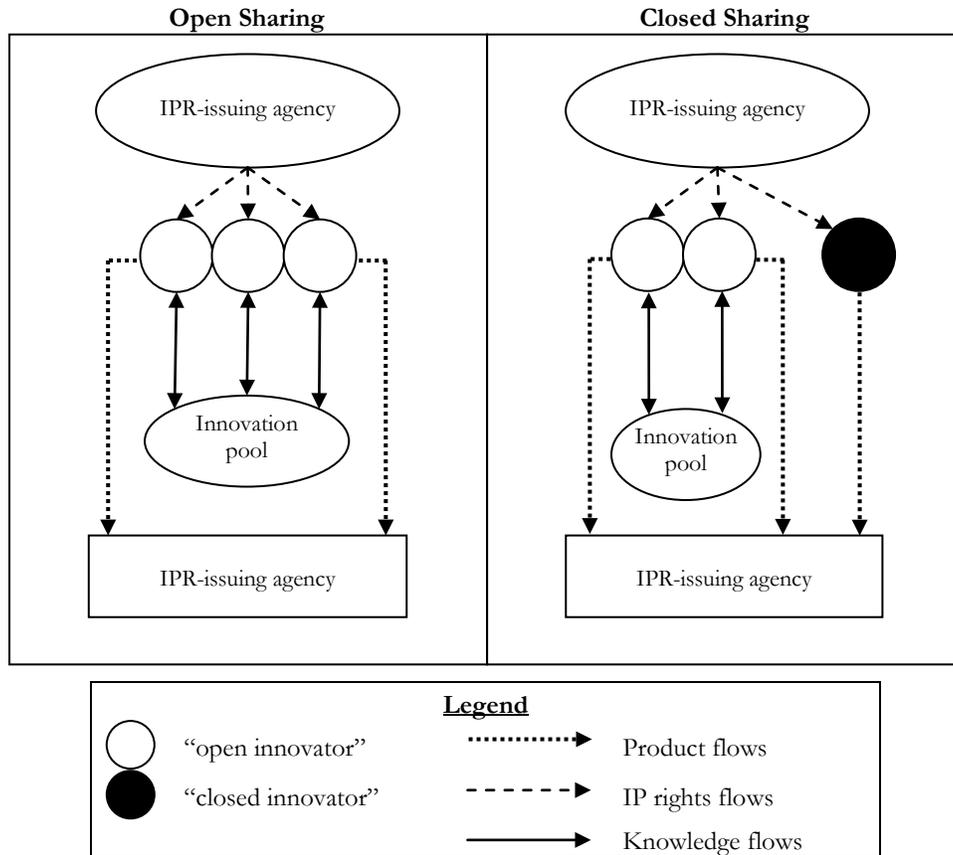
A. REGIME TAXONOMY

Tellingly, it is surprisingly difficult to locate innovation markets that implement a pure-form sharing regime where intellectual property protection is entirely absent. Hence, actual sharing regimes are best situated along an “access continuum.” This continuum ranges from “open” versions to “closed” or “semi-closed” versions. “Open” versions refer to regimes where intellectual property rights are formally available but weak, regularly waived, or otherwise largely unused, as a result of which at least some innovative output is deposited in a collective pool to which all participants have access. “Closed” or “semi-closed” versions refer to regimes that make substantial recourse to formal intellectual property rights but maintain innovation pools that are accessible to member firms subject to a mix of contractual and norm-based constraints.⁵⁶ Figure 5 below provides a graphical illustration of these two “mixed-form” sharing regimes (*open/closed sharing*), which may be usefully compared with the idealized pure-form sharing regimes (*sharing/property*) set forth earlier in Figure 1.⁵⁷

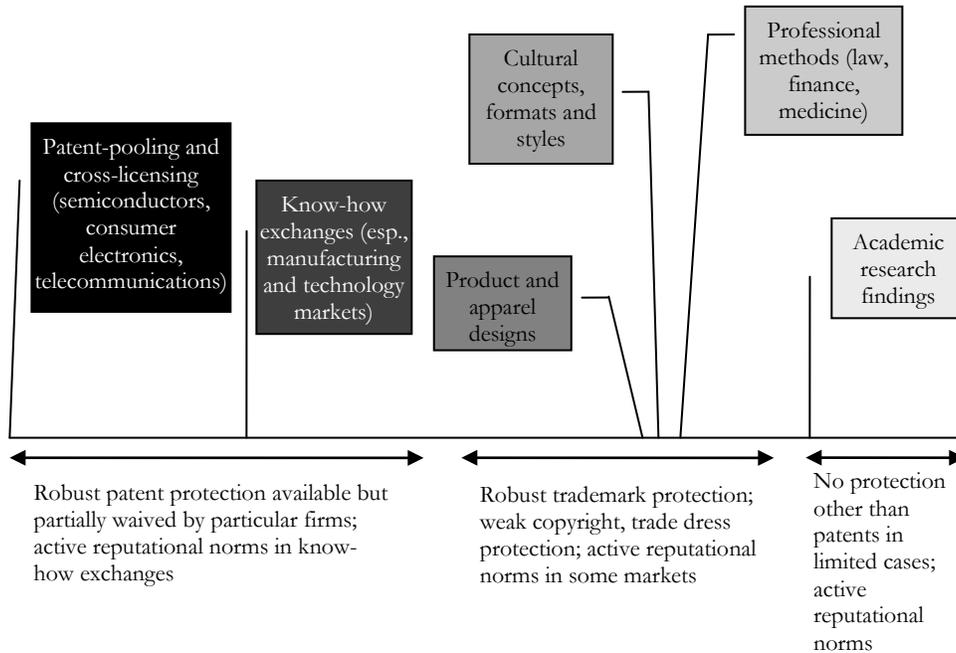
56. For a related distinction between informally organized and formally-organized knowledge-sharing mechanisms, see Penin, *supra* note 55, at 327.

57. Note that, following previous usage, “open” innovator refers to an innovator that participates in a nominal to low-cost exchange of intellectual assets; a “closed” innovator does not.

Figure 5: Alternative Regimes (Mixed Form)



This abstract distinction between closed and open sharing regimes translates as a practical matter into a graduated continuum of sharing regimes with different levels of non-negotiated third-party access, as set forth in Figure 6 below. Moving from right to left, access costs to the existing knowledge stock increase as the innovator population makes increasing use of property rights and decreasing use of the reputational reward and sanction mechanisms that support a norm-governed sharing regime. Approximately as the Figure moves from low capital-intensity markets in the research, design, professional, and cultural fields, to high capital-intensity markets in the technology and manufacturing fields, participants make greater use of property rights in general. These high capital-intensity markets make greater use of the strongest forms of intellectual property rights (moving from trademark and trade dress to copyright to patents) in particular, and lesser use of reputational norms to support innovation incentives.

Figure 6: Actual Sharing Regimes⁵⁸

The economic logic seems clear. As the innovator population (or at least, its stronger members) places greater capital at risk as a result of technological requirements, it anticipates higher expected losses in the event it incorrectly elects *cooperate* and a competitor elects *defect(copy)*. This in turn induces the strongest portions of the innovator population to act preemptively by electing *defect(property)*. Innovators abandon a norm-governed innovation regime (which can secure innovation returns at low capital intensities by recourse to reputational rewards and sanctions, to a law-governed regime). In short, increased losses in the event of expropriation justify the increased transaction-cost burdens imposed by the strongest form of legal protection.⁵⁹

58. For simplicity, this chart ignores the limited availability of patent protection for financial-method innovations, which has existed since 1998. Given the Federal Circuit's decision in *In re Bilski*, 545 F.3d 943 (Fed. Cir. 2008), and the Supreme Court's subsequent decision in *Bilski v. Kappos*, 130 S. Ct. 3218 (2010), the patentability of certain financial-method patents continues to be subject to uncertainty. Following earlier usage, increasingly dark coloration corresponds to increasing proprietization, and vice versa.

59. Scientific research (which does require substantial investment) is the exception to this relationship, which in turn accounts for the extensive subsidies provided to this market. For further discussion of this last point, see *infra* Section V.B. Note that the observed relationship does not imply that a norm-governed innovation regime could not operate at

Hence, common observations that certain low capital-intensity environments (on the right side of the spectrum) sustain innovative output, without substantial recourse to intellectual property, presumptively fail to generalize (to the left side of the spectrum) to higher capital-intensity environments, which are unlikely to induce investment in the absence of a secure barrier against third-party expropriation.

B. OPEN SHARING REGIMES

Open sharing models persist in forms substantially untouched by property rights with respect to an important set of product attributes and therefore come closest to realizing the pure-form model of a sharing regime. The most economically salient markets that fall within this category can be classified into four broad categories: (1) research—i.e., scientific and other academic research, where abstract ideas are ineligible for patent protection; (2) design—i.e., fashion and product design, where design patents, copyrights and trade dress generally offer unreliable protection for utilitarian components of any garment or industrial design;⁶⁰ (3) culture—i.e., plots, routines, formats and certain other conceptual elements used in film, television and theatrical productions, where there is weak to no protection against non-literal style and format imitation; and (4) the professions—i.e., methods or procedures used in law, finance, accounting and the medical professions.⁶¹ Legal protections against imitation in these markets are generally absent, weak, or ineffective, and, as a result, there is widespread and regular circulation of concepts, methodologies and/or designs, which are then modified and re-circulated without any remuneration flowing directly to the original contributor. Consistent with the theoretical model of a norm-based sharing regime, it should be expected that reputational rewards and sanctions would be deployed to cover the incentive shortfall generated by incomplete intellectual property coverage. This in turn ensures both a rough parity of net contributions to the innovation pool over time and a premium for original contributions over derivative applications, thereby precluding the

higher capital intensities assuming other relevant environmental variables were hospitable to it, but it tilts the odds against this possibility considerably.

60. More specifically: (1) design patent protection is usually practically ineffective given the associated delays and costs, (2) in light of *Wal-Mart Stores, Inc. v. Samara Bros., Inc.*, 529 U.S. 205 (2000), trade dress protection usually requires showing “secondary meaning,” and (3) copyright protection is unavailable for any utilitarian articles (and generally, any “conceptually inseparable” component thereof).

61. It is possible to patent medical procedures; however, this is now practically moot in light of a 1996 amendment to the Patent Code that immunizes physicians and medical facilities from liability for infringement of any medical procedure patent. There is some patent protection for financial methods, although that now is uncertain. *See supra* note 58.

under-innovation result. As described in greater detail subsequently with respect to academic research,⁶² formal and informal mechanisms for allocating inventive credit assure that original contributors accrue appropriate reputational rewards. However, in certain market segments, gross imitation triggers reputational penalties (or, to the extent trademark protections apply, legal penalties) for excessively close replications of successful originals.

A small body of scholarship documents the imitation norms that govern cultural and other market segments covered by weak or minimal intellectual property protections. These include luxury furniture design, luxury French restaurants, extreme-sports equipment hobbyists, magicians, stand-up comics, and online fan fiction contributors.⁶³ Consistent with theoretical expectations, these innovation communities tend to be relatively small in number, demand low capital investment, appear to have relatively homogenous endowments, and maintain informal mechanisms for administering reputational rewards and penalties, which in turn elicit contributions to, and regulate withdrawals from, the innovation pool. To illustrate in some more detail, consider the luxury furniture industry in Italy and the Netherlands. Designers operate with little effective protection against imitation other than unreliable copyright protections; however, they nonetheless abide by industry norms that limit excessive imitation and reward original contributions in the form of reputational credit (which is then sometimes monetized in the form of increased market premia for the most creative designers). This reputational economy is in turn facilitated by regular informal and formal communications among competing designers that can stigmatize any firm that violates these imitation norms.⁶⁴ As this market

62. See *infra* Section V.B.

63. See Emmanuelle Fauchart & Eric von Hippel, *Norm-Based Intellectual Property Systems: The Case of French Chefs*, 19 *ORG. SCI.* 187 (2008) (describing luxury French restaurants); Gerda Gemser & Nachoem Wijnberg, *Effects of Reputational Sanctions on the Competitive Imitation of Design Innovations*, 22 *ORG. STUDIES* 563 (2001) (examining Dutch and Italian luxury furniture design); Greg Lastowka, *Digital Attribution: Copyright and the Right to Credit*, 87 *B.U. L. REV.* 41 (2007) (digital forms of literary creation); Jacob Loshin, *Secrets Revealed: Protecting Magicians' Intellectual Property Without Law*, in *LAW AND MAGIC: A COLLECTION OF ESSAYS* 123 (Christine A. Corcos ed., 2007); 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) (referring to stand-up comedy routines); Rebecca Tushnet, *Payment in Credit: Copyright Law and Subcultural Creativity*, 70 *LAW & CONTEMP. PROBS.* 135 (2007) (discussing online fan fiction); Sonali K. Shah, *From Innovation to Firm Formation in the Windsurfing, Skateboarding, and Snowboarding Industries* (Univ. of Ill., Working Paper No. 05-0107, 2005), available at <http://apps.olin.wustl.edu/faculty/conferences/cresgort/pdf/6SonaliShah.pdf> (describing U.S. amateur extreme-sports hobbyists and small-business owners).

64. See Gemser & Wijnberg, *supra* note 63.

illustrates, social reward and sanctioning mechanisms can apparently substitute, at least in part, for weak intellectual property protection, and thereby induce innovation that is otherwise subject to replication. Interestingly, extensive use of honorific practices in research, design and other cultural markets may not be a sociological accident. Rather, it may be symptomatic of the fact that these markets induce innovation through the lower transaction-cost structure of a sharing regime that rewards contribution substantially through reputational rewards, as opposed to the higher transaction-cost structure of a property regime that rewards contributions primarily in monetary remuneration.

Following commons-styled thinking, it might be tempting to generalize these markets (which are certainly not short on innovation) as a paradigm case for the proposition that intellectual production typically does not require intellectual property or other access limitations. But at least one important characteristic common to all these markets immediately counsels against any such interpretation. Namely: none of these markets constitute pure stand-alone sharing regimes as envisioned in our theoretical discussion. That is, there is always some positive level of intellectual property protection available. In research markets, copyright protects against literal replication of verbal content and patent protection limits third-party usage of some applied-science findings. In design markets, trademark protects against unauthorized reproductions of name and logo (and, in non-apparel design markets, patents and trade secrets may limit unauthorized usage of other components of the relevant product). In cultural markets, trademark protects against use of name and logo and copyright protects (at least) against literal reproduction of written, visual or musical expression. In the professions, trademark protects against use of name and logo and, in finance, trade secrets (and, more recently but still to a much lesser degree, patents) may play an important role in limiting unauthorized usage of technical methods and other valuable knowledge. Moreover, even where intellectual property protections are especially minimal or ineffective, there often exist substantial levels of tacit knowledge (e.g., research methods), technological opacity (e.g., magic tricks or cuisine), associated products, services, or other business capacities (e.g., financial methods that are packaged together with the reputational capital of an established financial institution)⁶⁵ that frustrates easy or perfect imitation of the total product or services bundle provided by the original contributor. This fact is critical: if there exists some nontrivial level of exclusionary protection, whether provided legally or extra-legally, then some

65. See Peter Tufano, *Financial Innovation and First-Mover Advantages*, 25 J. FIN. ECON. 213 (1989).

product attributes are not thrown into the collective innovation pool. This precludes exact replication, and therefore allows consumers to distinguish between originators and imitators. That in turn enables the reliable operation of the attribution technology that supports the allocation of reputational awards and sanctions, which in turn generates collateral streams of monetary returns for original contributions, which in turn supports innovation investment. This is entirely consistent with the conventional incentive model! So, at best, these markets are really paradigm cases for the important proposition that intellectual production sometimes does not require a lot of intellectual property (or some practical equivalent).

In substantial conformity with theoretical expectations, this preliminary survey of open sharing markets yields a highly qualified proposition that sets strict bounds to any practical realization of the commons model. Namely, intellectual production at low capital intensities among small-number populations with substantially equivalent-value innovation endowments usually does not require strong levels of intellectual property, which is largely (but not completely) replaced by social norms that impose imperfect constraints on unauthorized imitation. This narrow proposition implies in turn that this norm-based incentive structure is unlikely to generalize to capital-intensive innovation environments, which, subject to other identified factors, therefore do require robust forms of exclusionary protection. Subject to further case-specific inquiry, social norms are unlikely to substitute adequately for intellectual property or other exclusionary protections in large-scale innovation markets characterized by high capital-intensity investments, large numbers, high endowment heterogeneity and high economic values for the relevant asset class. But this does not consign sharing mechanisms to the exotic margins of innovation markets. This proposition has an important positive implication that reserves an important place for sharing practices even in large-number and capital-intensive environments. Namely, these practices are unlikely to substitute for intellectual property, but are likely to operate as a complementary mechanism for reducing the transaction-cost burden inherent to property-rights protections.⁶⁶ Just as property has staying power even in innovation markets characterized by low levels of capital

66. Sharing arrangements and other forms of inter-firm cooperation can play other important purposes in innovation markets, including most notably, achieving gains from collective cost-sharing and risk-sharing mechanisms. For an exploration of the former possibility, see WILLIAM J. BAUMOL, *THE FREE-MARKET INNOVATION MACHINE: ANALYZING THE GROWTH MIRACLE OF CAPITALISM* Chs. 6–7 (2002); for an exploration of the latter, see Jonathan M. Barnett et al., *The Fashion Lottery: Cooperative Innovation in Stochastic Markets*, 39 J. LEGAL STUD. 159 (2010).

investment, so too do sharing practices have staying power even in innovation markets characterized by high levels of capital investment.

C. CLOSED SHARING REGIMES

Closed sharing models operate in innovation markets that (1) widely adopt intellectual property protections, (2) decline to enforce these rights with respect to knowledge exchanges with certain (usually, substantially similar peer) competitors, but (3) do enforce these rights to restrain access by other (usually, substantially dissimilar non-peer) competitors, or by any other firm, over some other class of intellectual goods. This tailored enforcement of intellectual property rights effectively constructs an innovation pool to which only member firms have access, subject to any contractual agreement as to contribution requirements, withdrawal limitations, and collateral royalty or other payments. These closed sharing arrangements are endemic in some of the most economically significant industries and appear in two forms, as broadly defined below. First, as a large social-science literature documents, geographic clusters exist (and historically existed) in crafts, industrial design, high-technology, and some manufacturing industries where rivals' employees engage in informal exchanges of technological know-how.⁶⁷ Alternatively, know-how may be embodied in fluid human capital that regularly shifts between employers.⁶⁸ These cluster formations effectively waive trade-secrecy protections over certain classes of intellectual goods in a certain segment of a larger industry (which, in some cases, otherwise does make use of patent protections).⁶⁹ Second, a wide variety of manufacturing and high-technology industries employ, or have employed, formal sharing arrangements in the form of cross-licensing or patent-pooling schemes that

67. For an indicative reference source, see A HANDBOOK OF INDUSTRIAL DISTRICTS (Giacomo Becattini et al. eds., 2009).

68. See ANNALEE SAXENIAN, REGIONAL ADVANTAGE: CULTURE AND COMPETITION IN SILICON VALLEY AND ROUTE 128 (1994) (describing the high-technology industry in Silicon Valley and Boston area); Ronald Gilson, *The Legal Infrastructure of High Technology Industrial Districts: Silicon Valley, Route 128 and Covenants Not to Compete*, 74 N.Y.U. L. REV. 575 (1999) (describing the same subject).

69. See, e.g., Robert C. Allen, *Collective Invention*, 4 J. ECON. BEHAV. & ORG. 1 (1983) (describing blast furnace industry in 19th-century Cleveland, England); Alessandro Nuvolari, *Collective Invention During the British Industrial Revolution: The Case of the Cornish Pumping Engine*, 28 CAMBRIDGE J. ECON. 347 (2004) (describing the steam-engine industry in the mining district Cornwall, England); Eric von Hippel, *Cooperation Between Rivals: Informal Know-How Trading*, 16 RES. POL'Y 291 (1987) (describing minimills in steel industry). This is an incomplete list of know-how exchange and similar arrangements. For a discussion of some additional examples, see BAUMOL, *supra* note 66, at 86–90.

implement a partial effective waiver of certain patent protections.⁷⁰ Industries using such agreements and licenses include: (1) the consumer electronics industry, which widely operates on the basis of arrangements that pool “essential patents” contributed by participating firms in connection with a variety of industry standards for fundamental audio and video compression and transmission technologies;⁷¹ (2) the biotechnology industry, which widely uses strategic technology alliances and other multi-firm cooperative research and other arrangements;⁷² and (3) the semiconductor industry, which relies on cross-licensing arrangements that provide large peer competitors with reciprocal access to an agreed-upon pool of patented assets.⁷³

A closed sharing arrangement that makes recourse to formal property rights in order to exclude non-members is substantially more stable than an open sharing arrangement that does not make use of any such exclusionary mechanism. As a result, it can induce innovation at substantially higher capital intensities. The reason is straightforward. Contract plus property rights backed up by the threat of state coercion provide a far more powerful technology for maintaining regime stability. In particular, this model presents

70. Professor Robert Merges has pioneered research in this area. See Robert P. Merges, *Contracting into Liability Rules: Intellectual Property Rights and Collective Rights Organizations*, 84 CAL. L. REV. 1293, 1340–54 (1996); Robert P. Merges, *Institutions for Intellectual Property Transactions: The Case of Patent Pools*, in EXPANDING THE BOUNDARIES OF INTELLECTUAL PROPERTY: INNOVATION POLICY IN THE KNOWLEDGE SOCIETY 123 (Rochelle Dreyfuss, Diane L. Zimmerman, & Harry First eds., 2001) [hereinafter Merges, *Patent Pools*]. I am excluding from this discussion performance rights organizations that pool copyrights relating to musical compositions (e.g., BMI and ASCAP), the reason being that these organizations simply pool copyrights in order to economize on licensing and enforcement costs and do not seek to facilitate knowledge-sharing among competing producers. Some, but not all, patent-pooling entities may share this characteristic.

71. For detailed discussion of some of these arrangements, see U.S. PATENT & TRADEMARK OFFICE, PATENT POOLS: A SOLUTION TO THE PROBLEM OF ACCESS IN BIOTECHNOLOGY PATENTS? (2000), available at <http://www.uspto.gov/web/offices/pac/dapp/opla/patentpool.pdf>; David Serafino, *Survey of Patent Pools Demonstrates Variety of Purposes and Management Structures*, KNOWLEDGE ECOLOGY INTERNATIONAL (June 4, 2007), <http://keionline.org/content/view/69/1>; and MPEG LA, <http://www.mpegla.com> (containing information regarding MPEG LA, the leading administrator of patent pools in the electronics industry). For overviews of patent-pooling arrangements in the consumer electronics and other industries, see Anne Layne-Farrar & Josh Lerner, *To Join or Not to Join: Examining Patent Pool Participation and Rent Sharing Rules* (Working Paper, 2007), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=945189; DAVID. J. TEECE, MANAGING INTELLECTUAL CAPITAL: ORGANIZATIONAL, STRATEGIC, AND POLICY DIMENSIONS, app. A.1.1 (2000).

72. In the biotechnology industry alone, over 20,000 alliances were reported as of 1996. See Maryann P. Feldman, *Strategic Research Partnerships in Biotechnology*, available at www.nsf.gov/statistics/nsf01336/p1s7.htm.

73. See TEECE, *supra* note 71, at app. A.

a far more powerful technology for deterring individually rational defections than the leaky technology supplied by social norms and the threat of reputational sanctions. Consistent with the incentive structure described previously, the broad extension of closed sharing arrangements across a wide variety of innovation markets follows from the fact that property rights enable participating firms to preserve stability through contractual requirements that regulate group size and endowment heterogeneity. Cross-licensing or patent-pooling arrangements use two principal instruments to regulate group composition so as to preserve regime stability. First, these arrangements usually implement access limitations that regulate endowment heterogeneity by assessing the value of firms' contributions to the collective pool. This is accomplished through an expert certification mechanism that evaluates whether any submitted patent is "essential" for the relevant technological standard.⁷⁴ Second, these arrangements often implement contractual requirements that correct for endowment heterogeneity through calibrated royalty payments that reflect substantially higher or lower-value contributions to the collective pool.⁷⁵ Additionally, governing contractual agreements limit defection opportunities into the surrounding property regime through grant-back provisions that require all members (and, typically, non-member licensees) to contribute to the pool all improvements deemed to be "essential" to the licensed technology.⁷⁶

These contractual mechanisms, as grounded in state-provided property rights, enable firms to satisfy the reciprocity principle that otherwise would dissuade participation by firms that could accrue higher gains by defecting into the surrounding property regime. The outcome: a limited number of participating firms with substantial endowment homogeneity who have little rational incentive to elect *defect(copy)* or *defect(property)*, thereby resulting in a high level of regime stability. Evidence on participation patterns in patent-pooling, cross-licensing, and know-how exchanges is consistent with this expectation: (1) firms with especially valuable technological assets sometimes opt out of participating in a patent pool (especially if a value-sensitive royalty

74. This is a typical element of patent pool agreements, especially in the consumer electronics industry. See Layne-Farrar & Lerner, *supra* note 71, at 9; Merges, *Patent Pools*, *supra* note 70, at 29–30, 34–35.

75. See Layne-Farrar & Lerner, *supra* note 71, at 3.

76. See Merges, *Patent Pools*, *supra* note 70, at 30–31, 35; see also Serafino, *supra* note 71, at 18, 22, 23, 26 (noting that MPEG-2 patent pool for video compression technology, the MPEG-4 patent pool for audio and visual compression technology, and the DVD3C and DVD6C patent pools for audio and video storage technology include grant-back commitments that all future essential patents held by licensors will be licensed back into the pool).

formula is lacking but even when it is present in some cases),⁷⁷ (2) firms are willing to enter a patent pool subject to a value-insensitive royalty formula when patent contributions are roughly symmetrical across firms,⁷⁸ and (3) in industries where even direct competitors routinely exchange proprietary know-how, firms are more likely to do so with firms who have comparably valued technology resources and often defect from the sharing norm by using property rights to safeguard the highest-value knowledge assets.⁷⁹ Moreover, formal property rights allow prospective members to safely and credibly disclose to each other endowment levels with a reduced risk of expropriation, which may be a necessary precondition to entering into a cooperative arrangement that seeks to maintain membership homogeneity in order to ensure satisfaction of the reciprocity principle.

It might be argued that this thesis does not fully characterize some multi-firm cross-licensing, standard-setting, and patent pooling arrangements, which sometimes cover a broad range of market participants with heterogeneous endowments. But this discrepancy actually reflects the stability of a closed sharing arrangement, which overcomes two vulnerabilities in an open sharing arrangement that operates without state-provided property rights. First, on the “high end” of the endowment distribution, closed sharing arrangements are able to generate a calibrated cooperation payoff that induces some participation by the strongest innovators. This occurs through tailored royalty-stream allocations and other payment mechanisms that reward exceptional contributions (sometimes complemented by allowances that permit those participants to exclude the most valuable patents).⁸⁰ Second, on the “low end,” these sharing arrangements induce some participation by weak innovators due to the exclusionary mechanisms that at least partially eliminate any anticipated defection payoff (that is, increase the cost of remaining outside the resource pool to which members can restrict

77. See Layne-Farrar & Lerner, *supra* note 71, at 3, 20–21. The authors cite the example of Lucent, who chose not to participate in the MPEG-2 patent pool, unlike most other major players in the industry, apparently on the view that it could extract greater value by licensing its especially valuable patents independently. *See id.*, at 7, 14. It turned out to be mistaken and, based on the “MPEG LA” website, is now a member. *See* MPEG LA, *supra* note 71.

78. See Layne-Farrar & Lerner, *supra* note 71, at 3.

79. For an indicative example, see von Hippel, *supra* note 69, who documents information sharing among competing steel “minimills” but additionally observes that an “outlier” firm declined to participate in this practice; not surprisingly, that firm appears to possess technical expertise that cannot be reciprocated by its competitors. *See id.* at 296. On further references to studies of know-how exchanges, see *supra* note 69.

80. See Merges, *Patent Pools*, *supra* note 70.

access)⁸¹ while contractual devices may be able to accommodate weak innovators without unduly eroding the cooperation payoff that flows to the existing pool of strong innovators. This is a somewhat paradoxical result: selective use of property rights (together with contract law) allows the sharing arrangement to capture the most “dangerous” lowest and highest fringes of the innovator population. While this increases endowment heterogeneity within the sharing arrangement, it decreases the defection payoff for weak innovators and increases the cooperation payoff for strong innovators. It thereby protects the cooperation payoff for “average” innovators against both weak innovators who would otherwise elect *defect(copy)*, and strong innovators who would otherwise elect *defect(property)*, which in turn could threaten the stability of the sharing arrangement.

D. SUMMARY

This Part has undertaken two tasks. First, it has provided a reasonably comprehensive taxonomy of innovation markets that maintain substantial and documented arrangements for the inter-firm exchange of valuable knowledge. Two core categories have been identified: (1) open sharing regimes characterized by knowledge exchange governed largely by social norms, and (2) closed sharing regimes characterized by knowledge exchange governed largely by multilateral contractual instruments grounded in intellectual property rights. Second, it has demonstrated that both sharing regimes substantially conform to an underlying reciprocity principle: whether through norms, contract, property rights, or some combination thereof, actually implemented sharing regimes are mixed arrangements that make extensive efforts to regulate the number and composition of the participants in any sharing arrangement. Moreover, consistent with theoretical intuitions, those access regulations increase in force and sophistication—largely as indicated by the move from norms to contract and some limited implementation of property rights—as numbers, endowment heterogeneity, capital investment and asset values increase. In the next Part, three selected markets will be studied in detail to assess further the robustness of these observations.

81. On the cost of remaining outside a technology-sharing consortium, see BAUMOL, *supra* note 66, at chs. 6–7; WILLIAM J. BAUMOL, *ENTREPRENEURSHIP, MANAGEMENT, AND THE STRUCTURE OF PAYOFFS* ch. 10 (1994). Baumol makes the important point that, in contrast to ejection from a price-setting cartel (where the ejected member can continue to profit from the supra-competitive prices set by the cartel), ejection from a technology-sharing consortium results in no benefits except to the extent there are information spillovers. This contingency obviously improves the cooperation payoff in the latter scenario.

V. CASE STUDIES: THREE ILLUSIONS OF THE COMMONS

This Part provides case studies of sharing arrangements in three disparate markets—pre-modern craft guilds, academic research, and open-source software—that are often referenced as paradigm illustrations for the commons thesis that intellectual production can proceed without limitations on access.⁸² This discussion provides the final component in the cumulative sequence of theoretical and empirical argument that I have presented to assess the reliability of the commons model for understanding innovation markets and making innovation policy. The commons model fails to reliably account for the mechanisms that support innovation even in these apparently open-access markets. This failure is consistent with both (1) theoretical expectations based on the hypothetical construct of a sharing regime, as presented in Part II, and (2) the general tendencies in actual sharing regimes, as presented in Part III. Innovation investment in these weakly propertized markets relies on, and would be unlikely to persist without, collateral instruments that restrict access and thereby generate remunerative streams to reward contributions to the innovation pool. By dispensing with any aspirational view that these markets successfully sustain (or sustained) innovative output in a free-appropriation environment unencumbered by exclusionary protections, it is possible to observe a remarkably consistent pattern in the mixed implementation of property-based and sharing-based strategies. Together these strategies generate the nuanced hybrid regimes that govern (or governed) these otherwise historically and technologically disparate markets. Remarkably, all three markets exhibit a nested mixed-form structure consisting of: a “sharing core” that enables low transaction-cost exchanges of intellectual assets among peer innovators, which is embedded within a “property perimeter” consisting of legal entitlements that enable the sharing core by regulating access and thereby preserving the conditions that support rational contributions to the innovation pool. To appreciate the analytical ground that has been covered, one may compare the complexity of these actual innovation regimes (each of which is presented graphically in Figures 7, 8 and 9 in the following discussion) with the idealized pure-form and generic mixed-form sharing regimes presented previously in Figures 1 and 5, respectively.

82. *See supra* note 8. While craft guilds in particular are not commonly referenced in support of the case that intellectual property is an unnecessary incentive mechanism, it is often asserted that intellectual production existed prior to the advent of intellectual property, which in turn is then purported to cast doubt on the case for intellectual property. *See, e.g.,* KRANICH, *supra* note 8. Craft guilds are used in this analysis as an illustrative case of pre-modern intellectual production.

A. CRAFT GUILDS

Various forms of sharing regimes appear to have been the standard governance structure for innovation markets in pre-modern Europe,⁸³ as illustrated in particular by the guilds and similar collective organizations that characterized Western European crafts industries for approximately 500 years ending in the nineteenth century.⁸⁴ The mechanisms used to regulate innovation within craft guilds, and the ultimate demise of that structure, are remarkably consistent with this Article's thesis. First, in the absence of intellectual property rights, craft guilds relied on norm-based (and technological) constraints to stimulate the production of, and regulate access to, valuable knowledge. Eventually, the norm-based mechanisms behind craft guild innovation were challenged and ultimately displaced by state-provided property rights as outside economic values, group size, and endowment heterogeneity increased.

At the cost of over-generalization, the basic structure of a craft guild was as follows: the guild was usually assigned an exclusive (or semi-exclusive) license to provide a certain product in a certain territory; the guild was empowered to enforce its regulations with respect to its members; and the guild regulated, among other things, the employment and training of apprentices and the conformity of working processes and finished products with guild standards. In place of property entitlements held by individual innovators, guilds avoided under-innovation outcomes through substantial compliance with community norms to the extent maintained by business and other social sanctions among guild members (often but not always tied

83. See Stephan R. Epstein, *Property Rights to Technical Knowledge in Premodern Europe, 1300–1800*, 94 AM. ECON. REV. 382, 383 (2004) [hereinafter Epstein, *Property Rights*] (noting that, in Western Europe, “much premodern craft and engineering knowledge appears to have been shared . . . within industrial districts”).

84. Crafts guilds (associations of artisans) and merchant guilds (associations of traders) were leading forms of economic organization in pre-modern Europe. The historical literature is vast and can only be referenced selectively. For overviews, see PAMELA O. LONG, OPENNESS, SECRECY, AUTHORSHIP: TECHNICAL ARTS AND THE CULTURE OF KNOWLEDGE FROM ANTIQUITY TO RENAISSANCE 72–101 (2001); S. R. Epstein, *Craft Guilds, Apprenticeship and Technological Change in Preindustrial Europe*, 58 J. ECON. HIST. 684, 689–90, 705–07 (1998) [hereinafter Epstein, *Craft Guilds*]; Sylvia Thrupp, *The Guilds*, in 3 THE CAMBRIDGE ECONOMIC HISTORY OF EUROPE 231 (M. M. Postan et al. eds., 1963). For an important prior contribution that explores the importance of guild institutions for intellectual property scholarship, see Robert P. Merges, *From Medieval Guilds to Open Source Software: Informal Norms, Appropriability Institutions and Innovation* (Working Paper, 2004), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=661543 [hereinafter Merges, *Guilds*].

together by neighborhood, religious and/or kin relationships)⁸⁵ and between guilds, and as complemented further by collateral benefits in the form of collective branding, knowledge-sharing, risk-spreading, financing, and cost-sharing mechanisms.⁸⁶ Each guild adhered (or claimed to adhere) to norms that promoted mutual (albeit perhaps incomplete) disclosure of technical knowledge.⁸⁷ This disclosure norm followed the basic construct of a sharing regime, and yielded a collective pool from which members could make withdrawals and to which members could make contributions, thereby reducing the transaction costs of knowledge exchanges and the input costs of knowledge generation among individual craftsmen. Just as contemporary observers rally against the extension of patent rights as endangering commonly-held knowledge resources, English guilds advertised precisely the virtues of these informal knowledge-sharing practices in arguing against patent protection for certain mechanical inventions in the late seventeenth and early eighteenth centuries.⁸⁸ While that may appear to be a self-serving defense of monopolistic guild privileges, it can be defended as, at least in part, a good-faith attempt to preserve the low transaction-cost structure of a sharing regime against the administrative burdens of a property rights regime.⁸⁹

85. See Epstein, *Craft Guilds*, *supra* note 84, at 701. On the role of social capital in inducing compliance with guild norms, see Sheilagh Ogilvie, *Guilds, Efficiency and Social Capital: Evidence from German Proto-Industry*, 57 *ECON. HIST. REV.* 286, 286–333 (2004).

86. On these collateral benefits, see Epstein, *Craft Guilds*, *supra* note 84, at 686–88. For further discussion, see S.R. Epstein & Maarten Prak, *Introduction*, in *GUILDS, INNOVATION, AND THE EUROPEAN ECONOMY, 1400–1800*, at 1 (S.R. Epstein & Maarten Prak eds., 2007), and Ulrich Pfister, *Craft Guilds and Proto-Industrialization in Europe, 16th to 18th Centuries*, in *GUILDS, INNOVATION, AND THE EUROPEAN ECONOMY, 1400–1800*, *supra*, at 11–24.

87. See CHRISTINE MACLEOD, *INVENTING THE INDUSTRIAL REVOLUTION: THE ENGLISH PATENT SYSTEM, 1660–1800*, at 83 (1988); see also Epstein & Prak, *supra* note 86, at 18 (noting that shipwrights' guilds promoted knowledge sharing through regular meetings at which attendance was compulsory); Liliane Pérez, *Inventing in a World of Guilds: Silk Fabrics in Eighteenth-Century Lyon*, in *GUILDS, INNOVATION, AND THE EUROPEAN ECONOMY, 1400–1800*, *supra* note 86, at 232, 256–57 (noting Lyon silk guilds' ethos that encouraged the free circulation of knowledge); Thrupp, *supra* note 84, at 274 (noting that cost-reducing process innovations would be shared among members of the guild and kept secret from outsiders).

88. See MACLEOD, *supra* note 87, at 188. Guilds in other jurisdictions similarly opposed the extension of patent protection as “privatizing” common knowledge. See Epstein, *Property Rights*, *supra* note 83, at 384.

89. Professor Robert Merges views guilds as a form of “collective invention” whereby members used secrecy practices and other mechanisms to appropriate returns from innovation activities, which may have efficiency benefits that are overlooked by the conventional dismissal of the guilds as being a pure rent-seeking enterprise. See Merges, *Guilds*, *supra* note 84. On the conventional view of craft guilds (and its limitations), see Epstein & Prak, *Introduction*, *supra* note 86, at 1–2.

The commons literature sometimes makes reference to pre-modern forms of intellectual production to support the thesis that innovation can be sustained even without the expectation of monetary or other remuneration.⁹⁰ But this assumes that no exclusionary mechanisms were employed by pre-modern markets prior to the advent of formal intellectual property, an assumption that (to this author's knowledge) has received virtually no inquiry and, at least with respect to the craft guild, would be seriously misleading. The craft guild never operated as a stand-alone incentive structure as contemplated by the idealized construct of a norm-governed sharing regime. Rather, every guild operated under the protection of a state-granted exclusive license (or one of a restricted set of licenses) that protected the relevant guild against imitation by non-members, as complemented by secrecy procedures and statutory authorizations to enforce guild rules through compulsory membership and other sanctions. As shown in Figure 7 below, a guild is best viewed as a voluntarily formed sharing arrangement (denoted by the box with bolded lines) embedded within a formal property regime constituted by exclusionary entitlements allocated by the state, which in turn generated revenue streams that sustained innovation incentives for the guild as a whole. While there were few intellectual property protections at the individual level, these protections were robust at the group level. Through this modified property-rights regime, the guild avoided the transaction costs of a full-fledged property regime but, through grant of an exclusive or semi-exclusive license, encouraged innovation by permitting guild members to internalize as a collective entity some of the social gains generated by private investment.⁹¹

This statement is obviously an imperfect account of the incentive structure that supported guild innovation in the absence of robust intellectual property rights. Even where the state-granted license securely blocked entry by non-guild competitors (not universally the case),⁹² it still did not address an inherent defect that threatened the guild with under-innovation. While the guild license sustained collective incentives to innovate, it did not provide any support for individual incentives to innovate. To encourage investments

90. See *supra* note 8.

91. Obviously grant of a monopoly license may to a certain extent *depress* innovation given the absence of any potential entry threat. In particular, guilds would appear to have had an incentive to oppose labor-saving innovations, which would have eroded their competitive advantage over non-member artisans. The evidence suggests, however, that this reputation is partly undeserved. See Epstein, *Craft Guilds*, *supra* note 84, at 694–96; Ulrich Pfister, *Craft Guilds and Technological Change: The Engine Loom in the European Silk Ribbon Industry in the Seventeenth and Eighteenth Centuries*, in *GUILDS, INNOVATION, AND THE EUROPEAN ECONOMY, 1400-1800*, *supra* note 86, at 172; Thrupp, *supra* note 84, at 271–79. For a defense of the conventional view, see Ogilvie, *supra* note 85.

92. See Epstein, *Craft Guilds*, *supra* note 84, at 705–06; Thrupp, *supra* note 84, at 276–78.

in new process technologies or in the transfer of technical knowledge to apprentices, some further remunerative mechanism was necessary. A partial remedy for this defect (which may account for the guilds' reputation for technical conservatism)⁹³ may have been provided by the technical requirements for guild membership to the extent that they screened out weak innovators through the apprenticeship process.⁹⁴ This screening process assisted in preserving some approximate parity between contributions and withdrawals from the collective innovation pool. But this effective protection against knowledge spillovers to weak innovators still did not provide a rational incentive for a strong innovator to incur the costs of generating innovations (and transmitting innovations to apprentice labor) that would then be thrown into the collective pool with no direct remuneration for the contributing innovator. Guilds appear to have used a variety of devices to address precisely this vulnerability, including: (1) barring poaching of apprentices by guild members (which allowed each artisan to recoup his training investment);⁹⁵ (2) permitting members to extract some return on private innovations by implicitly allowing the use of secret cost-reducing technical processes provided the final product conformed to the guild standard;⁹⁶ (3) quasi-bartering schemes whereby innovative artisans exchanged secret technical improvements;⁹⁷ (4) inviting exceptional non-members in possession of technical innovation to join the guild (often in exchange for not opposing issuance of a patent);⁹⁸ (5) providing individuals with special remuneration or prizes for exceptional innovations that would then be available to guild members generally;⁹⁹ and, in certain cases, (6) even

93. See MACLEOD, *supra* note 87, at 113 (same, with respect to English guilds in particular); Epstein, *Craft Guilds*, *supra* note 84, at 693 (noting and partially contesting this impression).

94. See Epstein & Prak, *Introduction*, *supra* note 86, at 7–9.

95. See Epstein, *Property Rights*, *supra* note 83, at 383.

96. This point is emphasized in Epstein, *Craft Guilds*, *supra* note 84, at 693–95. For further discussion with respect to fifteenth-century Venetian glass-making guilds, see LONG, *supra* note 84, at 91–92 and Merges, *Guilds*, *supra* note 84, who observe that guilds sometimes allowed members to keep technical processes secret.

97. See MACLEOD, *supra* note 87, at 188.

98. See *id.* at 83–84.

99. For examples of these policies in the eighteenth-century Lyon silk-weaving industry, see Dominique Foray & Liliane Hilaire Perez, *The Economics of Open Technology: Collective Organization and Individual Claims in the “Fabrique Lyonnaise” During the Old Regime*, in *NEW FRONTIERS IN THE ECONOMICS OF INNOVATION AND NEW TECHNOLOGY* 239, 245 (Cristiano Antonelli et al. eds., 2006).

assisting patenting by individual members, who in turn assumed certain teaching responsibilities within the guild.¹⁰⁰

These internal regulatory mechanisms functioned to preserve the reciprocity principle that falters in any sharing community as endowment heterogeneity increases: strong innovators will rationally constrain participation in the absence of calibrated reward mechanisms that reflect exceptional contributions to the innovation pool. Consistent with theoretical expectations, erosion of the reciprocity principle posed a key threat to the longevity of any guild organization. Unless substantial parity between contributions and withdrawals among differently-endowed innovators could be assured, either by regulating entry into the guild and/or allocating compensatory side-payments to strong innovators, the latter group would rationally constrain contributions or, when feasible, defect into a state-provided property regime where appropriate remuneration for original contributions could be obtained on the open market. Several historical incidents illustrate this risk. The eighteenth century Lyon silk-weaver guilds, which both emphasized the free circulation of knowledge and operated (with state assistance) a limited quasi-patent regime for certain silk-weaving inventions, sometimes experienced disputes with the best inventors over the grant and/or terms of an “exclusive privilege” (a quasi-patent right).¹⁰¹ More generally, highly innovative guild members were sometimes bought out (that is, induced to defect) by rival jurisdictions or guilds in exchange for a one-time royalty payment (functionally equivalent to a lump-sum payment for an intellectual property right), a not uncommon occurrence as higher-value supra-regional markets developed and offered increased economic rewards for technological advances.¹⁰² Not coincidentally, the rapid growth of these lucrative markets in the early nineteenth century, and the resulting ability of talented artisans to better withdraw intellectual goods from the collective pool constituted by craft guilds, seems to have played some part in the ultimate decline of the guild institution and increased usage of the formal

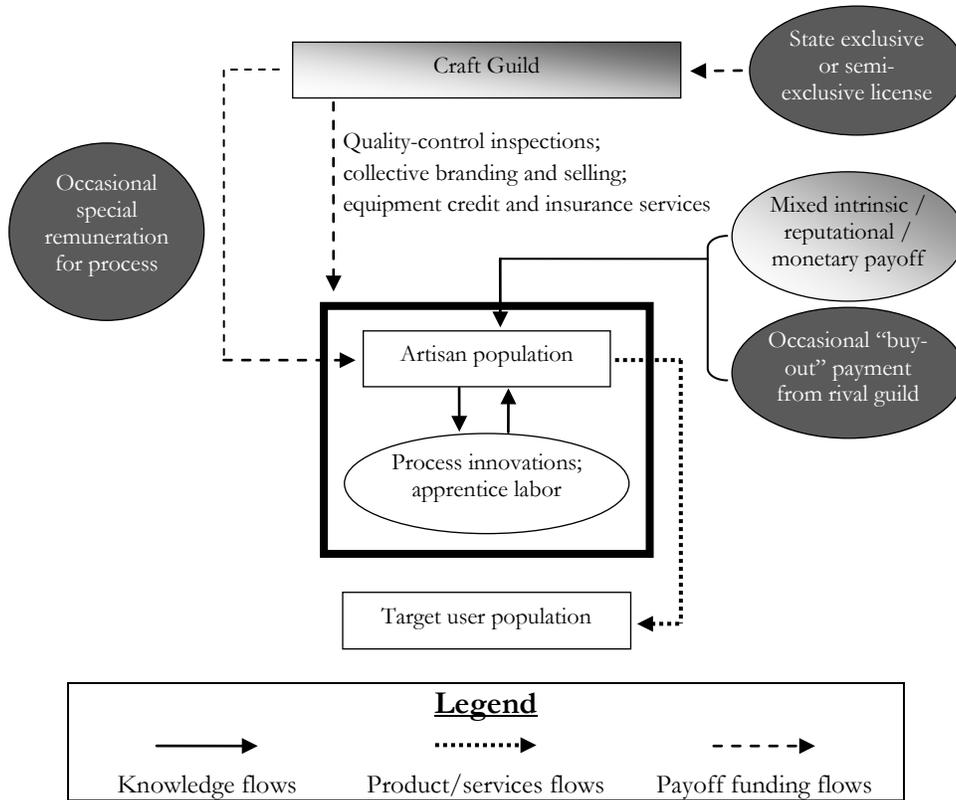
100. See Pérez, *supra* note 87, at 235–36 (making this observation with respect to the 18th-century Lyon silk weaving industry).

101. See Foray & Perez, *supra* note 99. Through state support, the eighteenth-century Lyon silk guilds maintained an early system of patent-like protection for silk-weaving inventions, which provided remuneration for inventors based largely on market success. See Pérez, *supra* note 87, at 232, 247–51, 258–59 (noting that best inventors were unhappy to operate in collective invention system that did not provide appropriate remuneration); *id.* at 262 (noting that guild system had persistent difficulty in appropriately remunerating the best inventors).

102. See MACLEOD, *supra* note 87, at 147; Epstein, *Craft Guilds*, *supra* note 84, at 703–05.

patent system.¹⁰³ Consistent with this Article's general thesis, as outside economic values, group size, and endowment heterogeneity increased, the most talented innovators rationally withheld contributions to the pool, which then stagnated. As a result, the guild institution unraveled.

Figure 7: Mixed-Form Sharing Regime (Craft Guilds)¹⁰⁴



B. ACADEMIC RESEARCH

Academic research is one of the clearest examples where intellectual production thrives in the absence of property rights. Closer scrutiny shows that it is remarkably consistent with this Article's thesis: the limited

103. See Epstein, *Craft Guilds*, *supra* note 84, at 705–07. Other commentators argue that the capital accumulation in a mature industry enabled individual merchant-manufacturers to undertake production of certain goods without recourse to the cost-sharing and risk-spreading advantages of the guild mechanism. See, e.g., Ulrich Pfister, *Craft Guilds and Proto-Industrialization in Europe, 16th to 18th Centuries*, in *GUILDS, ECONOMY AND SOCIETY* 172 (Clara Eugenia Nunez ed., 1998).

104. Consistent with prior usage, darker coloration denotes practices indicative of a property regime; lighter coloration denotes practices indicative of a sharing regime; intermediate coloration denotes mixed practices indicative of both regimes.

availability and use of intellectual property rights compels academic institutions to induce investment in research through a mix of reputational mechanisms, tax-funded subsidies and the sale of an inherently excludable product elsewhere in the consumption bundle. Namely: teaching.

Basic research results have generally not been subject to formal property-rights protection, aside from patent protection for some applied results in the hard sciences. Furthermore, in virtually all academic fields, the free exchange of research findings is a widely encouraged practice, and the hoarding of research results is a widely discouraged practice. This results in rapid dissemination of knowledge assets. These norms generate a common innovation pool from which researchers at competing institutions make withdrawals subject to attribution to the contributing author, and to which researchers make contributions in the form of research findings. Setting aside for a moment the limited availability (and even more limited use) of patent protection in some fields of scientific research, what propels researchers to invest time and effort in intellectual production even though the resulting positive externalities cannot even be partially internalized? The answer, as sociologists of science have observed, conforms precisely to the hypothetical construct of a norm-driven sharing regime. Social practices operate in virtually all disciplines to award reputational rewards that sustain output in academic research, where researchers follow openness norms. These norms mandate uncompensated forfeiture of private knowledge in exchange for the prospect of reputational prestige for innovation success, which is in turn supported by norms that mandate giving credit to prior innovators (and sanction harshly those who fail to give credit).¹⁰⁵ Reputationally driven contribution norms in the academic research market rest on a transparent and low-cost attribution technology—namely, the citation—that facilitates the fine allocation of credit among contributing researchers based on citation

105. See ROBERT MERTON, *THE SOCIOLOGY OF SCIENCE: THEORETICAL AND EMPIRICAL INVESTIGATIONS* 286–324 (1968); JEROME R. RAVETZ, *SCIENTIFIC KNOWLEDGE AND ITS SOCIAL PROBLEMS* 41–42, 245–59 (1971). For further discussion, see Partha Dasgupta & Paul David, *Toward a New Economics of Science*, 23 *RES. POL'Y* 487 (1994); Paul A. David, *The Historical Origins of 'Open Science': An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution* 3 *CAPITALISM & SOC'Y* no. 2, art. 5 (2008), <http://www.bepress.com/cas/vol3/iss2/art5> [hereinafter David, *Patronage*]; Christopher Kelty, *Free Science*, in *PERSPECTIVES ON FREE AND OPEN SOURCE SOFTWARE* 415, 416–27 (Joseph Feller et al. eds., 2005); Paula E. Stephan, *The Economics of Science*, 34 *J. ECON. LIT.* 1199 (1996). The role of informal reputation-based norms in academic research is mentioned in the canonical work on norm-based substitutes for legal regulation, see ELLICKSON, *supra* note 21, at 59–64, 258–64.

counts, peer-review processes, and journal placement.¹⁰⁶ Reputational capital has two further benefits. First, it is a naturally compounding asset, meaning that substantial accruals of reputational capital may enable a researcher to pay the “fee” required to gain access into the most elite professional circles that discuss the most advanced methodologies or findings in the relevant field.¹⁰⁷ Second, researchers can partially monetize reputational capital in certain disciplines into higher salaries, outside publishing contracts, consulting engagements, and other material benefits.¹⁰⁸

Consistent with an open sharing model that relies heavily on reputational carrots and sticks to overcome any potential threat of excessive withdrawals from the common innovation pool, regular use of this attribution technology in conformity with the governing norm is supported by potentially severe reputational sanctions: perfect imitation without attribution (i.e., plagiarism) can result in career-ending reputational or other institutional penalties, while failure to make contributions halts further career advancement. Consistent with the sharing model, original researchers who widely disclose valuable knowledge can accrue substantial reputational rewards, allocated both through professional prestige, continuously operating mechanisms for peer review, and a wide variety of formal honors. Through this combination of market norms, and a well-developed enforcement apparatus of peer-review journals, grant-making institutions, and other entities that make appropriate allocations of reputational capital to outstanding researchers, the academic research market provides the most vivid contemporary example of an innovation pool sustained largely without recourse to state-provided property rights.

Based on these observations, it might be tempting to conclude that academic research shows that innovation can proceed without recourse to formal property rights or any other exclusionary instrument, which are

106. On attribution and reputational norms in scientific and other academic scholarship, see Rebecca S. Eisenberg, *Proprietary Rights and the Norms of Science in Biotechnology Research*, 97 YALE L. J. 177, 181–84 (1987); Catherine Fisk, *Credit Where It's Due: The Law and Norms of Attribution*, 95 GEO. L. J. 49, 64–65, 81–85 (2006); Robert P. Merges, *Property Rights Theory and the Commons: The Case of Scientific Research*, in SCIENTIFIC INNOVATION, PHILOSOPHY, AND PUBLIC POLICY 145, 148–52 (Ellen Frankel Paul et al. eds., 1996) [hereinafter Merges, *Scientific Research*]; Arti Kaur Rai, *Regulating Scientific Research: Intellectual Property Rights and the Norms of Science*, 94 NW. U. L. REV. 77, 88–94 (1999).

107. See Paul A. David, *Communication Norms and the Collective Cognitive Performance of “Invisible Colleges,”* in CREATION AND TRANSFER OF KNOWLEDGE: INSTITUTIONS AND INCENTIVES 115, 128–29 (G. Barba Navaretti et al. eds., 1998).

108. See Stephan, *supra* note 105, at 1202–03.

replaced by reputational rewards.¹⁰⁹ But this account overlooks a simple fact: academic research in any recognizable form is supported by collateral revenue streams that are excludable, which implies that researchers only partially rely on reputational payoffs in electing whether to make innovation investments. Both historical and contemporary practices in the production of academic knowledge conform to this proposition. At its inception during and shortly after the Renaissance, early forms of scientific research demanded relatively low levels of capital investment and could subsist on the monetary infusions supplied by aristocratic patrons or the independent resources of gentlemen scholars.¹¹⁰ In its modern and highly capital-intensive form, scientific research is supported by four principal revenue streams, together amounting to tens of billions of dollars annually: (1) cash grants from government agencies or large philanthropic institutions, (2) tuition payments by students, (3) alumni donations, and (4) part-time or post-career employment in the private sector. Federal research grants to academic research account for the largest component of this funding bundle. In 2007, these amounted to approximately \$28.5 billion in the United States,¹¹¹ which constituted almost 90% of total research expenditures. Scholarly commentators who advance commons-styled understandings of “pre-property” academic research generally recognize this awkward fact in passing,¹¹² but then fail to observe that it actually demonstrates that any apparently property-free model rests on either property-based appropriation

109. See, e.g., DOMINIQUE FORAY, *ECONOMICS OF KNOWLEDGE* 147 (2004) (stating that “open science model” shows that knowledge production can take place in an “IPR”-free zone, although notes that universities must rely on public funding). For similar thoughts that academic research functioned well prior to the advent of intellectual property, which is then viewed as endangering the free dissemination enabled by traditional norms in the research community, see Rai, *supra* note 106.

110. For an extensive description of these patronage arrangements, see David, *Patronage*, *supra* note 105.

111. See NAT'L SCIENCE FOUND./DIV. OF SCIENCE RESOURCES STATISTICS, *SURVEY OF FEDERAL SCIENCE AND ENGINEERING SUPPORT TO UNIVERSITIES, COLLEGES AND NONPROFIT INSTITUTIONS, FY 2007 (SEPT. 2009)*, available at <http://www.nsf.gov/statistics/nsf09315/pdf/tab1.pdf>. Note that this figure does not include state or private contributions to academic research.

112. For an example of an open-access advocate who takes this fact seriously in designing an academic “knowledge commons,” see Peter Suber, *Creating an Intellectual Commons Through Open Access*, in *UNDERSTANDING KNOWLEDGE AS A COMMONS: FROM THEORY TO PRACTICE* 171, 175–76 (Charlotte Hess & Elinor Ostrom eds., 2007). For prior contributions that recognize the importance of public funding and other capital inflows to sustain scientific research, see Merges, *Scientific Research*, *supra* note 106, at 155, and Scott F. Kieff, *Facilitating Scientific Research: Intellectual Property Rights and the Norms of Science—A Response to Rai and Eisenberg*, 95 NW. UNIV. L. REV. 691 (2001).

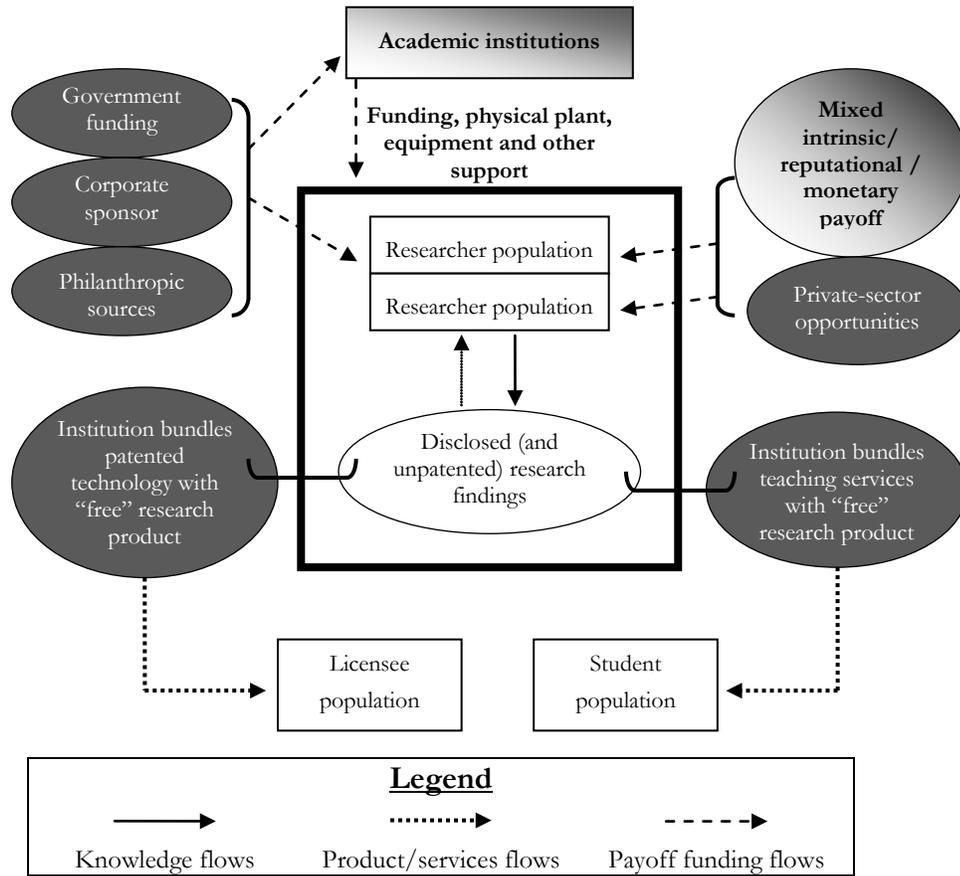
instruments, or coercive taxation to compel the necessary contributions to the public good constituted by scientific knowledge.

Properly construed, the university operates as an embedded sharing arrangement that is supported by public-goods contributions from either a coercive taxing authority (i.e., the government) and/or voluntary philanthropic institutions. The university then generates intellectual goods that can be used in three ways. First, the goods can be allied to an educational enterprise that provides an excludable good in the form of teaching services in return for which it receives an excludable stream of cash remuneration from its student clientele. Second, the goods can be “sold” on various terms to corporate partners that enter into research contracts or other forms of sponsorship with the university. Or alternatively, as enabled in part by the Bayh-Dole Act of 1980,¹¹³ which permitted universities to patent the results of federally funded research, the goods can be allied to a licensing enterprise that generates cash returns from licensees of the university’s patented technology, which is then obviously available to the collective innovation pool. So understood, the university is a knowledge-production enterprise that voluntarily participates in a sharing arrangement where it pools some intellectual goods with competing institutions for mutual advantage (represented by the “sharing core” denoted by the bolded box at the center of Figure 8 below). That is in turn funded by the proprietary sale of excludable goods to paying students, corporate sponsors, and corporate licensees. From this perspective, the university research environment looks entirely different from its standard characterization as a property-free zone that supports innovation through the publicly-interested spirit of academic pursuit. The “free” exchange of knowledge assets, which at first appears to be the key characteristic of academic research, is sustainable as a result of both (1) “internal” norm-based governance that allows for the regular allocation of reputational rewards and penalties based on a freely-exchanged body of research findings, and (2) collateral revenue streams generated by coercive redistribution (taxes), voluntary redistribution (philanthropy), and the sale of excludable assets to paying students and corporate sponsors under a conventional property-rights regime. Without these collateral revenue streams, the academic research enterprise would be compelled to migrate to a property-based model, which sustains innovation at high transaction costs (as exists in corporate research and existed in part prior to university-based

113. Bayh-Dole Act, Pub. L. No. 96-517, 94 Stat. 3019 (1980) (codified as amended at 35 U.S.C. §§ 200–211 (2000)).

academic research), or degenerate into an open-access commons, which fails to sustain innovation altogether.¹¹⁴

Figure 8: Mixed-Form Sharing Regime (Academic Research)¹¹⁵



C. OPEN-SOURCE SOFTWARE

Open-source software is one of the most widely-cited cases for the view that intellectual production can proceed and even thrive in the absence of

114. Hardly speculation: prior to the full development of the modern system of peer-reviewed scientific journals, the history of science is rife with concealment of results or partial communications of new findings in order to preserve returns from research investments, facts consistent with a modified open-access commons. See David, *Patronage*, *supra* note 105; RAVETZ, *supra* note 105, at 247–49. Based on the analytical framework set forth above, these earlier practices are easily explained: without a robust funding mechanism to close the incentive shortfall, researchers rationally declined to make valuable contributions to a shared innovation pool from which commensurate withdrawals were not forthcoming.

115. Consistent with prior usage, darker coloration denotes practices indicative of a property regime; lighter coloration denotes practices indicative of a sharing regime; intermediate coloration denotes mixed practices indicative of both regimes.

intellectual property or other access restrictions. That intuitively seems right: open-source software is defined precisely by the voluntary (if partial) disclaimer of intellectual property rights. Closer scrutiny, based largely on the extensive empirical inquiries conducted to date, shows that funding models and participation patterns tend to depart from that intuition. Consistent with this Article's thesis, open-source projects that achieve scale—that is, the most successful projects, tend to be implemented through mixed arrangements that combine giveaways of software code with the sale of complementary services and goods to institutional users and a mix of monetary, reputational and other benefits to paid and volunteer programmers.

Open-source software¹¹⁶ is an industry segment where software products and the underlying source code are released at no fee subject to relaxed contractual restrictions on use and distribution.¹¹⁷ The code is then subsequently improved by “volunteer” programmers (the reason for the quotation marks will soon become clear).¹¹⁸ In an open-source environment, the principal recourse to the state-provided property regime arises insofar as open-source software is released subject to contractual licenses. These licenses require inclusion of the developers' copyright notice (for attribution purposes), and sometimes (as in the case of the most widely-used “GNU General Public License” (GPL) license and variants thereof)¹¹⁹ obligate the user to distribute any derivative applications under the same “open source” terms as the original license, which substantially complicates the commercial distribution of derivative applications.¹²⁰ The GPL license and its variants

116. The scholarly literature is already extensive. For a critical overview, see Stephen M. Maurer & Suzanne Scotchmer, *Open Source Software: The New Intellectual Property Paradigm* (Nat'l Bureau of Econ. Research, Working Paper No. 12148, 2006), <http://www.nber.org/papers/w12148>. For a readable book-length overview, see STEVEN WEBER, *THE SUCCESS OF OPEN SOURCE* (2004). For a widely-known popular history, see ERIC S. RAYMOND, *THE CATHEDRAL AND THE BAZAAR: MUSINGS ON LINUX AND OPEN SOURCE BY AN ACCIDENTAL REVOLUTIONARY* (1999).

117. Source code refers to the human-readable instructions that compose a computer program. By contrast, proprietary software is released in non-human-readable object-code form (which is a translation of source code made using compiler software) for a fee and under strict contractual restrictions on use and distribution.

118. That is a simplified definition; as described below, actual market practice in the terms of open-source software licenses can vary considerably. However, the industry generally relies on an “official” definition supplied by the Open Source Initiative, which effectively sets a minimum threshold that must be satisfied by any OSI-certified license. For more information, see *The Open Source Definition*, OPEN SOURCE INITIATIVE, <http://www.opensource.org/docs/osd> (last visited Nov. 7, 2010).

119. See FREE SOFTWARE FOUND., GNU GENERAL PUBLIC LICENSE, *available at* <http://www.fsf.org/copyleft/gpl.html>.

120. Other open-source software uses the Berkeley Software Distribution (“BSD”) license or close variants thereof, which do not impose these constraints on subsequent

rely on contract law in order to deter individually rational defections into the surrounding property regime: it bars exclusive distribution of derivative applications of any open-source code because its drafters correctly anticipate that this would effectively constitute a withdrawal of assets from the shared innovation pool, and would ultimately undermine incentives by other innovators to make further contributions, thereby precipitating project failure.

Under any of the standard licenses, the open-source model exhibits much of the characteristics of a sharing regime. It generates a common innovation pool in the form of unprotected code, to which some developers regularly make contributions and from which other developers and end-users make withdrawals, in each case at minimal transaction costs given the voluntary waiver of most (but, critically, not all) property-rights protections. This model is a modified continuation of the informal “hacker” culture at the university computer science departments, and quasi-academic corporate research labs where a good deal of software development was initially launched, which encourages knowledge sharing and distributes reputational rewards for original contributions.

The open-source model has achieved substantial successes in certain markets, including most notably: the GNU/Linux operating system (used by some corporate and government entities), the Apache web server (which currently runs most internet websites), the Perl programming language, the SendMail internet e-mail engine (which is used to send a large portion of e-mail traffic over the internet), and the Mozilla web browser.¹²¹ In some popular, trade, business, and scholarly discussions, these successes have been used to support the claim that innovation in the software industry may be sustainable without bearing the high transaction-cost structure of a fully deployed property regime.¹²² But any account of the open-source software market as a stand alone environment that prospers without property rights or other imitation barriers seriously misunderstands the development, distribution, governance, and organizational structures at use in this market.

distributions. See WEBER, *supra* note 116, at 179–85. For a detailed description of the various types of licenses, see MARTIN FINK, *THE BUSINESS AND ECONOMICS OF LINUX AND OPEN SOURCE* (2003); LAWRENCE ROSEN, *OPEN SOURCE LICENSING: SOFTWARE FREEDOM AND INTELLECTUAL PROPERTY LAW* (2005).

121. See Ronald J. Mann, *Commercializing Open-Source Software: Do Property Rights Still Matter?*, 20 HARV. J.L. & TECH. 1, 9–10 (2006).

122. For the leading scholarly statement of this position in the legal literature, see BENKLER, *supra* note 10; Benkler, *Coase's Penguin*, *supra* note 10. For similar views, see James Boyle, *The Second Enclosure Movement and the Construction of the Public Domain*, 66 J.L. & CONTEMP. PROBS. 33, 45–46 (2003).

As can get lost in enthusiasm over what appears to be a weakly propertized but economically sustainable environment for innovation investment among a large mass of voluntary contributors,¹²³ the open-source model must confront and resolve the basic dilemma of any sharing regime: in the absence of restrictions on use and distribution (and, hence, any direct remuneration for original contributors), it must induce contributions from innovators who rationally demand returns in excess of development costs. This requires taking steps to regulate membership size and composition in any open-source project, which in turn sustains a roughly equal parity between contributions and withdrawals from the shared innovation pool (as corrected by side-payments or the equivalent thereof), thereby yielding a cooperation payoff in the form of reputational and/or monetary benefits that elicits migration from the surrounding property regime into this embedded sharing regime.

This expectation is remarkably consistent with actual practice. Open-source projects are sometimes described as mass-collaboration enterprises among hundreds to even thousands of diversely knowledgeable individual participants that somehow converge on a spontaneous order.¹²⁴ The unusually lavish scholarly attention devoted to the open-source market in its short history, however, has yielded virtually the opposite conclusion. Empirical researchers who look “behind the curtain” consistently find that open-source projects (or more precisely, the small minority of successful projects among the thousands of abandoned projects) are typically maintained by a small, core group of experienced developers (to which entry is often strictly constrained through internal control hierarchies) who exhibit high levels of technical sophistication and operate subject to reputational and other norm-governed pressures that elicit high effort.¹²⁵ For example, while

123. For a review of the literature and a similar observation, see Joseph Lampel & Ajay Bhalla, *The Role of Status Seeking in Online Communities: Giving the Gift of Experience*, 13 J. COMPUTER-MEDIATED COMM. (2007) (observing that “a fascination with the utopian aspects of virtual communities has strongly influenced research in this area,” which tends to be “highly attuned to features of virtual communities that highlight egalitarian and altruistic motivation”).

124. For the most well-known example in the popular literature, see RAYMOND, *supra* note 116, and for somewhat more nuanced versions in the legal literature, see BENKLER, *supra* note 10, at 66–67; Benkler, *Sharing Nicely*, *supra* note 10, at 332–39; Boyle, *supra* note 122, at 44–46.

125. See FINK, *supra* note 120, at 138–57; WEBER, *supra* note 116, at 70–71; Andrea Bonaccorsi & Cristina Rossi, *Why Open Source Software Can Succeed*, 32 RES. POL'Y 1243 (2003); Charles M. Schweik, *Free/Open-Source Software as a Framework for Establishing Commons in Science*, in UNDERSTANDING KNOWLEDGE AS A COMMONS, *supra* note 112, at 277, 285; see also Rishab Aiyer Ghosh, *Understanding Free Software Developers: Findings from the FLOSS Study*, in PERSPECTIVES ON FREE AND OPEN SOURCE SOFTWARE, *supra* note 105, at 23, 35 (noting

the Apache web server is used directly or indirectly by a broad pool of firms and other users, the maintenance and enhancement process is controlled by approximately twenty-five core developers, subject to formalized review and approval procedures to ensure system integrity (while larger groups of users submit “problem reports”).¹²⁶ Likely reflecting in part the disproportionate costs borne by these small groups of dedicated developers, open-source projects often fail to achieve scale beyond an initial “pioneer” effort, resulting in a high abandonment rate¹²⁷ (a fact sometimes obscured by casual references to “tens of thousands” of projects). This is an entirely unsurprising result in the case of a sharing regime that lacks an exclusionary mechanism to deliver remunerative streams that reflect differential individual contributions.

This observation still does not immediately rule out a commons-styled account of the open-source phenomenon (although high failure rates should immediately cast some doubt). This is because it fails to identify any rational support for the costly investments of time and effort even by these smaller groups of dedicated programmers in the small minority of successful projects, which therefore appear to operate on a largely or purely voluntary basis. But two further observations show this anomaly to be substantially overstated. First, there simply is no puzzle with respect to roughly half of all

that “[m]easures of source-code authorship show that a few individuals are responsible for disproportionately large fractions of the total code base” and referencing other studies that reach similar results). For membership and screening procedures as described in great detail with respect to the Debian project, see Fabrizio Ferraro & Siobhan O’Mahony, *Managing the Boundary of an ‘Open’ Project* (Harvard NOM, Working Paper No. 03-60, 2004), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=474782 (noting that contributors to open-source projects must provide “joining scripts” to show commitment to the project and describing cryptographic and other technical tools used to regulate access to the code base), and for a similar study with respect to the Freenet project, see Georg von Krogh et al., *Community, Joining Script and Specialization: A Case Study*, 32 RES. POL’Y 1217 (2003) (describing detailed admission requirements and apprenticeship and similar training periods to regulate admission into “core” developer group).

126. See Audris Mockus et al., *Two Case Studies of Open Source Software Development: Apache and Mozilla*, in PERSPECTIVES ON FREE AND OPEN SOURCE SOFTWARE, *supra* note 105, at 163, 171–75. For similar, more general observations, see ROSEN, *supra* note 120, at 43–45. See also Bonaccorsi & Rossi, *supra* note 125, at 1247 n.10 (referencing studies of contributions to the Apache, GNOME and other active open-source projects, which show heavy concentration of contributions among core group of developers).

127. See Brian Fitzgerald, *Has Open Source Software a Future?*, in PERSPECTIVES ON FREE AND OPEN SOURCE SOFTWARE, *supra* note 105, at 93, 96–97 (noting that in a study of over 400 registered open-source projects most had two or fewer developers and the vast majority appeared to be abandoned); Mockus et al., *supra* note 126, at 187 (noting that open-source projects sometimes fail to scale because core developers cannot handle and coordinate the quasi-administrative tasks of finding and repairing defects, resulting in code of suboptimal quality).

open-source programmers, who are employed or sponsored by for-profit software incumbents or not-for-profit foundations (which are usually sponsored by for-profit companies).¹²⁸ The “half” figure is a gross understatement, however, since it counts all contributors equally. However, the most recent evidence (including a study by the Linux Foundation) shows that paid programmers are more productive than unpaid volunteer programmers—that is, they contribute more per-capita than unpaid developers.¹²⁹ This is an entirely unremarkable finding within a rational-choice framework but directly contrary to the standard altruistic model used to explain participation in open-source projects. Second, available survey evidence tends to suggest that even the residual population of unpaid programmers are motivated by a miscellany of factors, including intrinsic interest in intellectual enjoyment, need for a customized program that did not yet exist in the market, the opportunity to improve programming skills, and,

128. See Rishab Aiyer Ghosh et al., *Survey of Developers, in FREE/LIBRE AND OPEN SOURCE SOFTWARE: SURVEY AND STUDY* 64–65 (2002), available at http://www.flossproject.org/report/FLOSS_Final4.pdf (finding that, based on online survey of 2784 developers active in “open source” or “free source” projects, 54% receive some kind of monetary rewards for code, administrative or other contributions); Maurer & Scotchmer, *supra* note 116, at 7 (reviewing similar results); see also WEBER, *supra* note 116, at 68–69 (noting that most developers involved in open-source projects appear to come from the private-sector rather than the academic sector); GREG KROAH-HARTMAN ET AL., *THE LINUX FOUND., LINUX KERNEL DEVELOPMENT* 10–12 (2009), available at <http://www.linuxfoundation.org/publications/whowriteslinux.pdf> (finding that 70% of total code contributions to the Linux kernel came from developers who are being paid to do so by for-profit companies, including Red Hat, Google, Novell, Intel, Oracle and IBM); Karim R. Lakhani & Robert G. Wolf, *Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Project, in PERSPECTIVES ON FREE AND OPEN SOURCE SOFTWARE*, *supra* note 105, at 3 (based on survey of 684 software developers, finding that 40% of the sample received direct financial compensation from employer for participation in open-source projects). Most current participants in open-source software arrangements are for-profit firms. See James Bessen, *Open Source Software: Private Provision of Complex Public Goods, in THE ECONOMICS OF OPEN SOURCE SOFTWARE DEVELOPMENT* 1, 6–7 (Jurgen Bitzer & Philipp J.H. Schroder eds., 2006).

129. See KROAH-HARTMAN ET AL., *supra* note 128, at 10; see also Evangelia Berdou, *Managing the Bazaar: Commercialization and Peripheral Participant in Mature, Community-Led Free Open Source Software Projects* 150 (2007) (unpublished dissertation, submitted to London School of Economics and Political Science) (based on study of GNOME and K Desktop open source projects, finding that paid developers are more likely to contribute to, and maintain, critical parts of the code base, as well as participate in community events); Jan Eilhard & Yann Ménière, *A Look Inside the Forge: Developer Productivity and Spillovers in Open Source Projects* 1, 16 (Working Paper, 2009), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1316772 (based on panel of 10,553 open source projects registered at SourceForge during February 2005–May 2007, finding that paid corporate developers are on average more productive than unpaid academic or private developers (as measured by file releases) but that the interaction between unpaid and paid developers on the same project can cause aggregate inefficiencies).

as some researchers emphasize, reputational capital that can translate into improved career prospects.¹³⁰ The potential reputational value attached by individual contributors to participation in high-profile open-source projects is illustrated by the fact that most projects have highly detailed attribution procedures—akin to the citation technology in the academic context—to apportion reputational credit to contributing programmers.¹³¹ This is not to deny that some programmers are motivated partly by payoff-insensitive ideological or other non-instrumental considerations, but it does not appear that it can reasonably be described as the prevailing motivating factor that drives participation by most open-source programmers, or more precisely, by unpaid open-source programmers.

Even the incentive effects of reputational utility and its monetizable by-products can be overstated as the key to resolving the “open source puzzle.” It is now clear that the most economically significant portions of the open-source software segment are most accurately viewed as a mutually beneficial venture among a restricted group of participant firms dedicated to the joint development of an open-access infrastructure, which will in turn support the provision of complementary products under an allied proprietary model.¹³² It is hard to underestimate the contribution—both in terms of cash, code and, most importantly, personnel—made by proprietary software companies to facilitate the development and adoption of open source’s largest successes to date. Consider Linux, perhaps the most successful open-source application. In 2001, IBM made a \$1 billion funding commitment to Linux open-source software development (which it claims to have recouped by 2002).¹³³ That

130. For studies that emphasize reputational effects, see Eric von Hippel & Georg von Krogh, *Open Source Software and the “Private-Collective” Innovation Model: Issues for Organization Science*, 14 *ORG. SCI.* 209 (2003); Josh Lerner & Jean Tirole, *The Simple Economics of Open Source* (Nat’l Bureau of Econ. Res., Working Paper No.7600, 2000), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=224008. Other studies are more mixed, generally finding weaker support for ideological values and stronger support for extrinsic incentives such as accrual of reputational capital and improving programming skills and intrinsic incentives such as user-based enjoyment, see, for example, Rishab Aiyer Ghosh, *Understanding Free Software Developers: Findings from the FLOSS Study*, in *PERSPECTIVES ON FREE AND OPEN SOURCE SOFTWARE*, *supra* note 105, at 23; Lakhani & Wolf, *supra* note 128. For a survey of empirical studies, see Bonaccorsi & Rossi, *supra* note 125, at 1246–50; Siobhán O’Mahony, *Guarding the Commons: How Community Managed Software Projects Protect Their Work*, 32 *RES. POLY* 1179 (2003).

131. See, e.g., FINK, *supra* note 120, at 28, 55; Lakhani & Wolf, *supra* note 128, at 7.

132. Elsewhere I describe in greater detail the extent to which leading open source software projects in the enterprise and mobile computing markets are substantially governed, funded and staffed by proprietary software, hardware, telecom, handset makers, and chip firms. See Barnett, *Host’s Dilemma*, *supra* note 14.

133. See Stephen Shankland, *IBM: Linux Investment Nearly Recouped*, CNET NEWS, Jan. 29, 2002, <http://news.cnet.com/2100-1001-825723.html>.

investment continues: today IBM employs 600 programmers at the IBM Linux Technology Center to maintain and improve the Linux operating system.¹³⁴ Together with other corporate sponsors, IBM sought to protect this investment in open-source development through the formation of (and the donation of 500 patents to) the Open Invention Network, a non-profit entity that acquires and warehouses Linux-related patents so as to preclude “hold-up” by third-party claimants.¹³⁵

The privately interested objective behind these apparently publicly interested investments is straightforward. By promoting an installed base constituted by a commoditized open-source operating system or software application (which offers an alternative to existing proprietary platform systems or applications), sponsor firms can accrue premia on proprietary applications, hardware and/or packaging, support, and documentation services that run on that base (e.g., IBM servers that run the Linux operating system).¹³⁶ Any rational choice anomaly largely disappears: each repeat-player firm incurs short-term costs (principally, losses attributable to disclosure of the source code and donated employee time and capital) in exchange for anticipated gains in the form of increased sales on complementary products and services. The practical result is reflected in the hybrid scheme set forth in the Figure below: an unprotected “sharing core” characterized by the free-exchange (and partially reputation-driven) practices typical of a sharing regime (denoted by the box in bold) is allied with complementary revenue streams on allied products and services that are protected by a legal or extralegal exclusionary instrument typical of a “conventional” property regime.¹³⁷

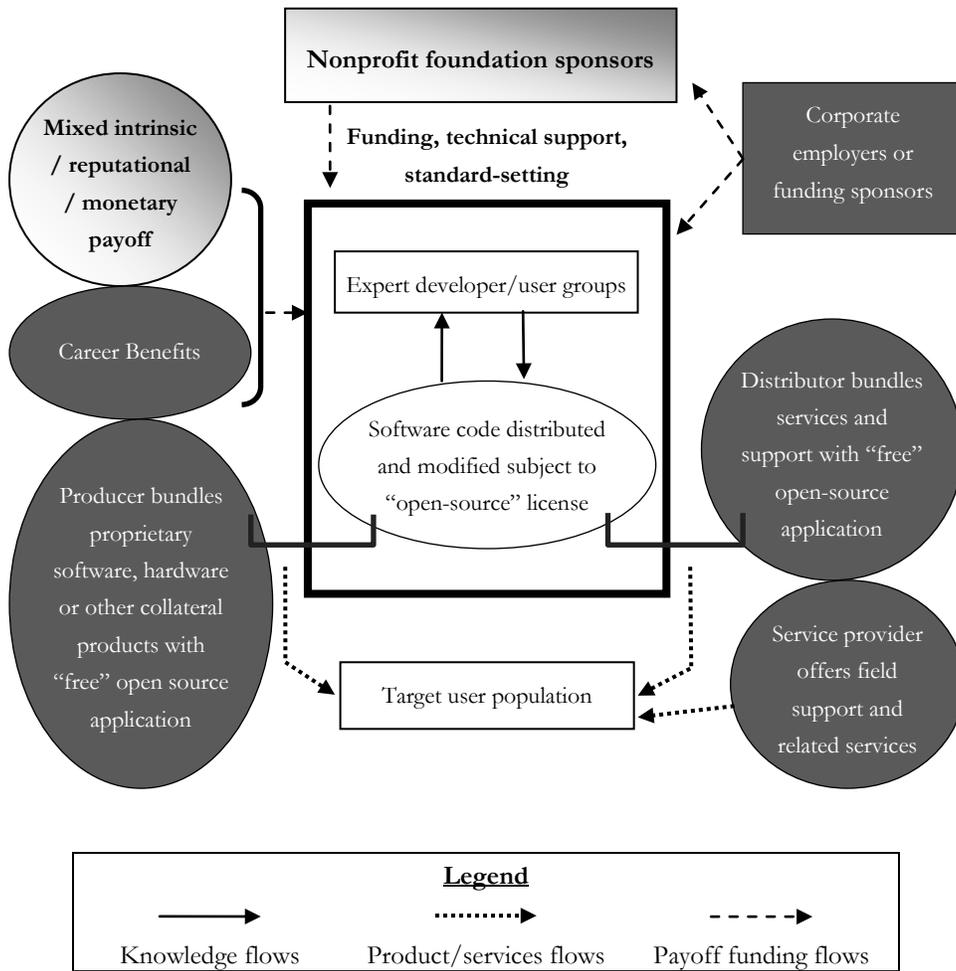
134. See David Kirkpatrick, *IBM Shares Its Secrets*, FORTUNE, Sept. 5, 2002, available at http://money.cnn.com/magazines/fortune/fortune_archive/2005/09/05/8271402/index.htm.

135. See Mann, *supra* note 121, at 29–30, 31 n.129 (2006). For further information, see OPEN INVENTION NETWORK, <http://www.openinventionnetwork.com> (last visited Nov. 7, 2010).

136. See WEBER, *supra* note 116, at 74–76; James Bessen, *Open Source Software: Free Provision of Complex Public Goods*, in THE ECONOMICS OF OPEN SOURCE SOFTWARE DEVELOPMENT 57 (Jürgen Bitzer & Philipp J.H. Schröder eds., 2006).

137. The generic taxonomy set forth in Figure 9 is a simplification of the diverse business models being deployed in the open source market. For further discussion, see FINK, *supra* note 120, at ch. 11; Joel West & Scott Gallagher, *Patterns of Open Innovation in Open Source Software*, in OPEN INNOVATION: RESEARCHING A NEW PARADIGM, 82 (Henry Chesbrough et al. eds., 2006).

Figure 9: Mixed-Form Sharing Regime (Open-Source Software)¹³⁸



The mixed-form structure of the most successful open-source projects conforms to this Article’s basic proposition. Any sharing regime that sustains economically significant investment must make recourse to the state-provided property system, or some other exclusionary instrument, in order to induce contributions. The open-source phenomenon certainly demonstrates the meaningful ability of reputational incentives (and related career benefits) to elicit certain levels of “voluntary” individual contributions to the innovation pool—a long-familiar phenomenon in academic research and

138. Consistent with prior usage, darker coloration denotes practices indicative of a property regime; lighter coloration denotes practices indicative of a sharing regime; intermediate coloration denotes mixed practices indicative of both regimes.

other “low-IP” settings. However, it equally demonstrates that, to sustain innovation projects that can scale to commercially significant levels, these reputational incentives must be accompanied by the conventional lure of monetary and other material benefits. This in turn necessitates recourse to some other legal or extralegal exclusionary instrument. Contrary to the tenor of some scholarly and trade commentary, but fully consistent with the prevailing findings in empirical research, the open-source market poses a relatively minor puzzle (if at all) for rational choice models of intellectual production. While an open-source project makes little recourse to the surrounding property regime to limit access to the innovation pool, it overcomes the free-rider threat by supplying an appropriation platform that can then generate demand for secondary products or services to which access is limited following a standard property model. As such, the open-source model is best understood as the most recent installment in an ongoing sequence of various combinations of sharing-based and property-based regimes whereby innovator populations seek to secure investment returns while minimizing the transaction-cost burdens that attend a formal property-rights regime. The true novelty of the open-source model lies in the fact that it represents a highly sophisticated tradeoff between the low transaction-cost burden of a sharing regime (mitigated by relaxed licensing of a common software platform) and the high innovation incentives of a property regime (sustained through remunerative streams from the sale of collateral proprietary products and services).

VI. CONCLUSION

In this Article, I set out to formulate and then assess a broadly representative version of a related set of positions that I group under the rubric of the commons thesis. This is understood to refer to the view that innovation markets can and do operate by recourse to reputation-driven norms in lieu of intellectual property or other exclusionary barriers. This intuition is normatively attractive and, presumptively, has some respectable factual grounding. Casual empiricism identifies innovation markets that thrive with little intellectual property and a great deal of rapid imitation; multiple case studies document the regulatory force of social norms in selected innovation markets; law-and-economics scholars and, in the common-pool resource context, political scientists and institutional economists, have documented the regulatory force of social norms (in lieu of legal instruments) in multiple settings outside of intellectual property. Sustained analysis seriously contests the reliability of this intuition. Theoretical and empirical scrutiny shows that the observation that some innovation markets apparently proceed vigorously without intellectual

property protections does not so easily yield the conclusion that economically-intensive forms of innovation can be sustained without some legal or other barrier against imitation. Basic rational-choice analysis anticipates that this norm-based model has a narrow expected scope of application: only under strict parameters is it plausible to believe that innovation will proceed without some legal or other barrier against imitation. In a certain respect, this discovery is entirely unsurprising: it is simply an extended application of the well-known claim that private contributions to a collective good in large-number settings will inevitably fail in the absence of material incentives to reward contributors and material sanctions to deter non-contributors. Empirics exhibit a tight fit with these expectations and diverge markedly from the commons thesis and its variants. A novel overview of actual sharing regimes shows that any apparently open-access environment for intellectual production either (1) tends to support economically insubstantial levels of innovation investment, or (2) actually does rely on some other exclusionary barrier, usually in connection with an allied product or service component that generates a positive remunerative stream to reward innovation. In other words: either the exception proves the rule or what appears to be the exception turns out to follow the rule after all!

This line of argument confines the scope of application of the commons thesis to small-scale or “little IP” environments characterized by low capital-intensity, low endowment heterogeneity and small group size—if, but only if, it is taken to stand for the proposition that sharing regimes can independently sustain innovation incentives without any substantial limitations on third-party access to the relevant product bundle. However, more constructively for purposes of future research, this line of argument exposes a far broader landscape of large-scale or “big IP” environments in which to expect that sharing practices will flourish as embedded mechanisms for alleviating the transaction-cost burdens that attend an extensively-deployed property regime. At least in innovation settings that demand substantial capital investments, it is of greater practical interest to adopt the following intermediate proposition: (1) sharing regimes confer substantial collective gains in the form of reduced transaction-cost burdens, but (2) outside of limited settings, are unlikely to persist unless supplemented by state-provided property rights or some other exclusionary mechanism of functional equivalence. This nuanced thesis explains both why (1) “stand alone” sharing regimes tend to be confined to low capital-intensity activities that tend to stand at the margins of economic activity, but (2) sharing practices and other nominal-cost exchange arrangements persist in embedded form in broad portions of the high-technology industries that operate at the heart of the current information-based economy. It is easy to see why the

commons model beckons so strongly. There do appear to be sharing communities that apparently sustain innovative output without robust legal barriers against imitation. However, sustained examination mostly bears out the wary intuitions of the rational choice skeptic—though not in a straightforward manner. Stand-alone sharing arrangements typically are only able to survive in low capital-intensity settings that tend to lie outside the most economically significant forms of technological and creative production; where this condition is not satisfied, these arrangements often persist in some form but are embedded within an environment secured by property rights or other access barriers.

This view might seem inconsistent with the law-and-economics literature on communities that maintain “order without law” and the social science literature on “limited-access commons regimes” that solve or ameliorate public-goods problems without recourse to state enforcement. But there is no such discrepancy. Scholars have identified settings where norms operate in lieu of law to address collective-action failures: this tends to occur in small-number communities consisting of a restricted membership of repeatedly-interacting players with similar endowments and interests (Shasta County ranchers,¹³⁹ New York diamond merchants,¹⁴⁰ and Maine lobstermen,¹⁴¹ to name a few). Not coincidentally, the most prudently constructed commons models in innovation environments are confined to low-stakes environments for cultural expression that do not require capital-intensive investments in research, production or distribution.¹⁴² These conditions are by definition unsatisfied by innovation markets of economic significance involving large numbers of differentially-endowed agents and large investment requirements, which must therefore make recourse to exclusionary instruments in order to sustain innovation incentives.

But the staying power of property in innovation markets does not banish sharing practices to the fringes of intellectual property scholarship. To the contrary: mechanisms for the low-cost exchange of intellectual goods persist at the very heart of innovation markets that widely implement intellectual

139. See ELLICKSON, *supra* note 21.

140. See Lisa Bernstein, *Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry*, 21 J. LEGAL STUD. 115 (1992).

141. See JAMES M. ACHESON, *THE LOBSTER GANGS OF MAINE* (1988).

142. See Lastowka, *supra* note 63 (arguing that amateur creators of original works of authorship are motivated by nonmarket considerations and that creation can be sustained without copyright protections so long as attribution norms are respected); Tushnet, *supra* note 63 (describing how individual contributions to “fandom” literature, which modifies and extends storylines in existing literary or other entertainment content, are governed by a mix of social norms and copyright protections, and suggesting that this offers a future model for user-generated cultural production).

property rights. Just as rational self-interest defeats any stand-alone sharing regime as it attempts to scale up to economically intensive settings, rational self-interest drives the formation of sharing arrangements to lower the transaction-cost burden attendant to a formal property-rights regime. Most fundamentally, property rights and sharing arrangements are not antithetical concepts. Contrary to the “new” conventional wisdom, *property rights are a complement to sharing arrangements*: that is, it is only by recourse to property rights or other exclusionary instruments that sharing arrangements can persist in economically intensive markets characterized by endowment heterogeneity, large numbers and high capital-intensity requirements. And, contrary to the “old” conventional wisdom, *sharing arrangements are a complement to property rights*: that is, it is only by recourse to sharing arrangements that innovator populations can substantially alleviate the transaction-cost burdens imposed by property rights or other exclusionary barriers.

THE ROLE OF THE RULE OF LAW IN VIRTUAL COMMUNITIES

Nicolas Suzor[†]

TABLE OF CONTENTS

I.	INTRODUCTION	1818
II.	THE DETERMINISTIC TREND IN CYBERLAW THEORY	1820
	A. THE DEVELOPMENT OF CYBERLAW THEORETICAL DISCOURSE.....	1821
	1. <i>The Flawed Assumptions in the Development of Cyberlaw Theoretical Discourse</i>	1825
	2. <i>Situating Cyberspace in Order to Overcome the Dichotomy Between Regulation and Autonomy</i>	1830
	B. USING THE RULE OF LAW TO BETTER CONCEPTUALIZE PRIVATE GOVERNANCE POWER	1836
III.	GOVERNANCE LIMITED BY LAW	1842
	A. RULE OF LAW LIMITATIONS ON ARBITRARY PUNISHMENT	1843
	B. PROTECTION OF SUBSTANTIVE INTERESTS	1845
	1. <i>Freedom of Expression Concerns: Discrimination, Speech, and Protest</i> ..	1850
	2. <i>Property Rights</i>	1856
	3. <i>Right to Privacy</i>	1861
	4. <i>Rights of Legal Enforcement</i>	1863
	5. <i>Summary of Substantive Values and a More General Application</i>	1865
IV.	FORMAL LEGALITY	1866
	A. PREDICTABILITY	1866
	1. <i>Clear Rules</i>	1867
	2. <i>Changing Rules</i>	1868
	3. <i>Emergent Behavior and Uncertain Rules</i>	1871
	4. <i>Inconsistent Application and Discretionary Enforcement</i>	1874
	B. PROCEDURAL FAIRNESS	1875

© 2010 Nicolas Suzor.

[†] Lecturer, QUT School of Law. I would like to extend my sincere gratitude to Kylie Pappalardo, Brian Fitzgerald, Greg Lastowka, Julie Cohen, Matthew Rimmer, Sal Humphreys, John Banks, Louise Longdin, Brendan Nerdal, Michael Risch, and participants at the *State of Play VI* Conference, June 2009, hosted by New York Law School, for comments on drafts of this paper.

V.	THE ROLE OF CONSENT AND DEMOCRACY.....	1878
VI.	CONCLUSION	1885

I. INTRODUCTION

As the Internet continues to develop and people become more invested in social networks and virtual communities, the role of private entities in governing these spaces becomes more important. Current cyberlaw theory and doctrine, however, is not particularly well-suited for analyzing the legitimacy of private governance. There is, accordingly, a substantial risk that emerging tensions in virtual communities will not be adequately recognized or addressed. This Article proposes a framework based upon rule of law theory through which to better conceptualize virtual community governance and suggest appropriate regulatory responses. The rule of law, as a discourse that highlights the potential for abuse of power and the legitimacy of governance, provides a particularly useful tool for examining the exercise of private power in virtual communities.

This project follows partially from A. V. Dicey's argument that in the absence of a substantive, written constitution, rule of law principles in the United Kingdom were protected by the evolution of private law doctrines that secured the substantive rights of citizens.¹ This work also builds upon Paul Berman's and Brian Fitzgerald's recognitions that public constitutional values are threatened by unrestrained private governance.² Essentially, this Article argues that if private law rules are used to regulate the governance of virtual communities, then those private rules should be influenced by public governance principles—specifically, those of the rule of law, which provide the most appropriate discourse on the regulation of governance power.

Part II of this Article examines the development of cyberlaw theory over the last two decades and argues that there is a severe tendency to delegitimize state intervention in private governance. In the mid-nineties, this delegitimization was accomplished predominantly by cyberspace exceptionalists, who argued that the Internet was so different from physical space that state laws should not apply. Gradually, this exceptionalism has given way to a recognition that while the Internet is regulable, the best mode

1. A. V. DICEY, INTRODUCTION TO THE STUDY OF THE LAW OF THE CONSTITUTION 187–88 (10th ed. 1959).

2. See Paul Schiff Berman, *Cyberspace and the State Action Debate: The Cultural Value of Applying Constitutional Norms to Private Regulation*, 71 U. COLO. L. REV. 1263, 1269 (2000); Brian F. Fitzgerald, *Software as Discourse: The Power of Intellectual Property in Digital Architecture*, 18 CARDOZO ARTS & ENT. L.J. 337, 384 (2000).

of regulation is generally to create and enforce strong property rights in internet resources in order to enable self-regulation. This Article argues that while autonomy is critical in virtual communities, both of these types of deterministic reasoning about governance are dangerous. By creating a false dichotomy between regulation and liberty, much of the current cyberlaw discourse risks misunderstanding the tensions that revolve around the legitimacy of governance in virtual communities.

In Part III, this Article examines substantive conceptions of the rule of law as they relate to the governance of virtual communities. As a first step, rule of law ideals suggest that we ought to be wary of claims that providers require absolute control and absolute discretion over a community. One of the oldest strands of the rule of law requires legal authorization for the exercise of power. Incorporating this insight into cyberspace self-governance implies that the contracts that underpin participation in virtual communities ought to be enforceable against the providers of those communities as well as the participants. This proposition highlights some shortcomings in the ways that these contracts are drafted. Namely, they are drafted overwhelmingly in favor of the providers, grant wide discretionary powers, and greatly limit any potential liability to the providers. If a restraint on the arbitrary exercise of power is warranted, there should be concern about the enforcement of such agreements as written.

Part III also considers the role of substantive external values in limiting the scope of cyberspace self-rule. This Part argues that the private law that is used to regulate private governance should be informed by public governance principles and that these substantive values should aid in determining appropriate limits to self-rule. This analysis canvasses a small number of substantive values: equality, freedom of speech, freedom of peaceful assembly, the right to privacy, protection of property, and rights of legal enforcement. While not all public governance principles should be directly applicable to virtual communities, rule of law ideals suggest that these principles should at least be taken into account when attempting to resolve tensions between participants and providers.

Part IV contrasts the modern liberal conceptions of the rule of law that revolve around formal legality with the uncertainties of governance in virtual communities. These modern ideals of the rule of law require that laws be clear, consistent, general, equal, and certain—characteristics that private contractual governance in virtual communities do not generally possess. Accordingly, Part IV argues that in communities where predictability is important, it may be desirable, at least in some circumstances, for territorial

courts not to defer to rules that fail to live up to these ideals. This Part also considers the role of due process and procedural fairness in the administration of virtual communities and suggests that states may be able to encourage more legitimate internal governance mechanisms by examining the exercise of discretion in exceptional circumstances.

Part V considers the role of consent in the governance of virtual communities. Since the theory supporting self-governance relies upon the consent of the participants in virtual communities to create better rules, we should be suspicious of contractual interpretations that conflict with internal norms. Where there is no conflict with substantive or formal values of the broader society, we ought to defer to the internal norms of the community in evaluating regulatory disputes. However, in cases where there is a conflict between internal consensual norms and a strict literal interpretation of the contractual terms of service, it may sometimes be appropriate for courts to refuse to uphold the contractual terms as written. This Part argues that if self-governance is encouraged for the creation of consensual rule sets, then providers may find themselves bound by the norms of the community they help to create, notwithstanding contractual provisions to the contrary.

This Article concludes that the rule of law discourse highlights important tensions in virtual communities that standard legal liberal contractual doctrine is unable to adequately address. As the role of private virtual community governance becomes greater in the lives of its participants, reliance on standard contractual doctrine risks marginalizing public governance values. In evaluating responses to disputes between participants and providers of virtual communities, it is desirable to read governance principles into the private law that bounds cyberspace self-rule. In doing so, significant care must be taken to ensure that no harm is unduly done to the autonomy of virtual communities. Any legal framework must be sensitive to the real needs of the participants and providers of virtual communities and should avoid regulatory solutions that diminish the value and potential of the community. As these governance issues are contextually sensitive, a significant degree of flexibility is required in determining appropriate legal responses. States should not, however, allow private governance to override core governance values in ways that are detrimental to the interests of their citizens.

II. THE DETERMINISTIC TREND IN CYBERLAW THEORY

The legitimacy of governance within virtual communities is not easily assessable within the framework of current cyberlaw theory. Over the last

two decades, the cyberlaw discourse has greatly evolved, but a deterministic trend remains that tends to delegitimize state interference in the governance of virtual communities. While this trend is most visible in the early cyber libertarian approaches, it persists as a set of flawed assumptions in the later anti-separatist theory. Section II.A traces the development of cyberlaw theory to highlight these deterministic tendencies and the false dichotomy that has emerged between regulation and autonomy. Section II.B then introduces the rule of law as a useful framework for evaluating regulatory approaches in a way that is sensitive both to the legitimacy of private governance and the importance of autonomy in the development of online communities.

A. THE DEVELOPMENT OF CYBERLAW THEORETICAL DISCOURSE

The early cyber libertarians argued that cyberspace was a new, different space—one devoid of scarcity, whose boundless possibilities would provide better rules than any state-made law.³ This utopian vision delegitimized the role of state law in regulating cyberspace and asserted that self-rule of autonomous virtual communities was both freer and more legitimate than any law imposed by the territorial state. This is best understood as a recognition of the malleability of cyberspace—the seductive opportunity to shape these brave new worlds into ideal communities.⁴ In 1996, John Perry Barlow famously declared the independence of cyberspace, calling it “the new home of Mind.”⁵ David Johnson and David Post followed in the legal literature, making the argument in their 1996 article, *Law and Borders*, that “the fundamental principle” of internet governance should be that

[i]f the sysops and users who collectively inhabit and control a particular area of the Net want to establish special rules to govern conduct there, and if that rule set does not fundamentally impinge upon the vital interests of others who never visit this new space, then the law of sovereigns in the physical world should defer to this new form of self-government.⁶

3. See David R. Johnson & David G. Post, *Law and Borders—The Rise of Law in Cyberspace*, 48 STAN. L. REV. 1367 (1995) [hereinafter Johnson & Post, *Law and Borders*]; David G. Post, *The Unsettled Paradox: The Internet, the State, and the Consent of the Governed*, 5 IND. J. GLOBAL LEGAL STUD. 521 (1997) [hereinafter Post, *Unsettled Paradox*]; John Perry Barlow, *A Declaration of the Independence of Cyberspace* (Feb. 8, 1996), <http://homes.eff.org/~barlow/Declaration-Final.html>; David G. Post & David R. Johnson, *The Great Debate*, 11 FIRST MONDAY (Feb. 2006), <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1311/1231> [hereinafter Post & Johnson, *The Great Debate*].

4. Julie E. Cohen, *Cyberspace as/and Space*, 107 COLUM. L. REV. 210, 217 (2007).

5. Barlow, *supra* note 3.

6. Johnson & Post, *Law and Borders*, *supra* note 3, at 1393.

In this ideal world, whenever the rules that govern participation in any given community become undesirable, a user has a practically unfettered ability to move to another community with a different rule set, or to create and grow a new community.⁷ The ability to easily move in and out of virtual communities will create a market for rule sets, resulting in rules that are more responsive to the demands of participants. This allows participants to self-select into communities whose rule sets more closely reflect their needs and desires.⁸ The lack of scarcity and ease of exit in virtual communities provides “a more legitimate ‘selection mechanism’ by which differing rule sets will evolve over time.”⁹

This exceptionalist treatment of cyberspace as completely separate from physical space gave way, largely, to the recognition that cyberspace was subject to the same regulatory forces as physical space, and indeed, was no different from physical space.¹⁰ It became clear that the utopian libertarian dream was premised not on self-governance and the delegitimization of the state but upon the creation and maintenance of state-granted property rights.¹¹ Closer analysis of the software code through which communication was mediated showed not only that the architecture of cyberspace had a regulatory function,¹² but that it could, and in some cases should, be bent to the will of the state.¹³ The recognition that cyberspace was already regulated

7. *Id.* at 1383, 1398–99.

8. See Post, *Unsettled Paradox*, *supra* note 3, at 539 (arguing that the decentralized generation of law online is made by “the aggregate of the choices made by individual system operators about what rules to impose, and by individual users about which online communities to join”).

9. Johnson & Post, *Law and Borders*, *supra* note 3, at 1398–99.

10. See James Boyle, *Foucault in Cyberspace: Surveillance, Sovereignty, and Hardwired Censors*, 66 U. CIN. L. REV. 177, 178 (1997); Jack L. Goldsmith, *Against Cyberanarchy*, 65 U. CHI. L. REV. 1199 (1998) (challenging both normative and descriptive claims against public regulation of cyberspace); Lawrence Lessig, *The Law of the Horse: What Cyberlaw Might Teach*, 113 HARV. L. REV. 501 (1999) (explaining four modalities of regulation of cyberspace); Margaret Jane Radin & R. Polk Wagner, *The Myth of Private Ordering: Rediscovering Legal Realism in Cyberspace*, 73 CHI.-KENT L. REV. 1295 (1998) (pointing out that self-governance in cyberspace is always reliant on background legal rules); Timothy Wu, *Application-Centered Internet Analysis*, 85 VA. L. REV. 1163 (1999) (discussing public and private regulation of internet applications).

11. See Radin & Wagner, *supra* note 10, at 1296–97.

12. See Lawrence Lessig, *The New Chicago School*, 27 J. LEGAL STUD. 661, 676 (1998) [hereinafter Lessig, *The New Chicago School*]; see also LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE (1999) [hereinafter LESSIG, CODE] (describing the interaction of code, norms, the market, and the law in the regulation of cyberspace).

13. See Graham Greenleaf, *An Endnote on Regulating Cyberspace: Architecture vs. Law*, 21 U. NEW S. WALES L.J. 593 (1998); Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553, 557 (1998).

and always regulable struck a blow to the utopian vision and severely damaged the project to isolate cyberspace from the interference of the territorial state.

In the place of a utopian technological determinism, however, rose a determinism of market rule—a suggestion that cyberspace could be best regulated through the creation and enforcement of strong and clear property rights.¹⁴ The exceptionalist nature of cyberspace had disappeared, but the end result was very similar: state interference in the governance of cyberspace was delegitimized in the name of autonomy and innovation.¹⁵ Judge Frank Easterbrook signalled the beginnings of this change, in a famous exhortation to cyberlaw scholars in 1996, when he argued that the risk of legal error in regulating cyberspace meant that the best regulatory approach would be to create new property rights, allowing for efficient bargaining between users.¹⁶ For Easterbrook, if rules are clear, if strong property rights exist, and if institutions can be created to facilitate bargaining, then Coasean determinism will prevail and an efficient result will emerge irrespective of the initial allocation of entitlements.¹⁷ If society could just let “the world of cyberspace evolve as it will,” everyone could simply “enjoy the benefits.”¹⁸

Allowing virtual communities to determine their own rules is intuitively appealing to the liberal ideal of autonomy and self-determination. The claim is that “cyberspace self-governance more fully embodies the liberal democratic goals of individual liberty, popular sovereignty, and the consent of the governed than does the top-down administration of even the most democratic nation states.”¹⁹ In an ideal world, individuals will be able to self-select into communities that reflect their needs and desires, thus allowing a

14. See Frank H. Easterbrook, *Cyberspace and the Law of the Horse*, 1996 U. CHI. LEGAL F. 207, 212 (1996); Richard A. Epstein, *Intellectual Property: Old Boundaries and New Frontiers*, 76 IND. L.J. 803, 818–19 (2001); Trotter Hardy, *Property (and Copyright) in Cyberspace*, 1996 U. CHI. LEGAL F. 217, 236 (1996).

15. See F. Gregory Lastowka & Dan Hunter, *The Laws of the Virtual Worlds*, 92 CALIF. L. REV. 1, 72–73 (2004) (concluding that while virtual worlds are subject to legal regulation, courts should “recognize that virtual worlds are jurisdictions separate from our own” in order to allow internal governance to develop); see also R. Polk Wagner, *On Software Regulation*, 78 S. CAL. L. REV. 457 (2004) (arguing that creating strong property rights supports the desirable development of cyberspace self-regulation).

16. Easterbrook, *supra* note 14, at 212.

17. *Id.* at 212–13 (citing R. H. Coase, *The Federal Communications Commission*, 2 J.L. & ECON. 1, 212 (1959)).

18. *Id.* at 216.

19. Neil Weinstock Netanel, *Cyberspace Self-Governance: A Skeptical View from Liberal Democratic Theory*, 88 CALIF. L. REV. 395, 402 (2000).

range of diverse communities to cater to each individual taste.²⁰ In this ideal world, the norms that develop in virtual communities are generally *better* than any law that could be imposed by the state because they can be tailored for and by the participants themselves.²¹ Johnson and Post argued that individuals are “more likely to be in possession of the relevant information regarding . . . their own welfare” than will elected officials.²² Therefore, individuals can use their ability to enter and exit virtual communities to reflect their needs and desires, potentially resulting in rule sets that can react faster and more flexibly to changing environments and externalities imposed by other communities.²³

This emphasis on self-determination has been taken up in the virtual worlds literature, particularly by Lastowka and Hunter.²⁴ Their arguments express the concern that “the complexity of ascertaining a virtual world’s emerging legal rules and balancing them” with participant and provider interests will result in bad decisions by real-world courts on virtual disputes.²⁵ As virtual communities develop their own rules, “[c]ourts will need to recognize that virtual worlds are jurisdictions separate from our own, with their own distinctive community norms, laws, and rights.”²⁶ As these (“cyborg”) communities develop, the role of territorial law will fade: “If these attempts by cyborg communities to formulate the laws of virtual worlds go well, there may be no need for real-world courts to participate in this process. Instead, the residents of virtual worlds will live and love and law for themselves.”²⁷

Other academics have noted that allowing providers to create expressive or entertaining spaces requires substantial autonomy to determine internal

20. Johnson & Post, *Law and Borders*, *supra* note 3, at 1398; Post, *Unsettled Paradox*, *supra* note 3, at 539; Post & Johnson, *The Great Debate*, *supra* note 3.

21. Post & Johnson, *The Great Debate*, *supra* note 3.

22. David G. Post & David R. Johnson, *Chaos Prevailing on Every Continent: Towards a New Theory of Decentralized Decision-Making in Complex Systems*, 73 CHI.-KENT L. REV. 1055, 1087–88 (1997).

23. *See id.*

24. Lastowka and Hunter were careful to distinguish Johnson and Post’s “precyberskeptical ambitious thinking about ‘cyberspaces’ as a separate jurisdiction” on the basis that the Internet, more broadly, “never became an independent community.” For Lastowka and Hunter, law and self-rule would only evolve where there was a real community, and the best example of new communities forming was in virtual worlds, which meant that “the emergence of virtual law within those worlds [was] much more likely.” Lastowka & Hunter, *supra* note 15, at 69.

25. *Id.* at 71.

26. *Id.* at 73.

27. *Id.*

norms.²⁸ This approach has also found favor in law and economics discourse. Richard Epstein makes the argument that an absolute right to exclude, which centralizes the power in the hands of the provider, forms the basis for the development of private rules and that “private voluntary arrangements will outperform forced interactions in the long run.”²⁹

Each of these arguments shares a common thread, specifically that governance by the local virtual community is likely to be better than rules imposed from external sources. Many of these arguments for self-governance are based upon ideal world assumptions where there is little to no scarcity, where participants can come and go without friction, where new communities can quickly and cheaply be established when existing rule sets are no longer appropriate, and where participants are empowered to choose communities whose rules suit their needs and desires. In a non-ideal world, these assumptions are all suspect; there are significant limits to self-governance that must be addressed in any regulatory framework.³⁰

1. *The Flawed Assumptions in the Development of Cyberlaw Theoretical Discourse*

First is the problem of exit. Fundamentally, the assumption that a marketplace for norms will emerge is severely limited if participants are not able to easily leave one community for another.³¹ Providers of virtual communities, however, have an incentive to make the community difficult to leave.³² Subscription and ad-supported communities earn more revenue for each participant and most communities benefit from the network effects of

28. See Jack M. Balkin, *Virtual Liberty: Freedom to Design and Freedom to Play in Virtual Worlds*, 90 VA. L. REV. 2043, 2046 (2004) (discussing the freedom to design); Edward Castronova, *The Right to Play*, 49 N.Y.L. SCH. L. REV. 185, 196 (2004).

29. Epstein, *supra* note 14, at 819; see also Richard A. Epstein, *Intel v. Hamidi: The Role of Self-Help in Cyberspace*, 1 J.L. ECON. & POL'Y 147, 157 (2005) (arguing that supporting self-help through strong property rights provides more efficient outcomes than state regulation of speech).

30. See Niva Elkin-Koren, *Copyrights in Cyberspace—Rights Without Laws*, 73 CHI.-KENT L. REV. 1155, 1166 (1997) (critiquing the technologically deterministic economic predictions that self-rule will result in better norms); Netanel, *supra* note 19 (arguing that cyberspace self-rule should be limited to enhance liberal values).

31. See Dan L. Burk, *Virtual Exit in the Global Information Economy*, 73 CHI.-KENT L. REV. 943 (1997); Netanel, *supra* note 19, at 426.

32. See Sal Humphreys, “You’re In Our World Now” *Ownership and Access in the Proprietary Community of an MMOG*, in INFORMATION COMMUNICATION TECHNOLOGIES AND EMERGING BUSINESS STRATEGIES 76, 85 (Shenja Van Der Graaf & Yuichi Washida eds., 2007) (arguing that “[t]he stronger the ties, the longer the engagement, and the longer the monthly subscription rolls in for the publisher”).

having more participants. As such, the total utility increases significantly with each additional connected individual.³³

Thus, exit is nowhere near frictionless. A participant's ability to leave a community is constrained by the social connections she has developed or strengthened with other people, with whom she would lose an important point of contact, context, and common interest.³⁴ Any investment she has made in social capital, reputation, or virtual property within the community, none of which is easily transferable to other communities, makes it harder for her to leave. Further, and this hints at the next problem, exit is constrained by the availability of other communities that offer reasonably substitutable experiences.

The second problem is the considerable barriers to the creation of new communities. When Johnson and Post were first writing in the mid-nineties, barriers to the establishment of new communities were reasonably low, as it was relatively trivial to create a new channel or new server for Internet Relay Chat, a new Usenet newsgroup, or a new text-based virtual world.³⁵ The ease with which new services could be offered suggested that any harm caused by poor governance could readily be overcome by joining or creating a new community. Modern virtual communities, however, are much less readily created. Millions of dollars of investment in coding, artwork, testing, and marketing go into the creation of large-scale virtual worlds. Even where communities can be built on relatively simple technology, they will often fail to reach or maintain the critical mass required to sustain a large-scale community. Small-scale communities may still be created relatively trivially, but the importance of network effects generally ensures that these

33. Michael L. Katz & Carl Shapiro, *Network Externalities, Competition, and Compatibility*, 75 AM. ECON. REV. 424 (1985) (defining network externalities); S. J. Liebowitz & Stephen E. Margolis, *Network Externality: An Uncommon Tragedy*, 8 J. ECON. PERSP. 133 (1994) (explaining network externalities with respect to software and computer networks); see also danah m. boyd & Nicole B. Ellison, *Social Network Sites: Definition, History, and Scholarship*, 13 J. COMPUTER-MEDIATED COMM. 210, 218 (2008) (explaining that most, but not all, social networking sites aim for exponential growth).

34. T. L. TAYLOR, *PLAY BETWEEN WORLDS: EXPLORING ONLINE GAME CULTURE* 135 (2006) (explaining that despite participants' opposition to structures of technological systems, it is not easy to refrain from participating); Sal Humphreys, *Ruling the Virtual World: Governance in Massively Multiplayer Online Games*, 11 EUR. J. CULTURAL STUD. 149, 163 (2008).

35. Johnson & Post, *Law and Borders*, *supra* note 3, at 1395 ("The ease with which internal borders, consisting entirely of software protocols, can be constructed is one of Cyberspace's most remarkable and salient characteristics; setting up a new Usenet newsgroup, or a 'listserver' discussion group, requires little more than a few lines of code.").

communities remain at the fringes, rarely attracting enough participants to seriously compete with larger communities.³⁶

The third major problem with the ideal of a marketplace of norms is that it does not exist. Firms that provide virtual communities tend to draft contracts that are not designed to be easily read or understood by subscribers.³⁷ These contracts are typically dense, long, full of legalese, and presented in a way that discourages readers from actually reading the contract.³⁸ Individual subscribers are at a significant disadvantage compared to the providers, who have the ability to amortize the high costs of understanding, drafting, and changing these agreements over a very large number of transactions.³⁹ Finally, perhaps because subscribers are discouraged from reading and understanding the terms of service, there is very little competition in the market. Consequently, there is a high degree of homogeneity in the terms of service available from the various providers of virtual communities.⁴⁰

These factors illustrate the assumption that cyberspace self-governance will always provide better results than externally imposed regulation is deeply flawed. It may be that virtual communities will develop legitimate rules for themselves.⁴¹ Nevertheless, there can be no guarantee that providing strong property and contract rights and allowing communities to govern themselves will necessarily lead to desirable outcomes.⁴² Indeed, legitimate self-

36. TAYLOR, *supra* note 34, at 135 (“We might also consider the ways participating in particular forms or places always are tied up with questions of power. Separate does not mean equal, and sometimes we can see quite clearly the benefits that come from being in particular spaces.”).

37. See Dale Clapperton & Stephen Corones, *Unfair Terms in “Clickwrap” and Other Electronic Contracts*, 35 AUSTL. BUS. L. REV. 152 (2007).

38. See Margaret Jane Radin, *Regulation by Contract, Regulation by Machine*, 160 J. INSTITUTIONAL & THEORETICAL ECON. 142 (2004); see also Humphreys, *supra* note 34, at 165 (“Many players never read the EULA or Terms of Service, lengthy documents written in legal discourse impenetrable to most of the world outside the legal profession.”).

39. Victor P. Goldberg, *Institutional Change and the Quasi-Invisible Hand*, 17 J.L. & ECON. 461, 483–85 (1974); see also Clapperton & Corones, *supra* note 37; Andrew Robertson, *The Limits of Voluntariness in Contract*, 29 MELB. U. L. REV. 179, 194 (2005) (explaining the imbalanced bargaining power between the consumer and the provider).

40. Andrew Jankowich, *EULAn: The Complex Web of Corporate Rule-Making in Virtual Worlds*, 8 TUL. J. TECH. & INTELL. PROP. 1, 49 (2006); Radin & Wagner, *supra* note 10, at 1311–12.

41. Johnson & Post, *Law and Borders*, *supra* note 3, at 1388 (arguing that providers and subscribers “have begun explicitly to recognize that formulating and enforcing such rules should be a matter for principled discussion, not an act of will by whoever has control of the power switch”).

42. Netanel concludes:

governance is an extremely difficult ongoing process.⁴³ Theorists have demonstrated that of the four modalities of regulation identified by Lessig (code, law, the market, and social norms),⁴⁴ none are value neutral, and none can be relied upon to provide utopian results.⁴⁵

While the assumptions that underpin both the technological determinism of the exceptionalists and the Coasean market determinism of the law and economics scholars have not gone unchallenged,⁴⁶ there remains some conceptual difficulty that surrounds the autonomy of online communities. The normative claim that online communities can develop better norms for interaction than state-imposed rules⁴⁷ still holds some weight. Autonomy continues to be regarded as crucially important from a number of perspectives: that there is a fundamental right to design communities and to immerse oneself in spaces where the normal rules of the corporeal world do

An untrammelled cyberspace would ultimately be inimical to liberal democratic principles. It would free majorities to trample upon minorities and would serve as a breeding ground for invidious status discrimination, narrowcasting and mainstreaming content selection, systematic invasions of privacy, and gross inequalities in the distribution of basic requisites for netizenship and citizenship in the information age.

Netanel, *supra* note 19, at 498.

43. See A. Michael Froomkin, *Habermas@Discourse.Net: Toward a Critical Theory of Cyberspace*, 116 HARV. L. REV. 749 (2003) (demonstrating that the effort required to create a legitimate consensual governance regime was substantial, leading to an inference that not all communities will invest in the normative discourse required to create a workable and fair system).

44. Lessig, *The New Chicago School*, *supra* note 12, at 662–63.

45. See generally Mark A. Lemley, *The Law and Economics of Internet Norms*, 73 CHI.-KENT L. REV. 1257, 1292 (1997) (expressing concern about the legitimacy and lack of restraint on those enforcing non-legal norms); Netanel, *supra* note 19 (highlighting the inescapable inefficiencies and the substantial imbalances of power in the market); Radin & Wagner, *supra* note 10 (explaining that the legal property and contractual rights that underpin self-governance claims entrench particular decisions about the definition and allocation of entitlements); see also Cohen, *supra* note 4, at 255 (“Many important questions have tended to slip between the cracks in an analytical universe that seeks to unpack ‘code’ while taking ‘law,’ ‘norms,’ and ‘the market’ for granted.”).

46. See LESSIG, *CODE*, *supra* note 12; Julie E. Cohen, *Lochner in Cyberspace: The New Economic Orthodoxy of “Rights Management,”* 97 MICH. L. REV. 462 (1998); Elkin-Koren, *supra* note 30; William W. Fisher III, *Property and Contract on the Internet*, 73 CHI.-KENT L. REV. 1203 (1998); F. Gregory Lastowka & Dan Hunter, *Virtual Crimes*, 49 N.Y.L. SCH. L. REV. 293 (2004); Lemley, *supra* note 45; Jonathan F. Fanton, *Rights and Responsibilities Online: A Paradox for Our Times*, 13 FIRST MONDAY (Aug. 9, 2008), <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2196>.

47. See generally Johnson & Post, *Law and Borders*, *supra* note 3 (arguing that new rules will arise in virtual communities that differ from territorial, state-based rules, and that these rules should govern in virtual spaces).

not apply,⁴⁸ that individuals with similar interests should be able to bargain for their own rules of association,⁴⁹ that vibrant communities need regulatory freedom to create their own interesting norms,⁵⁰ and that flexibility to determine rules is of paramount necessity for “maintaining and improving the environment for innovation, experimentation, and entrepreneurship.”⁵¹

This emphasis on autonomy is balanced, to an extent, by a recognition that internal governance should be limited in certain circumstances. What exactly these circumstances are is equally varied. Some form of property right in virtual assets is often thought to be worth protecting by territorial states.⁵² At other times, theorists have suggested that limits on the ability of providers to control speech may be appropriate, at least where the borders (particularly the economic borders) between actions in the community and the real world are porous.⁵³ There have also been suggestions that the consent expressed by the contractual rules that bind participants should be procedurally protected from anti-competitive behavior⁵⁴ and “force and fraud.”⁵⁵ Unfortunately, however, no easy way to reconcile the need for autonomy with the disparate legitimate interests of participants has emerged.

This tension between autonomy and regulation has led to the emergence of a false dichotomy. It seems to be generally understood that states can (and sometimes should) impose their will on the providers of virtual communities, but that doing so is likely to limit the ability of communities to develop consensual norms and cause harm to the diversity and vibrancy of online spaces that makes them particularly attractive in the first place. The result is

48. Balkin, *supra* note 28, at 2043, 2062; Castronova, *supra* note 28, at 202.

49. Richard A. Epstein, *Cybertrespass*, 70 U. CHI. L. REV. 73, 81 (2003); Epstein, *supra* note 14.

50. Lastowka & Hunter, *supra* note 15, at 61, 72–73; *see also* Richard A. Bartle, *Virtual Worldliness: What the Imaginary Asks of the Real*, 49 N.Y.L. SCH. L. REV. 19, 22 (2004) (arguing that freedom to design and regulate is fundamentally important in virtual communities).

51. Wagner, *supra* note 15, at 506.

52. *See generally* Joshua A. T. Fairfield, *Virtual Property*, 85 B.U. L. REV. 1047 (2005) (arguing that creating property rights in virtual assets can help prevent anticommons that arise from fragmentation of rights to exclude); Andrew E. Jankowich, *Property and Democracy in Virtual Worlds*, 11 B.U. J. SCI. & TECH. L. 173 (2005) (arguing that creating enforceable property rights in virtual assets is necessary to help structure relationships and resolve disputes between participants and providers in virtual worlds); Lastowka & Hunter, *supra* note 15 (arguing that property theory generally supports creating property rights in virtual assets).

53. Balkin, *supra* note 28, at 2090; Castronova, *supra* note 28, at 204 (arguing for a strong legal distinction between play worlds and worlds with a porous economy).

54. Epstein, *supra* note 14, at 819.

55. Joshua A. T. Fairfield, *Anti-Social Contracts: The Contractual Governance of Virtual Worlds*, 53 MCGILL L.J. 427, 468 (2008).

that state intervention in online community governance continues to be delegitimized, even as the harm caused to participants in examples of bad governance becomes more visible.

2. *Situating Cyberspace in Order to Overcome the Dichotomy Between Regulation and Autonomy*

A theoretical approach that is more sensitive to the ways in which different sources of regulatory power interact in online communities should be adopted in order to move beyond the dichotomy between regulation and autonomy. A key problem with the current cyberlaw governance discourse is that it has been largely “predicated on a teleology of disembodiment”⁵⁶ that isolates participation in cyberspace from the remainder of lived experience. Much of current cyberlaw discourse positions law and the market as bounding forces that structure isolated zones of liberal self-governance. This discourse is flawed because it tends to focus on the existence and operation of rights to exclude based on property or contract law as the borders of acceptable regulation, particularly between public regulation and private governance.

This is particularly true in debates that center around cyberproperty and cybertrespass, where the analysis generally focuses on whether an enforceable right to exclude exists, but rarely considers the effect of such a right on community governance. There is a tendency to characterize such rights as absolutes, notwithstanding that their offline analogs are highly contextually sensitive and contain numerous complicated exceptions.⁵⁷ The mere existence of a right to exclude tells very little about any limitations that may be imposed on the exercise of such a right.⁵⁸

The cyberproperty debate is not, however, the only part of cyberlaw theory that maintains a relatively sharp dichotomy between regulated and unregulated zones of self-governance. These distinctions are also quite popular in the virtual worlds discourse where theorists often try to separate

56. Cohen, *supra* note 4, at 255.

57. Michael A. Carrier & Greg Lastowka, *Against Cyberproperty*, 22 BERKELEY TECH. L.J. 1485, 1498 (2007) [hereinafter Carrier & Lastowka, *Against Cyberproperty*]; see also F. Gregory Lastowka, *Decoding Cyberproperty*, 40 IND. L. REV. 23, 46–47 (2007) [hereinafter Lastowka, *Decoding Cyberproperty*] (questioning the assumption that private property rights in “cyberspace” is the best means of promoting the public good).

58. Carrier & Lastowka, *Against Cyberproperty*, *supra* note 57, at 1508–09 (arguing that limits are fundamentally important to property, but conceptions of cyberproperty tend not to include limits).

games and expressive spaces from more quotidian platforms.⁵⁹ In these conceptions, social spaces that allow property or cross-border real-money trades are typically treated as regulable, whereas “play” spaces are held to remain free from state interference.⁶⁰ The desire to protect the integrity of play or expressive spaces is understandable; there is clearly a threat that an overly limited capacity to mold the community experience will greatly jeopardize the enjoyableness or expressiveness of the spaces.⁶¹ While these concerns are significant, however, it does not necessarily follow that the providers of all such spaces require absolute power over the community. Some tensions will obviously be less relevant in play and expressive spaces, but there are still legitimate concerns about other potential abuses of private power.⁶²

Perhaps the most pronounced example of this disembodied dichotomy is the tendency to resort to contractual doctrine as a model for evaluating disputes in virtual communities.⁶³ While recognizing that contractual limits do exist and apply, this model obviates the need to evaluate internal governance within those boundaries. So long as the contractual documents that purport to govern participation are upheld, then regulating governance becomes a simple matter of contractual interpretation. Unfortunately, a predominantly private contractual model of governance imports all the familiar baggage of liberal contract theory and does a poor job of structuring the potentially conflicting interests of providers and participants in virtual

59. See Balkin, *supra* note 28, at 2072; Castronova, *supra* note 28, at 204; Lastowka & Hunter, *supra* note 15, at 70–72 (contrasting cyberspace in general with the claims for autonomy made by virtual communities); James Grimmelman, *Virtual World Feudalism*, YALE L.J. POCKET PART (Jan. 18, 2009), <http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/property-law/virtual-world-feudalism/>.

60. See generally Joshua A. T. Fairfield, *The Magic Circle*, 11 VAND. J. ENT. & TECH. L. 823 (2008) (explaining and critiquing the concept of the “magic circle,” a metaphor used to isolate play spaces from the “real world”).

61. See Bartle, *supra* note 50, at 27; Castronova, *supra* note 28, at 202.

62. See TAYLOR, *supra* note 34, at 19 (“In much the same way we now see the relationship between on- and offline life as not a bounded one, in many ways a game/not-game dichotomy does not hold.”).

63. Fairfield, *supra* note 55, at 435 (noting that in virtual worlds, “questions of property law, tort law, and even criminal law are uniformly construed by the courts as contract disputes”). Fairfield hints at the inadequacy of reducing governance to a contractual framework by arguing that private contract law is unable to provide the stable default rules that societies need to govern interaction between participants. Fairfield’s analysis, however, focuses on the horizontal relationships between participants, and does not consider the relationship between participants and providers in any great detail. Fairfield does consider vertical relationships, but mainly notes that contractual law makes it much more difficult to define clear ownership rights in virtual property than does property law itself. *Id.* at 454–57.

communities. This model makes problematic assumptions about the way that individual participants bargain and contract to enter communities: rationally evaluating risk and retaining at all times the consumer sovereignty of being able to simply leave a community whose governance structure becomes objectionable. It imposes an assumption of market determinism that participants will express their demand for certain rule sets and this demand will be satisfied through standard economic forces, as long as property and contract rights are sufficiently well-defined and easily transferable.⁶⁴ Most importantly, perhaps, it reduces community participation to simple consumer transactions, which tends to downplay or ignore the set of tensions that revolve around the legitimacy of governance.⁶⁵

The critical insight here is that the dichotomy between absolute self-rule and a complete lack of autonomy is false. The borders of regulation are much more complex and interesting than is typically recognized,⁶⁶ and “[t]o admit only dreams of total freedom or total control seems too limiting.”⁶⁷ By continuing to conceptualize cyberspace governance as isolated zones of liberal self-rule, simply bounded by contractual doctrine, the tensions that revolve around private governance risk being misunderstood. Julie Cohen, drawing from science and technology studies, argues:

[T]he processes that construct power in networked space] are social and emergent, and have consequences both spatial and material. They operate in what Saskia Sassen terms “analytic borderlands”: between public and private, between technical and social, and between network and body. Mapping these borderlands requires descriptive and analytical tools that do not simply reduce them to borders.⁶⁸

This argument appears to be fundamentally correct. The relationships of power within virtual communities are important because the people within those communities “are real people, not simply disembodied virtual users.”⁶⁹ The contested interplay between the various forces at work, the borderlands of regulation, is of primary importance to the construction of power in

64. See Easterbrook, *supra* note 14, at 209–16; Epstein, *supra* note 14; Hardy, *supra* note 14, at 219, 236–58 (discussing transaction costs and costs of drawing boundary lines).

65. See Nicolas Suzor, *On the (Partially) Inalienable Rights of Participants in Virtual Communities*, 130 MEDIA INT’L AUSTL. 90 (2009).

66. See Radin & Wagner, *supra* note 10, at 1297.

67. Cohen, *supra* note 4, at 224.

68. *Id.* at 251 (citing SASKIA SASSEN, TERRITORY, AUTHORITY, RIGHTS: FROM MEDIEVAL TO GLOBAL ASSEMBLAGES 379–86 (2006)).

69. *Id.* at 221.

cyberspace.⁷⁰ These borderlands are the sites of the power struggles that directly affect the interests of the real people who participate in these communities, and “[t]he emergent geographies of power within networked space shape the conditions of possibility, the conditions of participation, and the conditions of material existence.”⁷¹

These emergent geographies of power in networked space need a different approach than the liberal framework through which the law typically views power relations. The core problem with the traditional legal liberal approach is that it tends to ignore the existence and flow of power in the private sphere. Brian Fitzgerald explains:

Traditionally, constitutionalism (which means the regulation of power) has focused on regulating or limiting the vertical exercise of government or public power over the citizen. On the other hand, the horizontal exercise of power between citizens has occurred in the private sphere and has been rarely analyzed in terms of power or constitutionalism, although the (largely common) law has played a mediating role.⁷²

Fitzgerald concludes that “[p]ower relations in the private sphere . . . are fundamental constitutional issues that should be informed by fundamental constitutional principles.”⁷³ This notion, also known as “constitutive constitutionalism” allows society to grapple with constitutional values that are otherwise marginalized by the public-private divide.⁷⁴ These public community values are sometimes threatened by the private exercise of power in cyberspace (and elsewhere), and these concerns should be explicit:

70. However, Epstein argues that we should address “the . . . basic outlines” of property rules first, rather than the details:

The success and the glory of any legal system is not how it resolves hard marginal cases, but rather how it sets out the rules that allow most routine transactions to go from cradle to grave without so much as a hint of litigation. . . . All the while, we must remember that even if sound legal principles do not eliminate every anomaly or answer every single question of system design, they can help us avoid major errors that could carry with them disastrous social consequences. We can live with gray areas, so long as we have black and white, but we cannot live with fundamental flaws in system design.

Epstein, *supra* note 14, at 827; *see also* Easterbrook, *supra* note 14, at 211 (arguing that the risk of legal error justifies the granting of property rights that are easy to reverse in private transactions, rather than attempting to determine the optimal allocation of entitlements).

71. Cohen, *supra* note 4, at 255.

72. Fitzgerald, *supra* note 2, at 382.

73. *Id.* at 384.

74. Berman, *supra* note 2, at 1269.

[I]f it is true that we already think of the Constitution as embodying such constitutive values of our society, it may seem quite natural to use the Constitution as a touchstone for evaluating a broader range of social interaction. Moreover, an argument based on constitutive constitutionalism may also be particularly persuasive in the context of debating online regulation, because in cyberspace it is perhaps easier to see how private entities can threaten cherished constitutional norms.⁷⁵

The point is not to directly extend constitutional regulation to the governance of virtual communities, as there are of course many important respects in which a virtual community is not like a real state and should not be regulated as one. What is important is the direct confrontation of constitutional values, while considering how they can inform the current regulatory discourse. It is these public values that are most under threat by private governance, and it certainly seems desirable to examine much more closely what effects marginalizing these values have on the people who participate in these spaces.

The attractiveness of cyberspace, its seductive appeal, is largely based upon the explicit promise of malleability—the largely unbounded choices that shape the world to be inhabited. The important question has been “what kinds of alternate social orderings do we imagine and seek to enable?”⁷⁶ Julie Cohen, however, makes explicit the second part of this question: “[w]hich attributes of real space do we seek to perfect and harness in the service of utopian ambitions?”⁷⁷ Cyberspace is neither wholly distinct nor wholly similar to regular space.⁷⁸ Answering these questions requires direct consideration of the constitutive limits that shape power relations in cyberspace, with the explicit goal of determining whether they are appropriate for the spaces society is trying to construct. One glaring omission in current regulatory approaches is the limitation of the exercise of power, leading to the fear expressed by a number of theorists about the potential for the rampant abuse of “private” power in a system that predicates legitimacy on a sharp distinction between public and private spheres.⁷⁹

75. *Id.* at 1270.

76. Cohen, *supra* note 4, at 222; see also Dan Hunter, *Cyberspace as Place and the Tragedy of the Digital Anticommons*, 91 CALIF. L. REV. 439, 443–44 (2003); Johnson & Post, *Law and Borders*, *supra* note 3, at 1378–79.

77. Cohen, *supra* note 4, at 222.

78. *Id.* at 219–21.

79. See LAWRENCE LESSIG, *CODE VERSION 2.0*, at 285–93 (2006) (contrasting the emergence of private norms with public governance and the potential tensions where those

The constitutional discourse serves to highlight some of what may be lacking in virtual governance. Contrasting the values believed to be important in corporeal states with the way in which virtual communities are governed can help to identify potentially desirable regulatory approaches.⁸⁰ In looking at virtual communities experientially to see both how they are different from real spaces and how they are the same, one of the striking realizations is that limitations on the exercise of power are conspicuously absent. Thus, as the use of virtual communities grows in importance in all aspects of a citizen's life, the public law of the state is slowly replaced by the private "law" of the provider.⁸¹ There is substantial danger in a world where contractual regimes promulgated by firms are enforced as written by the courts and largely accepted as effective by both participants and providers:

If we continue assuming . . . that the mass-market contractual regime is efficacious, then it is obvious that for a large subset of the social order . . . the law of the state . . . has been superseded by the promulgated contractual regime, the "law" of the firm. In the limiting case . . . the official constitutional/legislative/judicial regime is completely irrelevant. In situations short of the limiting case, but in which large numbers of people are subject to these superseding regimes, the official constitutional/legislative/judicial regime is severely eroded or marginalized.⁸²

As this process continues, important constitutional principles may begin to fade in relevance, to our collective detriment. In these cases, it may be more desirable to attempt to read these values into the regulatory framework that bounds self-governance in virtual communities.

norms are not just); Elkin-Koren, *supra* note 30, at 1186–87; Netanel, *supra* note 19, at 482–83; Radin, *supra* note 38, at 146–47.

80. See JONATHAN ZITTRAIN, *THE FUTURE OF THE INTERNET AND HOW TO STOP IT* 174 (2008) (arguing that we need discussion and "lawyers who can help translate the principles of fairness and due process that have been the subject of analysis for liberal democracies into a new space where private parties and groups come together with varying degrees of hierarchy to try to solve the problems . . . in the digital space"); see also Fanton, *supra* note 46 (analyzing the "Internet's democratic promise and lack of democratic protections").

81. See Cohen, *supra* note 4, at 221 (arguing that as cyberspaces "increasingly replace (or displace) their real-space analogues, the rules governing them become increasingly important"); Humphreys, *supra* note 34, at 166 (arguing that "[a]s people with access to these technologies come to live more of their social lives (and work lives) in online environments, and to construct both their identities and communities in proprietary spaces, the terms under which they do so will become increasingly important"); Radin, *supra* note 38.

82. Radin, *supra* note 38, at 6 (using the extended propertization of copyright as an example).

B. USING THE RULE OF LAW TO BETTER CONCEPTUALIZE PRIVATE GOVERNANCE POWER

One of the most concerning characteristics of private governance in virtual communities is that it is very seldom transparent, clear, or predictable, and providers often purport to have absolute discretion on the exercise of their power to eject participants under both contract and property law.⁸³ If the absolute discretion of the provider tends to be upheld, participants are likely to be exposed to a lack of certainty and stability in their communities and will be potentially vulnerable to the arbitrary and malicious exercise of power by the providers. Private governance, understood in this absolutist sense, offers none of the safeguards of corporeal public governance.

In order to better conceptualize these issues, it is useful to analyze the power of providers through the constitutional lens of the ideals of the rule of law, which operates in Western democratic theory as a fundamentally important limitation on the abuse of public power. The rule of law is a contested set of ideals that consists of a number of different strands, none of which can be universally or directly applied to the governance of virtual communities, but each of which serves to highlight potential shortcomings in private governance. These strands include restraints on discretionary power,⁸⁴ substantive limits based upon individual rights,⁸⁵ formal limits on the creation and implementation of laws,⁸⁶ procedural safeguards and due process,⁸⁷ and an emphasis on consensual governance.⁸⁸ The rule of law discourse provides a rich set of theoretical critiques about the legitimacy of governance and, as such, provides an appropriate framework through which to evaluate the legitimacy of governance in virtual communities and the legal limits that could be imposed on the exercise of private governance power.

83. Jankowich, *supra* note 40, at 20 (arguing that the interpretation of contractual rules by proprietors “is more likely to be and appear arbitrary” than under the common law); *see also id.* at 45 (noting that three-quarters of virtual world contracts surveyed “allowed the proprietor to delete a player account at the proprietor’s discretion”).

84. *See* DICEY, *supra* note 1, at 187–88.

85. *See* T. R. S. ALLAN, CONSTITUTIONAL JUSTICE: A LIBERAL THEORY OF THE RULE OF LAW 27 (2001); RONALD DWORKIN, LAW’S EMPIRE 93 (1986).

86. *See* LON L. FULLER, THE MORALITY OF LAW (2nd ed. 1969); FRIEDRICH A. VON HAYEK, THE ROAD TO SERFDOM (50th anv. ed. 1994); Joseph Raz, *The Rule of Law and Its Virtue*, 93 L.Q. REV. 195 (1977).

87. *See* ALLAN, *supra* note 85, at 121; JOHN RAWLS, A THEORY OF JUSTICE 239 (9th ed. 1972); Raz, *supra* note 86, at 201–02.

88. *See* JURGEN HABERMAS, BETWEEN FACTS AND NORMS: TOWARD A DISCOURSE THEORY OF LAW AND DEMOCRACY 449 (William Rehg trans., 1996); John Locke, *Second Treatise*, in TWO TREATISES OF GOVERNMENT § 95 (Peter Laslett ed., Cambridge Univ. Press 3d ed. 1988) (1690).

If contractual governance is viewed as a purely private and autonomous enterprise and the creation and enforcement of internal norms is wholly deferred to providers, the role of law in shaping the lives of those within virtual communities becomes marginalized.⁸⁹ If law is not merely restrictive or wholly subject to the interests of the powerful, but can and does play a useful role in restraining the raw exercise of power,⁹⁰ then reducing the role of law poses a risk in that power within virtual communities is not subject to the rule of law. The rule of law provides a framework through which it is possible to contrast the ideals of governance in a liberal democracy with the reality of everyday private governance in virtual communities. As virtual communities grow in importance and become more central to the lives of a rapidly increasing number of users, the idea that governance is unimportant in these spaces because they are *private* is not just archaic, but dangerous. While the rule of law is often thought of as solely relevant to public law, this is not necessarily the case. A. V. Dicey argued that the rule of law “pervades” the English common law, as the “general principles of the constitution . . . are . . . the result of judicial decisions determining the rights of private persons in particular cases brought before the courts.”⁹¹ This point remains as important today as it was a century ago:

[T]he division between public and private law, though important, can never be safely invoked without reference to the specific context [T]here can be no clear-cut distinction between the state and other “quasi-public” bodies, or even private associations that exercise significant power over their own members. As the problems of abuse of power by non-governmental bodies becomes more clearly recognized, the common law is capable of generating appropriate requirements of fairness and rationality in private law.⁹²

The values of the rule of law have important ramifications for private law and private relationships, although the “countervailing public interest in protecting people’s constitutional freedom to define the terms of their own association as they see fit” must be recognized.⁹³ While rule of law values cannot provide a wholly determinative answer, they do provide an important

89. Radin, *supra* note 38, at 147.

90. E. P. THOMPSON, WHIGS AND HUNTERS: THE ORIGIN OF THE BLACK ACT 266 (1990).

91. DICEY, *supra* note 1, at 195.

92. ALLAN, *supra* note 85, at 11; *see also* T. R. S. ALLAN, LAW, LIBERTY, AND JUSTICE: THE LEGAL FOUNDATIONS OF BRITISH CONSTITUTIONALISM 4 (1993) (“[T]he ideas and values of which the rule of law consists are reflected and embedded in the ordinary common law.”).

93. ALLAN, *supra* note 85, at 12.

normative framework through which to conceptualize and evaluate tensions about private governance in virtual communities. By exposing the underlying tensions and presenting a framework that is sensitive to both legitimacy and autonomy, rule of law values can provide guidance for desirable outcomes when disputes arise around governance in virtual communities.

Because they provide an established discourse about the legitimate exercise of governance power, rule of law values form the most important component of the constitutional discourse that ought to inform the continued development of cyberspace governance and regulation.⁹⁴ Rule of law values are particularly useful in that they provide a framework that is much more familiar with the tensions of legitimate governance than is the contractual doctrine generally used to evaluate private governance disputes. A governance framework is required to evaluate tensions in virtual worlds (and virtual communities more broadly):

In virtual worlds, the relationship between platform owners and players is not simply one between producers and consumers. Rather, it is often a relationship of governors to citizens. Virtual worlds form communities that grow and develop in ways that the platform owners do not foresee and cannot fully control. Virtual worlds quickly become joint projects between platform owners and players. The correct model is thus not the protection of the players' interests solely as consumers, but a model of joint governance.⁹⁵

Mark Zuckerberg, founder of *Facebook*, expressed a similar point of view in early 2009 when *Facebook* decided to seek user input on its Terms of Service.⁹⁶ Zuckerberg explicitly recognizes the tension between contractual and governance discourses where the Terms of Service are used in a way that governs participation:

Our terms aren't just a document that protect our rights; it's the governing document for how the service is used by everyone across the world. Given its importance, we need to make sure the terms reflect the principles and values of the people using the service.

Our next version will be a substantial revision from where we are now. It will reflect . . . how people share and control their information, and it will be written clearly in language everyone can understand. Since this will be the governing document that we'll all

94. See Berman, *supra* note 2; Fitzgerald, *supra* note 2.

95. Balkin, *supra* note 28, at 2082.

96. See Caroline McCarthy, *Facebook's About-face: Change We can Believe In?*, SOCIAL (Feb. 18, 2009), http://news.cnet.com/8301-13577_3-10166663-36.html.

live by, Facebook users will have a lot of input in crafting these terms.⁹⁷

Because contractual terms of service play a constitutive role in virtual communities, it makes sense to use constitutional discourse to examine their effect on private governance. Historically, rule of law ideals, in particular, have pervasively shaped the evaluation of territorial governments. At least in Western liberal democracies, these ideals form a large part of what it means to have good governance.⁹⁸ The vocabulary of the rule of law seems to fit reasonably comfortably with emerging tensions in the governance of virtual communities. The concern that the law of the firm is superseding the law of the state means that the new regime “is not subject to democratic input and debate,”⁹⁹ and that the exercise of power is not “subject to continuing rebalancing and checking by the courts.”¹⁰⁰ As the importance of private rules increases in all aspects of social life, the lack of restraint on the abuse of private power threatens the practical ideals of the rule of law.¹⁰¹ Using a rule of law framework highlights these tensions and directly confronts the issues that arise from private governance.

The remainder of this Article will canvass three main themes that emerge from rule of law discourse: the proposition that governance ought to be

97. Mark Zuckerberg, *Update on Terms*, FACEBOOK (Feb. 17, 2009), <http://blog.facebook.com/blog.php?post=54746167130>.

98. BRIAN Z. TAMANAHA, *ON THE RULE OF LAW: HISTORY, POLITICS, THEORY* 1–3 (2004).

99. Radin, *supra* note 38, at 6.

100. *Id.*

101. Nikolas Rose explains that the dichotomy between public and private exercises of power is false and highlights the lack of legal and constitutional restraints on the exercise of power:

The strategies of regulation that have made up our modern experience of “power” formulate complex dependencies between the forces and institutions deemed “political” and instances, sites and apparatuses which shape and manage individual and collective conduct in relation to norms and objectives, but yet are constituted as “non-political.” They do not have law and constitutionality as their governing principle, but entail diverse ways in which legal mechanisms, agents, codes and sanctions are called upon and activated in different contexts. The lines between public and private, compulsory and voluntary, law and norm operate as *internal* elements within each of these complexes, as each links the regulation of public conduct with the subjective emotional and intellectual capacities and techniques of individuals, and the ethical regimes through which they govern their lives.

Nikolas S. Rose, *Government, Authority and Expertise in Advanced Liberalism*, 22 *ECON. & SOC’Y* 283, 286–87 (1993).

limited by law;¹⁰² the liberal emphasis on predictability and formal legality;¹⁰³ and the importance placed upon consent and democracy as a source for legitimacy in pluralistic communities.¹⁰⁴ Each of these themes draws out different concerns and tensions about private governance that are somewhat difficult to recognize under a classical contractual framework.

The extent to which each of these three themes is important is highly sensitive to the particular community context. Another recent attempt to evaluate the existence of the rule of law within virtual worlds concluded that virtual worlds exhibit few of the indicators of the rule of law.¹⁰⁵ This framework focused on predictability and formal legality, rule by law, as key drivers for business investment in virtual worlds, rather than tackling substantive conceptions.¹⁰⁶ The normative aspects of the rule of law are more difficult to apply to virtual worlds for four main reasons: that games do not lend themselves to freedoms; that requirements of democracy and legitimacy are difficult to reconcile with provider rule by fiat; that liberty is limited by technical constraints as participants are not able to easily leave the community; and that “to the extent that the rule of law fosters investment by setting expectations, liberal ideals are less important.”¹⁰⁷

Concerns about the applicability of substantive rule of law values reflect the concerns that virtual communities ought to be able to consensually develop in ways that do not reflect liberal values and that, from a business perspective, liberal values are less important than a stable framework for the enforcement of known rules. This view is correct in that it is difficult to map substantive rule of law values for virtual communities, but these issues are nonetheless worth examining. That a particular virtual community does not embrace certain rule of law values may not be a concern for the territorial regulation of virtual communities, but legitimacy is a key issue. The values of the rule of law are not universal within virtual communities. The promise of diverse communities includes the ability to participate by rule sets that are arbitrary, unpredictable, oppressive, or not reflective of liberal, democratic values. This promise, however, is conditioned upon the consent of those who participate within these spaces.¹⁰⁸ Where rule of law values are

102. *See infra* Part III.

103. *See infra* Part IV.

104. *See infra* Part V.

105. Michael Risch, *Virtual Rule of Law*, 112 W. VA. L. REV. 1, 50 (2009).

106. *Id.* at 12 (discussing “the positive, rather than normative, aspects of the rule of law in virtual worlds”).

107. *Id.* at 20–22.

108. *Id.* at 22 (noting that “subscribers democratically choose to have their avatars be subject to dictatorial laws”).

potentially impugned by community norms, territorial states may use consent and legitimacy as a primary indicator of whether or not those norms are harmful and, and if so, whether they ought to be constrained.

In this rule of law analysis, it is not always necessary to identify the sovereign source of law with precision. Regulation comes in a number of different forms, each of which affects participants.¹⁰⁹ Regulation also comes from a number of different sources: the moral force of the community, the imposed rule of the provider, and the laws of territorial states. Some tensions are best illustrated from a position internal to the rules and norms of the virtual community, while others are clearer from an external position.¹¹⁰ There are overlapping constraints from multiple sources, but “[w]hat matters is the cumulative effect of the law on its subjects.”¹¹¹ The interplay between internal and external perspectives and sources of regulation constructs the experience of participants, who are subject to all these forces at once.¹¹² This Article will proceed on the basis that rule of law values highlight tensions that can be located in different sources of regulation in virtual communities and provide insights that may be relevant to a number of different forms of governance.

A final caution is necessary before embarking on a normative account of the rule of law and the applicability of these values to virtual communities. It is important to remember that the ideals of the rule of law are deeply contested and are certainly not universal. This is particularly so for those rule of law values that exist primarily as Western liberal ideals, such as the emphasis on formality or the set of substantive rights familiar to Western constitutionalism.¹¹³ Much of the argument that follows will proceed from

109. Lessig, *The New Chicago School*, *supra* note 12, at 662–63.

110. *See generally* Orin S. Kerr, *The Problem of Perspective in Internet Law*, 91 GEO. L.J. 357 (2003) (arguing that the problem of perspective pervades internet law); Jonathon W. Penney, *Privacy and the New Virtualism*, 10 YALE J.L. & TECH. 194 (2007) (discussing the “new virtualism” approach that blends internal and external perspectives to address emerging issues in cyberspace regulation); James Grimmelman, *Virtual Borders: The Interdependence of Real and Virtual Worlds*, 11 FIRST MONDAY (Feb. 2006), available at <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1312/1232> (explaining virtual world tensions from internal and external perspectives).

111. Risch, *supra* note 105, at 25.

112. *See* Cohen, *supra* note 4, at 215.

113. There is an argument that the more basic form of the rule of law, as a constraint on the exercise of government power, is universal. *See* TAMANAHA, *supra* note 98, at 137; THOMPSON, *supra* note 90, at 266. Nevertheless, even this basic limitation cannot be universal in the context of virtual communities, where those communities, like some games, for which arbitrary governance is part of the appeal, must be protected. *See* Bartle, *supra* note

the basis of Western liberal understandings of the rule of law and deal with the impact that these understandings ought to have on the regulation of virtual communities by Western states. This frame unfortunately excludes the rapidly developing jurisprudence of many countries that are struggling with similar issues, particularly South Korea, whose familiarity with tensions arising out of virtual communities in many cases far surpasses that of Western countries,¹¹⁴ and for whom the Western liberal ideals of the rule of law do not have the same resonance. While the constitutional discourse and set of fundamental values may be different, however, future comparative work may show very similar struggles around legitimacy that extend beyond the Western framework.¹¹⁵

III. GOVERNANCE LIMITED BY LAW

As noted above, the rule of law discourse contains a number of separate and contested ideals. One of the primary clusters of values of the rule of law requires that governance operates within the limiting framework of the law. This means that those in positions of power must abide by the law and that the law should only be changed by appropriate procedures within appropriate limits.¹¹⁶ It is in this sense that the rule of law has been called a “universal human good,” as all societies benefit from restraints on the arbitrary or malicious exercise of power.¹¹⁷

Measuring governance within virtual communities against the principles of this conception of the rule of law highlights some interesting shortcomings in cyberspace self-governance. Most notably, the power of providers in virtual communities is not often restrained to acting in accordance with the rules. Additionally, however, these values suggest some

50 (arguing that creators of play spaces sometimes need absolute control over the environments in order to make them attractive).

114. See particularly the works of Judge Ung-Gi Yoon: Ung-Gi Yoon, *A Quest for the Legal Identity of MMORPGs—From a Computer Game, Back to a Play Association*, 10 J. GAME INDUSTRY & CULTURE (2005), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=905748; Ung-Gi Yoon, *Real Money Trading in MMORPG Items from a Legal and Policy Perspective*, 1 J. KOR. JUDICATURE 418 (2008) [hereinafter Yoon, *Real Money Trading in MMORPG*].

115. See Ung-Gi Yoon, *Connecting East and West*, Presentation at the State of Play V: Building the Global Metaverse Conference (Aug. 20, 2007) (video available at http://origin.eastbaymedia.com/embed/player.swf?height=350&width=500&streamer=rtmp://fms.ebmcndn.net/8004B6/origin.eastbaymedia.com&file=streamer=rtmp://fms.ebmcndn.net/8004B6/origin.eastbaymedia.com&file=nyls/flash/SOP/03_SOP_DVD_03/SOP_5/200807_04.flv).

116. TAMANAHA, *supra* note 98, at 115.

117. *Id.* at 137; THOMPSON, *supra* note 90, at 266.

substantive limits that may be appropriate to impose on virtual governance in order to safeguard the interests of participants.

A. RULE OF LAW LIMITATIONS ON ARBITRARY PUNISHMENT

The first conception of the rule of law is a prohibition on arbitrary governance, a requirement that power is exercised according to the law. This was famously set out by A. V. Dicey, whose “first and main articulation”¹¹⁸ of the rule of law was:

[N]o man is punishable or can be lawfully made to suffer in body or goods except for a distinct breach of law established in the ordinary legal manner before the ordinary courts of the land. In this sense the rule of law is contrasted with every system of government based on the exercise by persons in authority of wide, arbitrary, or discretionary powers of constraint.¹¹⁹

A limitation on the arbitrary exercise of power immediately raises questions about the power of providers in virtual communities. Essentially, providers have control over the code that creates the platform, allowing them to exercise absolute power within the community itself.¹²⁰ Any feature of the community can be changed at will by altering the code in some way. A provider accordingly has an unlimited technical ability to alter the virtual landscape—changing entitlements to virtual property, limiting the ability of participants to express themselves or communicate with others, or imposing punishments and excluding participants from the community altogether. These abilities can be exercised programmatically upon certain defined triggering conditions or ad-hoc by the direct intervention of a provider’s representative.

A provider’s technical ability to alter the virtual landscape is limited by a number of sources generalizable to the market, norms, and law.¹²¹ If internal governance is successful, sufficiently legitimate internal norms will develop to respond to the needs of the community.¹²² Alternatively, should internal norms fail, participants may vote with their feet, or wallets, and move to another community, thus allowing the market to efficiently regulate.¹²³

118. TAMANAHA, *supra* note 98, at 63.

119. DICEY, *supra* note 1, at 188.

120. See Bartle, *supra* note 50, at 27.

121. Lessig, *The New Chicago School*, *supra* note 12, at 662–63.

122. See David G. Post, *Governing Cyberspace: Law*, 24 SANTA CLARA COMPUTER & HIGH TECH. L.J. 883, 911–12 (2007).

123. See Elkin-Koren, *supra* note 30, at 1180–85; Epstein, *supra* note 14, at 17–18; Post, *supra* note 122, at 170.

Assuming, however, that neither of these forces provides a satisfactory guarantee, the law is expected to impose some limits on the absolute discretion of a provider where appropriate.

The most immediate legal limits on a provider's discretion usually lie in the contractual terms of service that purport to govern most communities. First, providers are expected to act in accordance with the terms of service since these contractual documents ought to be enforceable against providers and not merely for the benefit of providers.¹²⁴ This leads to some serious problems, particularly as most terms of service are drafted in a manner that greatly favors the interests of the provider.

Most importantly, terms of service generally include clauses that reserve a wide discretion to the provider.¹²⁵ In communities where the value of the rule of law against arbitrary power is significant, clauses that allow absolute discretion should be regarded suspiciously. Take, for example, the *Facebook* Terms of Use, as they were before they were updated due to user protest in May 2009.¹²⁶ The former terms provided that:

[Facebook] may terminate your membership, delete your profile and any content or information that you have posted on the Site . . . and/or prohibit you from using or accessing the Service or the Site . . . for any reason, or no reason, at any time in its sole discretion, with or without notice.¹²⁷

Facebook is an interesting example, as it eventually decided to create less harsh terms of service in response to user protest.¹²⁸ Assuming *Facebook* had not modified its terms, however, this conception of the rule of law may suggest that such broad discretionary powers ought to be restrained in appropriate cases. If a *Facebook* subscriber had her account terminated for no apparent reason, or for expressing criticism of *Facebook*, for example, could

124. See Risch, *supra* note 105, at 27–28.

125. See Jankowich, *supra* note 40, at 20.

126. Protest over a proposed change to the *Facebook* Terms of Use led *Facebook* to completely revise its terms in a manner that invites public input. See Zuckerberg, *supra* note 97; see also McCarthy, *supra* note 96 (explaining the controversy created by *Facebook*'s changed terms).

127. Facebook Terms of Use, TOSBACK (Feb. 24, 2009), <http://www.tosback.org/version.php?vid=156> (Sept. 23, 2008 revision).

128. It is important to note, however, that under the new terms of service, while absolute discretionary power is not explicitly claimed, any award for damages for breach is limited so tightly as to effectively close off the threat of contractual breach as a limit on discretion. See Statement of Rights and Responsibilities, FACEBOOK, § 15(3) (Oct. 4, 2010), <http://www.facebook.com/terms.php> (limiting damages to the greater of \$100 or the amount the subscriber has paid *Facebook* in the last twelve months, which is generally zero).

Facebook rely on the broad discretionary clause in the Terms of Service to avoid any potential liability?

The answer, of course, must be: “it depends.” There are competing tensions at stake, and it is possible that some communities rely on the ability to act arbitrarily and that participants in those communities may not always be harmed (at least in a way that ought to be legally recognizable) by the exercise of broad discretionary powers.¹²⁹ This conception of the rule of law, accordingly, does not seem to be universal, at least not with regard to the exercise of private virtual governance. However, there may be some communities where the existence of such a broad discretionary power is harmful to the point where it should be restricted.

Throughout history, this conception of the rule of law has been seen as important to help ward off tyrannical governance, and is a project that “will never be obsolete.”¹³⁰ This concern, if it is accurate, is not likely to dissipate simply because the loci of certain governance tensions move online to private virtual communities. Thus, the arbitrary or malicious exercise of power by the providers and their delegates ought to be cause for concern, at least for some communities. In a paradigmatic case, not only will the discretion of a provider be limited to that provided under the contract, but a contractual clause that claims absolute discretionary power may not be enforceable. To the extent that contractual documents are used to govern behavior in virtual communities, it is reasonable that providers be similarly bound by the same “law,” at least where doing so would not unduly harm the community.

B. PROTECTION OF SUBSTANTIVE INTERESTS

A requirement that governance be limited by law is somewhat empty if there are no substantive limits on the ability to create and modify the law.¹³¹ It follows for some rule of law theorists that if the exercise of power ought to be authorized by law, then the lawmaking power of the government

129. See Bartle, *supra* note 50 at 26–27 (explaining the arbitrary powers of some game administrators and game rules); see also Vili Lehdonvirta, *The Efficient Level of RMT in MMORPGs*, VIRTUAL ECON. RES. NETWORK (Aug. 23, 2007), http://virtual-economy.org/blog/the_efficient_level_of_rmt_in_ (arguing that people can react differently to varying levels of RMT in online gaming communities).

130. TAMANAHA, *supra* note 98, at 138–39.

131. JOHN AUSTIN, *THE PROVINCE OF JURISPRUDENCE DETERMINED* 225 (2d ed. 1861) (“[T]he power of a monarch properly so called, or the power of a sovereign number in its collegiate and sovereign capacity, is incapable of legal limitation.”); Brian Tierney, “*The Prince Is Not Bound by the Laws.*” *Accursius and the Origins of the Modern State*, in 5 COMP. STUDS. SOC’Y & HIST. 378, 385 (1963) (arguing that “[i]n constitutional states the eliciting of a consensus is just as important as the exclusion of caprice”).

should be limited over certain subject matter. Historically, these substantive limits come from a variety of sources such as natural law, divine law, custom, human rights, civil and political rights, and positive instruments like bills of rights.¹³² In this way, “the legality of a person’s treatment, at the hands of the state, depends on its being shown to serve a defensible view of the common good.”¹³³ The limits on the ability to create rules in virtual communities also come from a variety of sources, both legal and non-legal. Focusing only on legal limits, the constitutive limits are drawn not only from contract and property law but from the sum of all law that can potentially structure the relationship between participants and providers.

Dicey, writing at the turn of the twentieth century, was particularly concerned with showing how the constitutional values and rights of English citizens were protected by the general law without the need or existence of a written constitution.¹³⁴ While Dicey recognized that whether substantive rights were protected under a written constitution or by the common law was “a merely formal difference,”¹³⁵ he argued that the English approach was more useful than that of the French Constitution because it focused on remedies available to enforce rights rather than potentially empty declarations of the existence of rights.¹³⁶ Accordingly, values such as individual liberty, property, and freedom of speech are all protected by the operation of general law.¹³⁷ For example, “the right to express one’s opinion on all matters,” is protected by the common law, “subject to the liability to pay compensation for libellous or to suffer punishment for seditious or blasphemous statements.”¹³⁸

Dicey’s approach is of some assistance because there is no written constitution that governs virtual communities.¹³⁹ Assuming that public values are being displaced by private governance regimes,¹⁴⁰ then it may be desirable

132. TAMANAHA, *supra* note 98, at 118–19 (explaining that the applicable limits on legislative power come from a number of sources, but that “[t]he key . . . is simply a pervasive belief, on the part of the populace and officials,” that such limits exist).

133. ALLAN, *supra* note 85, at 2.

134. DICEY, *supra* note 1, at 187–88.

135. *Id.* at 198.

136. *Id.* at 198–99.

137. *Id.* at 201–02.

138. *Id.* at 201.

139. Raph Koster proposed a hypothetical Bill of Rights as a thought experiment in 2000. See Raph Koster, *Declaring the Rights of Players* (Aug. 7, 2000), <http://www.raphkoster.com/gaming/playerrights.shtml> (including *A Declaration of the Rights of Avatars*).

140. See Radin, *supra* note 38.

to address these concerns by ensuring that public values are read into the private law doctrines that regulate private governance.

Evolved as they have in the paradigm of freedom to contract and the sovereignty of private property, private contract and property doctrines do not currently reflect the needs or desires of participants in virtual communities. As a result, participants in virtual communities are unable to frame their interests in a manner that is recognizable in the legal system. As more cases are brought where participants seek to assert substantive rights, however, reading protection for certain interests into the regulatory framework of virtual communities may help to reduce the alienating effect of using private law rules to govern those communities.

Take, for example, the case of Peter Ludlow, who wrote a virtual newspaper called *The Alphaville Herald* about events in Electronic Arts' ("EA") virtual world, *The Sims Online*. Ludlow wrote some scathing commentary about EA's management and the lack of an appropriate response to "cyber-prostitution," and posted the story on an external website. After he posted a link to *The Alphaville Herald* on his in-world profile, EA subsequently ejected him from *The Sims* for a technical breach of the rules, which prohibit linking to external sites.¹⁴¹ Ludlow's concerns are free speech concerns. Viewed as a purely contractual dispute, however, the core issue here, the free speech argument, is not legally recognizable. The abstract way in which the legal system construes contractual disputes means that EA had absolute discretion in determining whether to accept Ludlow's breach of the Terms of Service or to terminate Ludlow's account.

This abstraction requires that Ludlow frame his concern in terms of a contractual argument, rather than being able to express his true concerns: that participants in a community ought to be able to express their dissatisfaction about how the community is governed. As critical legal scholars recognized, the imposition of reified legal categories alienates the real needs and desires of citizens by presenting legal abstractions rather than the underlying tensions.¹⁴² A better answer recognizes that contractual doctrine has a constitutive effect and that the limits imposed on the exercise

141. PETER LUDLOW & MARK WALLACE, *THE SECOND LIFE HERALD: THE VIRTUAL TABLOID THAT WITNESSED THE DAWN OF THE METAVERSE* 145–48 (2007).

142. See Peter Gabel, *Reification in Legal Reasoning*, 3 RES. L. & SOC. 25, 30–32 (1980) (discussing the abstraction of fact situations in the first stage of legal decision making); Mark Kelman, *Interpretive Construction in the Substantive Criminal Law*, 33 STAN. L. REV. 591 (1980) (arguing that under-analyzed choices made at the interpretive stage in legal arguments substantially affect the rational resolution of disputes).

of a technical right to terminate shape the boundaries of acceptable governance within the virtual community.

This is not to say that the process is in any way determinative. EA's interest in protecting the image and reputation of the community, and the value of its subscription fees, may outweigh any interest that Ludlow has in expressing his concerns about the in-world governance structures. At least, however, this conclusion will be reached with full knowledge of the values that are at stake, rather than ignoring the underlying tensions through the rote application of abstract doctrine. By expanding the frame of reference and considering the effects and by explicitly reading substantive values into legal doctrine, more appropriate outcomes can be achieved.

The Ludlow example shows that the task of confronting and evaluating substantive governance values is a fundamentally pragmatic exercise. Jettisoning universal natural law principles in favor of the subjectivity of value systems leaves the familiar liberal autonomy problem: the great difficulty of determining appropriate substantive normative limits in a pluralistic society. Several theorists have attempted to suggest some appropriate starting points by addressing the particular tensions that they perceive in cyberspace self-rule. One approach introduces "blanket non-waivability for certain well-defined exceptional categories of entitlements"¹⁴³ in order to allow a general regime of private bargaining to operate. As a "preliminary pass" to identifying some potential exceptions, this approach suggests "three categories for our attention: (1) rights related to legal enforcement; (2) human rights; [and] (3) rights that are politically weak."¹⁴⁴ Other theorists suggest different sets of substantive limits such as: the free speech rights of developers and players;¹⁴⁵ the encouragement of liberal democratic association;¹⁴⁶ property rights, personal and dignitary interests, and limiting fraud;¹⁴⁷ and anti-competitive barriers that would hinder the development of a marketplace of norms.¹⁴⁸

These are some of the governance values that form the substantive constitutional limits on the exercise of government power in Western democracies. Raph Koster drew many of these together as a thought experiment in 2000, when he proposed a hypothetical *Declaration of the Rights*

143. Radin, *supra* note 38, at 149–50.

144. *Id.* at 150.

145. Balkin, *supra* note 28.

146. *See* Netanel, *supra* note 19, at 455.

147. Fairfield, *supra* note 55, at 468.

148. Epstein, *supra* note 14, at 819.

of *Avatars*,¹⁴⁹ modelled on the French *Declaration of the Rights of Man and the Citizen*¹⁵⁰ and the U.S. Bill of Rights.¹⁵¹ Several substantive rights are claimed for participants in virtual communities, including equality,¹⁵² “liberty, property, security, . . . resistance to oppression,”¹⁵³ the right to contribute to the shaping of the internal rules,¹⁵⁴ freedom of speech,¹⁵⁵ freedom of assembly,¹⁵⁶ and privacy.¹⁵⁷ While certainly not exhaustive or authoritative, Koster’s list illustrates the range of substantive issues that, for various reasons, certain societies prohibit their citizens from opting out of, or at least enforce higher than normal thresholds of consent for their modification.¹⁵⁸

It is important to avoid a substantive construction of the rule of law that is so broad that it loses its potency.¹⁵⁹ While many of these constitutional values fit within a rule of law framework because they are said to be prerequisites for legitimate governance, there is no easy claim to universality. These are some of the values that constrain the autonomy of governance in Western liberal democracies, but in working through the list provided by Koster and the concerns raised by other theorists, it is evident that the application of any substantive values as limits to autonomy is heavily context-dependent. The type and extent of desirable substantive limits is, accordingly, likely to differ by territorial state, community, and time.

In the remainder of this Part, four core groups of substantive interests are examined: the interests that revolve around speech, discrimination, and protest; the recognition of property rights; the right to privacy; and the rights

149. Koster, *supra* note 139. Balkin reminds us that “the rights at stake are not really the rights of the avatars themselves. They are the rights of the players who take on particular (and possibly multiple) identities within the virtual communities.” Balkin, *supra* note 28, at 2083.

150. DECLARATION OF THE RIGHTS OF MAN AND THE CITIZEN (Fr. 1789), *translated in The Avalon Project: Documents in Law, History, and Diplomacy*, YALE LAW SCHOOL—LILLIAN GOLDMAN LAW LIBRARY, http://avalon.law.yale.edu/18th_century/rightsof.asp (last visited Nov. 11, 2010).

151. U.S. CONST. amends. I–X.

152. Koster, *supra* note 139, at art. 1.

153. *Id.* at art. 2.

154. *Id.* at art. 6.

155. *Id.* at art. 11.

156. *Id.* at art. 17.

157. *Id.* at art. 18.

158. See Margaret Jane Radin, *Market-Inalienability*, 100 HARV. L. REV. 1849 (1986) (discussing partial inalienability as a method to restrain harmful commodification but simultaneously allow beneficial trades or avoid causing further harm).

159. TAMANAHA, *supra* note 98, at 113 (“The rule of law cannot be about everything good that people desire from government. The persistent temptation to read it this way is a testament to the symbolic power of the rule of law, but it should not be indulged.”).

of legal enforcement. These limits are examples of the types of limits that states may choose to impose on autonomy in virtual communities and have attracted substantial academic interest. This list represents some of the core values that are important to western liberal conceptions of the rule of law. It cannot be either universal or exhaustive, but it is useful in providing an overview of how such interests can be thought of as constitutive limits to cyberspace self-governance and how malleable any such approach must be in order to take into account conflicting social interests.

1. *Freedom of Expression Concerns: Discrimination, Speech, and Protest*

Anti-discrimination law provides a useful example in highlighting the way in which a legislated protection of certain interests can shape the internal rules of a community. It also provides an example of the tailoring that occurs in trading off potential harms against the benefits of allowing communities some degree of autonomy. Take two examples of sexual discrimination in virtual communities: Blizzard threatening Sara Andrews with disconnection from *World of Warcraft*¹⁶⁰ for advertising a LGBT-friendly guild,¹⁶¹ and Microsoft banning Xbox Live¹⁶² players whose names included the word “gay.”¹⁶³ Both of these examples highlight reactions by some participants to communities that they find somewhat threatening—in these cases,

160. *World of Warcraft* is an extremely successful massively multiplayer online roleplaying game (“MMORPG”) by Blizzard Entertainment. Blizzard reports that the game currently has over twelve million active subscribers. See Press Release, Blizzard Entm’t, *World Of Warcraft Subscriber Base Reaches 12 Million Worldwide* (Oct. 7, 2010), <http://us.blizzard.com/en-us/company/press/pressreleases.html?101007>. But see Daeity, *Blizzard’s “Active Subscriptions” vs “Real Players,”* DIGITAL CASTRATION (Aug. 9, 2010), <http://daeity.blogspot.com/2010/08/blizzards-active-subscription-numbers.html> (arguing that the active subscriber figure is inflated due to Blizzard’s practice of banning approximately 100,000 accounts per month).

161. Brian Ashcraft, *Blizzard’s Reaction to Gay Guilds an “Unfortunate Mistake,”* KOTAKU (Mar. 9, 2006, 5:24 PM), <http://kotaku.com/159536/blizzards-reaction-to-gay-guilds-an-unfortunate-mistake>; Mark Ward, *Gay Rights Win in Warcraft World*, BBC NEWS (Feb. 13, 2006, 8:42 PM), <http://news.bbc.co.uk/2/hi/technology/4700754.stm>.

162. Xbox Live is Microsoft’s online gaming service. Microsoft claims the service reached twenty-three million subscribers in February 2010. See Marc Whitten, *An Open Letter from Xbox LIVE General Manager Marc Whitten*, XBOX.COM (Feb. 5, 2010), <http://www.xbox.com/en-US/press/2010/0205-whittenletter.htm>.

163. Luke Plunkett, *Xbox Live “Gay” Crackdown MIGHT Be Getting A Little Out Of Hand*, KOTAKU (May 21, 2008), <http://kotaku.com/392304/xbox-live-gay-crackdown-might-be-getting-a-little-out-of-hand>; Jay Slatkin, *“Gay” Player Name Banned By Xbox Live*, CONSUMERIST (May 14, 2008), <http://consumerist.com/5008908/gay-player-name-banned-by-xbox-live>.

prominent homophobia from other participants.¹⁶⁴ In order to minimize perceived conflict, the provider has in each case threatened to ban minority group participants for overtly expressing their sexuality in a way that could trigger negative reactions from other participants.¹⁶⁵

Under the Terms of Service, both Blizzard and Microsoft reserve a broad right to terminate access to participants.¹⁶⁶ These contractual clauses may, however, come into conflict with anti-discrimination laws.¹⁶⁷ Many territorial states already provide limits on discrimination for sexual orientation within private groups.¹⁶⁸ In the United States, anti-discrimination laws generally prohibit discrimination on protected grounds in “places of public accommodation.”¹⁶⁹ To date, courts have been reluctant to find that online communities are “places of public accommodation.”¹⁷⁰ As the importance of

164. See Justin Cole, *Op/Ed: The Impact Of Homophobia in Virtual Communities*, KOTAKU (July 11, 2009), <http://www.kotaku.com.au/2009/07/oped-the-impact-of-homophobia-in-virtual-communities/> (arguing that virtual communities express a widespread, normalized homophobia).

165. See Sara Andrews, Posting to *World of Warcraft Not Gaymer Friendly*, GAMERS EXPERIMENTATIONS (Jan. 16, 2006, 6:12 PM), <http://web.archive.org/web/20060221231447/http://gamers.experimentations.org/forums/index.php?showtopic=6852> (Internet Archive copy); *MS XBL Gay Equals Sex?*, LESBIAN GAMERS (May 23, 2008), <http://lesbiangamers.com/2008/05/ms-xbl-gay-equals-sex/>; PixelPoet, *Xbox Live Gaywood Drama Update Gay Gamer*, GAYGAMER (2008), http://gaygamer.net/2008/05/xbox_live_gaywood_drama_update.html.

166. See *World of Warcraft: Terms of Use Agreement*, WORLD OF WARCRAFT.COM, § 6 (Oct. 29, 2010), <http://www.worldofwarcraft.com/legal/termsofuse.html> (“BLIZZARD MAY SUSPEND, TERMINATE, MODIFY, OR DELETE ANY BNET ACCOUNT OR WORLD OF WARCRAFT ACCOUNT AT ANY TIME FOR ANY REASON OR FOR NO REASON, WITH OR WITHOUT NOTICE TO YOU.”); *Xbox LIVE and Games for Windows LIVE Terms of Use*, XBOX.COM, § 16 (Oct. 2010), <http://www.xbox.com/en-us/legal/livetou.htm> (“We may cancel or suspend your Service at any time. Our cancellation or suspension may be without cause and without notice.”).

167. Dagmar Schiek, *Freedom of Contract and a Non-Discrimination Principle: Irreconcilable Antonyms?*, in *NON-DISCRIMINATION LAW: COMPARATIVE PERSPECTIVES* 85–88 (Titia Loenen & Peter Rodrigues eds., 1999) (arguing for non-discrimination as a general principle of contract law).

168. *Id.* at 77–78.

169. See, e.g., 42 U.S.C. § 2000a(b) (2006) (defining public accommodations for the purpose of the 1964 Civil Rights Act); 42 U.S.C. § 12181(7) (2006) (defining public accommodations for the purposes of the Americans With Disabilities Act).

170. See *Noah v. AOL Time Warner, Inc.*, 261 F. Supp. 2d 532, 544 (E.D. Va. 2004), *aff'd*, No. 03-1770, 2004 WL 602711 (4th Cir. Mar. 24, 2004) (holding that because they lack physicality, “AOL’s online chat rooms cannot be construed as ‘places of public accommodation’ ” for the purposes of the Americans With Disabilities Act); *In Chambers — Court Order, Stern v. Sony Corp.*, No. 09-CV-7710 (C.D. Cal. Feb. 8, 2010), ECF No. 18, available at <http://www.scribd.com/doc/28950515/Stern-v-Sony-MTD-Order>

participation in online communities increases, however, the significance of being excluded on an unacceptable basis similarly becomes greater; if extended, such laws could prevent providers from discriminating as to who can join and remain in a community.¹⁷¹ To the extent that these rules are effective, they become constitutive limits by limiting the ways in which communities and groups can choose to discriminate.

The law in this area continues to develop. It seems likely that some form of rights to non-discriminatory access will be recognized in the future, at least for some communities, particularly those that have a more public character. In developing these rules, courts and legislatures should be particularly mindful of when certain forms of discrimination are tolerable for specific purposes and when such behavior crosses the line into impermissible discrimination or even vilification. Some level of discrimination is often beneficial where that discrimination goes to the heart of the community's purpose.¹⁷² The proper evaluation of whether discrimination is desirable must be determined by a thorough examination of the circumstances and social structure of the particular community.

This principle holds for other potential substantive limits; the relative importance of social values is contingent on the purpose and use of the community. For example, the importance placed on freedom of speech of participants needs to be weighed against the speech interests of the providers in virtual communities.¹⁷³ These freedoms can conflict at times and the developer's free speech rights to create an expressive game or other platform may outweigh any concerns about the legitimate interests of participants.

(holding that Sony's online games were not "places of public accommodation" for the purposes of the ADA).

171. See Balkin, *supra* note 28, at 2084–85; Colin Crawford, *Cyberplace: Defining a Right to Internet Access Through Public Accommodation Law*, 76 TEMP. L. REV. 225 (2003) (arguing that rules of public accommodation ought to be extended to virtual communities); Netanel, *supra* note 19, at 456 (arguing that "Cyberfora and networks that are generally open to the public should similarly be seen as 'places of public accommodation,' whether by statutory construction or legislative extension"); Joshua Newton, *Virtually Enabled: How Title III of the Americans with Disabilities Act Might Be Applied to Online Virtual Worlds*, 62 FED. COMM. L.J. 183 (2010) (arguing that virtual worlds should be treated as places of public accommodation and analyzed separately from websites).

172. Netanel argues that discrimination should be acceptable in circumstances where it is necessary in order to conduce meaningful and effective expression: "Some conversations lose their essential purpose and meaning unless limited to persons of a particular group. In such instances, the participants' interest in discriminating (and the allied public interest in promoting discursive expression and association) should prevail over the interest in preventing invidious status discrimination." Netanel, *supra* note 19, at 459–60.

173. Balkin, *supra* note 28.

However, sometimes the free speech rights of individual participants may be so important as to warrant regulatory protection.¹⁷⁴

The private nature of virtual communities, by placing the power to regulate speech in the hands of private property owners, has the capacity to significantly interfere with the liberty of individual citizens, particularly as online fora become more important to expression.¹⁷⁵ A purely negative reading of the First Amendment that does not extend in any way to private restrictions on speech in virtual communities substantially undercuts the protection historically afforded to speech, as there are few virtual spaces analogous to the protected corporeal public forums that exist to provide a platform for citizens to speak or be heard freely.¹⁷⁶ Harm can occur where a provider has absolute discretion over the content of communications within a virtual community, thus leading to a suggestion that at least for the more “public” types of communities, territorial states may have a legitimate interest in limiting the ability of the provider to regulate participant speech.¹⁷⁷

While it certainly seems desirable to protect participant speech, there is a significant difficulty in determining when private restraints on speech ought to be acceptable and when they should not. Limits on private restraints could be considered justified in “[i]nternet forums that are generally open to the public for free speech purposes,”¹⁷⁸ but this standard excludes the majority of virtual communities, which generally do not explicitly hold themselves out as free speech zones. Other approaches, on the other hand, suggest that regulatory boundaries can be drawn along the distinctions between commodified and non-commodified communities and the distinction between communities that encourage the free exchange of ideas and those that are developed to “realize the artistic or ideological vision of the platform owner.”¹⁷⁹ Balkin sets up a tension between state regulation and the free speech interests of virtual community providers, arguing that:

Regulating the platform owner’s right to design in order to protect the participants’ right to play is most justifiable when the virtual world serves as a public space for commerce, and when it is held open as a public space for the exchange of ideas. These two distinctions may not be perfectly clear in all cases; but they point

174. *Id.* at 2084.

175. Dawn C. Nunziato, *The Death of the Public Forum in Cyberspace*, 20 BERKELEY TECH. L.J. 1115, 1121–23 (2005).

176. *Id.* at 1117–18.

177. Balkin, *supra* note 28, at 2090; Nunziato, *supra* note 175, at 1161, 1167.

178. Nunziato, *supra* note 175, at 1166.

179. Balkin, *supra* note 28, at 2090.

the way to the boundaries of permissible state regulation on the one hand, and the free speech rights of platform owners on the other.¹⁸⁰

Accordingly, free speech interests of participants will be most important to recognize in virtual communities that act like a marketplace, and secondly, in communities that are “offered as a space for the free exchange of ideas.”¹⁸¹

The example of Peter Ludlow’s exile from *The Sims Online* highlights that any claim must be evaluated on the purposes for which the community was created and the way it was used.¹⁸² This seems to be fundamentally correct; no two communities can be treated alike,¹⁸³ and the free speech interests of participants need to be weighed against the free speech interests of the providers.¹⁸⁴ Nevertheless, in appropriate cases, providers may incur some responsibilities to allow dissenting voices, a claim that is somewhat stronger in more general use platforms.

In weighing the competing speech interests of providers and participants, it is desirable to avoid a strict dichotomy between communities that are, or are held out to be, free speech zones and those that are not. Such a distinction is likely to allow the majority of providers to exclude themselves from potential responsibility by simply disclaiming any participant interests. It is much more desirable to examine whether particular limits are appropriate for particular communities than to attempt a blanket determination of whether or not a community is exempt from all speech responsibilities. So, for example, EA could be required to tolerate an external link to a news article that is critical of its governance procedures but not an external link that exposes their users to unsolicited commercial communications. The purpose and use of a community will always be relevant to the types of speech restrictions that territorial states may consider appropriate for the provider to impose. If such speech concerns are serious enough to warrant territorial intervention, it is desirable to adopt a more subtle and critical method of evaluation, rather than attempting to rely on a binary classification of a community as either allowing free communication or not.

The analysis becomes more complicated if, in addition to the competing speech interests, the interests of the provider in the stability of the servers or

180. *Id.*

181. *Id.*

182. *Id.* at 2093.

183. *Id.* at 2084.

184. *Id.* at 2080.

network are considered. The interests of participants to peaceful assembly and protest,¹⁸⁵ for example, can sometimes directly conflict with the ability of other participants to enjoy the community and the interests of the provider in maintaining order and community uptime. The task of managing participants is a very complicated exercise for a provider, on which the fate of the entire community often rests.¹⁸⁶ Participants have a very large range of different motivations and interests in the community, and they contribute to the community in many ways. Passionate participants will often manifest their displeasure and, just as displeased citizens in territorial states, will seek to make their voices heard within the community. Virtual worlds have provided fertile platforms for protests over at least the last decade, with some protests about issues specific to the community and others that reflect external political struggles.¹⁸⁷

Individual or small-scale manifestations of dissent are speech concerns, but larger scale virtual protests raise interesting new tensions. Protests are often disruptive by their nature and design, and this is no different in virtual communities. The presence of a large number of people protesting in a virtual world, for example, can potentially prevent others from enjoying the world and can strain the platform.¹⁸⁸ The concentrated presence of a large number of participants in a small area can sometimes impose a severe load on the provider's network and software, which is sufficient to crash the platform and disable the community for a few hours. In addition to other reasons for suppressing dissent, providers accordingly often have a technical incentive to disband in-world protests.

By relying on a clause in the terms of service that prohibits disruptive behavior, providers may respond to protests by threatening to disconnect, suspend, or ban users if they do not disband. If a provider could be required to tolerate dissenting speech from individual subscribers, could it also be

185. See ALLAN, *supra* note 85, at 93–94 (arguing that the rule of law requires freedom to protest against injustice); Locke, *supra* note 88, § 149 (discussing the fundamental power of citizens to replace the government); JOHN STUART MILL, ON LIBERTY 76 (1863) (discussing the necessity of freedom of opinion).

186. See John Banks, *Co-Creative Expertise: Auran Games and Fury—A Case Study*, 130 MEDIA INT'L AUSTL. 77 (2009) (highlighting the different interests of distinct player groups and developers that shape the development of computer games and their commercial success).

187. See Bridget M. Blodgett, *And the Ringleaders Were Banned: An Examination of Protest in Virtual Worlds*, in PROCEEDINGS OF THE FOURTH INTERNATIONAL CONFERENCE ON COMMUNITIES AND TECHNOLOGIES 135, 135–36 (2009).

188. See, e.g., *id.* at 143 (providing examples of protests that resulted in strained platforms).

required to permit disruptive dissent that threatens the stability of the servers and the enjoyment of other members of the community? The tensions at play here are not only those involving speech, but also the provider's interest in the functioning of the platform and in maintaining a harmonious community.

2. *Property Rights*

Many of the speech tensions above are also reflected in the continuing debate about the ownership of virtual objects. In communities where participants are able to create or acquire virtual objects, participants may feel a sense of entitlement to those virtual objects.¹⁸⁹ There may be no descriptive or normative impediment to recognizing such virtual objects as legal property because not only are virtual objects indistinguishable from real world property interests, but the theoretical justifications for recognizing excludable property rights can be extended to virtual environments.¹⁹⁰

One model of virtual property argues that one of the most significant sources of substantive limits on self-governance ought to come from property law.¹⁹¹ Where participants create or acquire virtual property, the law may come to recognize their interests as the owner of that property.¹⁹² Property rights recognized in this manner would impose limits on the ability of providers to unilaterally exercise power over participants. The *Bragg* case, where Marc Bragg sued Linden Lab for terminating his *Second Life* account and confiscating his virtual property, provides an example of possible limits.¹⁹³ Linden alleged that Bragg cheated by purchasing land that was not technically for sale, at significantly under market value. Bragg maintained that he did not cheat and that, at any rate, the punishment was excessive, as it extended beyond the contested land to the remainder of his virtual assets.

Setting aside the circumstances of the dispute for the moment, it is arguable that “[b]ecause courts have not defined the relationship between EULAs and virtual property, the parties were not able to clearly articulate the deal they wished to make.”¹⁹⁴ If Bragg's property rights in his virtual land, objects, and Linden Dollars were recognized, Linden Lab would be unable to

189. Lastowka & Hunter, *supra* note 15, at 37.

190. *Id.* at 49–50.

191. *See generally* Fairfield, *supra* note 52 (arguing that property rights can help to better structure the vertical relationship between participants and providers).

192. Fairfield limits his argument to virtual assets that are rivalrous, persistent, and interconnected. *See id.* at 1053.

193. *Bragg v. Linden Research, Inc.*, 487 F. Supp. 2d 593 (E.D. Pa. 2007).

194. Fairfield, *supra* note 52, at 465.

unilaterally terminate his account without being required to pay compensation—or at least, the parties would be able to negotiate over appropriate rule sets.¹⁹⁵

Some version of property rights ought to be recognized within *Second Life* in particular. The “virtual” economy in *Second Life* is fluidly convertible to “real” currencies, like the U.S. dollar, and participants feel a sense of entitlement to their virtual property and currency. Linden Lab clearly encourages this behavior; its slogan is “Your world. Your imagination,” and its promotional materials refer to the possibility of “owning” virtual land and generally stresses the liquidity of the market.¹⁹⁶ Linden Lab encourages investment in virtual resources and substantially profits from that investment. In a press release announcing changes to the Terms of Service that vested intellectual property rights of in-world creations in subscribers, Linden claimed:

[O]ur new policy recognizes the fact that persistent world users are making significant contributions to building these worlds and should be able to both own the content they create and share in the value that is created. The preservation of users’ property rights is a necessary step toward the emergence of genuinely real online worlds.¹⁹⁷

Judge Robreno, in the *Bragg* case, noted this press release and other hype about the ownership of virtual property and land in *Second Life* and quoted the CEO of Linden Lab as boasting that “[t]he idea of land ownership and the ease with which you can own land and do something with it . . . is intoxicating. . . . Land ownership feels important and tangible. It’s a real piece of the future.”¹⁹⁸

Given these and other comments, it is hardly surprising that a participant such as Bragg would feel aggrieved if Linden Lab were to confiscate his virtual property and wealth. If a similar case were to proceed to final

195. *Id.* at 465.

196. At least until August 22, 2008, Linden Lab proudly proclaimed that residents could “Own Virtual Land” as part of the marketing material on their website. The page has since been removed. *Own Virtual Land Second Life*, LINDEN LAB, <http://web.archive.org/web/20080822144829/http://secondlife.com/whatis/land.php> (Internet Archive copy).

197. *Bragg*, 487 F. Supp. 2d at 596 (quoting Press Release, Linden Lab, Second Life Residents To Own Digital Creations: Linden Lab Preserves Real World Intellectual Property Rights of Users of its Second Life Online Service (Nov. 14, 2003), *available at* http://lindenlab.com/pressroom/releases/03_11_14).

198. *Id.* at 596 (quoting Philip Rosedale *in* Michael Learmonth, *Virtual Real Estate Boom Draws Real Dollars*, USA TODAY, June 3, 2004).

judgment,¹⁹⁹ it would not be difficult for a court to recognize that Linden Lab created an environment where participants derive some form of property rights in their virtual assets. While the appropriate scope of virtual property rights is not clear, they could prevent Linden Lab from destroying the property, or at least the value of the virtual property, whose creation Linden Lab encouraged.²⁰⁰

This principle, however, is not necessarily extendible to other environments. It is possible that a fantasy environment can be created where virtual objects exist and are possessed by participants but no legal property rights should be enforceable.²⁰¹ Because virtual communities are diverse, creating property rights may not be justified in every community. Not every instance of virtual property should be recognized as legal property, and there may be many valid reasons why both participants and providers would not benefit from the recognition of property rights.²⁰² The unfettered ability of the provider to control the community, including the ability to expel a participant and destroy her property, can be fundamentally necessary in order to create a community that is interesting, fun, or useful to its participants.²⁰³ Accordingly, there is a fundamental tension between the interests of participants in having a protected entitlement to what they see as their property and the ability of providers to regulate and develop the community.

Here, too, some distinctions based upon the level of commodification and purpose of a community may be useful.²⁰⁴ Perhaps, the more commodified a virtual community is and the more the provider encourages

199. The *Bragg* case settled on undisclosed terms after the decision in *Bragg*, 487 F. Supp. 2d 593.

200. BENJAMIN DURANSKE, VIRTUAL LAW: NAVIGATING THE LEGAL LANDSCAPE OF VIRTUAL WORLDS 39–40 (2008).

201. Nevertheless, Fairfield argues that property rights should be recognized first and that communal ownership schemes can be created on top of these rights: “[A]n overarching system of private property does permit communal property groups to continue to exist, if the community is able to make its social controls stick. The contrary is not true: the elimination of private property leaves, by definition, no room for private property.” See Fairfield, *supra* note 52, at 1101.

202. See Bartle, *supra* note 50, at 35–37 (arguing that administrators of game worlds need absolute power to prevent commodification in order to protect the game conceit); see also Lehdonvirta, *supra* note 129 (arguing that there are positive and negative effects of allowing trade in virtual goods which vary depending on the community).

203. Bartle, *supra* note 50, at 26–27. Balkin gives the example of *The Gulag Online*, a fictional game where participants experience a simulation of a Soviet-era prison camp; participants could not, in such a simulation, assert any virtual property or due process rights. See Jack M. Balkin, *Law and Liberty in Virtual Worlds*, 49 N.Y.L. SCH. L. REV. 63, 65 (2004).

204. Balkin, *supra* note 28, at 2090 (arguing that worlds that are more commodified should be subject to higher regulation to protect speech).

the creation and trade in virtual property, then the more likely it is that property rights in the virtual property ought to be enforceable. On the other hand, the more the provider is successful in genuine efforts to avoid commodification and the less porous the borders are in allowing real money trades, the more acceptable the provider's argument that the virtual property and virtual currency should not create real world entitlements. This approach allows for a more subtle and tailored examination of property interests. For the types of communities where the developer requires absolute control, and the participants understand this need, then enforceable property rights may have little relevance. Where participants have come to expect a sense of stability in their virtual possessions, however, the absolute ability of the provider to destroy those possessions may need to be curtailed.

One objection to this distinction is that it predicates property interests primarily on corporeal exchange value, ignoring, to an extent, the personal attachment that participants may develop to their virtual possessions and creations. The rise of user-generated content provides a prime example. There is an increasing trend in software development to create a bare platform and encourage the participants to create the assets—the objects and landscape that define the virtual environment.²⁰⁵ Even where this content is not commodified, participants attach particular value to their creations and possessions and may accordingly be entitled to some form of legal recognition for that attachment, either as the objects of their labor²⁰⁶ or as manifestations and expressions of their selves.²⁰⁷

Recognizing the interests of participants in virtual goods or expression may not necessarily require recognizing fully excludable property rights. There may be other approaches that are more suitable to dealing with the complex relationship between participants and providers. For example, recognizing participants' property interests in virtual items could be analogous to goodwill, which allows providers to modify environments while

205. See John Banks & Sal Humphreys, *The Labour of User Co-Creators: Emergent Social Network Markets?*, 14 CONVERGENCE 401, 402 (2008).

206. See Locke, *supra* note 88, § 31.

207. See GEORGE WILHELM FRIEDRICH HEGEL, *ELEMENTS OF THE PHILOSOPHY OF RIGHT* 39, 41, 43 (Allen W. Wood ed., Hugh Barr Nisbet trans., 1991); see also Neil Netanel, *Alienability Restrictions and the Enhancement of Author Autonomy in United States and Continental Copyright Law*, 12 CARDOZO ARTS & ENT. L.J. 1, 20 (1994) (discussing the influence of Kantian's and Hegelian's perspectives on authorship on continental intellectual property law).

giving some level of security to participants.²⁰⁸ In other circumstances, particularly where participants are contributing heavily to the value of the platform by creating content, there may be a real harm when participants are directly brought into the value chain and are thereafter treated unfairly by the provider. For example, unjust enrichment could be an appropriate remedy or, in jurisdictions where they are available, moral rights of attribution and integrity for virtual creations may be enforced. Possibly the most useful approach would be rooted in contractual doctrine and its limitations on the exercise of discretionary powers, such as the requirement of good faith.²⁰⁹ If these remedies prove unsuitable or inappropriate, it may be possible to recognize new interests that prevent a participant from being alienated from her creations, as would happen when a provider terminates a participant's access to a community but continues to use her in-world assets. There may even be situations where it would be desirable to allow a participant-author to enforce a right akin to the French moral right of withdrawal, allowing her to prohibit further uses of the material she creates after she has left a community.²¹⁰

Fundamentally there is unlikely to be a single solution that addresses property-type interests in virtual communities. The level of protection that a participant ought to be entitled to, if any, will be highly dependent on the particular circumstances of the community. The recognition of property rights may be detrimental in some communities, particularly where it would be prohibitively expensive for the community to develop comprehensive procedures to regulate property disputes or in a community where limiting the discretion of the provider to deal with virtual assets would greatly undermine the value and essential qualities of the community. Nevertheless, territorial states have a legitimate interest in articulating a set of entitlements that participants in particular types of virtual communities have in their virtual items or creations and in limiting the corresponding autonomy of

208. In this conception, a trade does not transfer title to the virtual object itself, but to the value of the participant's labor in obtaining the object. *See generally* Ung-Gi Yoon, *Real Money Trading in MMORPG*, *supra* note 114.

209. *See, e.g.*, Dan E. Lawrence, *It Really Is Just a Game: The Impracticability of Common Law Property Rights in Virtual Property*, 47 WASHBURN L.J. 505, 529–30 (2008) (arguing that “Bragg demonstrates that contract law, even in the absence of independent property rights in virtual property, can provide a remedy for an end-user wrongfully deprived of virtual property”); Michael Meehan, *Virtual Property: Protecting Bits in Context*, 13 RICH. J.L. & TECH. 7, ¶¶ 57–60 (2006) (discussing the potential applicability of good faith to restrain providers from devaluing virtual property).

210. The French right of withdrawal is subject to payment of compensation to the user of the work for any harm caused by its exercise. *See* CODE DE LA PROPRIÉTÉ INTELLECTUELLE [C. PROP. INTELL.] art. L121-4 (Fr.).

providers to destroy or modify those entitlements. Enforceable interests in virtual property, seen in this way, act as substantive constitutive restrictions on the scope of cyberspace self-governance.

3. *Right to Privacy*

The potential threats that the use of networked technologies pose to the privacy of participants has been the subject of much discussion in recent decades. The growing importance of participation and the increasing computational power and storage capacity of computer networks highlights immediate concerns about the collection, use, and distribution of personal information. Because all actions that occur “within” a virtual community are essentially reduced to information flows, they are all easily recorded and stored. Actions that are ephemeral in the corporeal world perversely take on a more tangible form when mediated through virtual information networks. Information that is not displayed or carried out synchronously must necessarily be processed and stored for later use; personal messages left on bulletin boards and profile pages are kept indefinitely on the provider’s server, for example. Even information that is used synchronously, however, is vulnerable to capture, including all actions, searches, information and products browsed, real-time chats, and exchanges between participants are potentially logged and stored.²¹¹

The data collected presents a treasure chest of potentially valuable information if it can be analyzed and repurposed in sufficiently innovative ways. Amazon, for example, has built a very successful business model by collating the browsing and purchasing habits of its customers in order to deliver targeted advertising and product recommendations.²¹² When *Facebook* decided to implement a similar system, called Beacon, that would advertise a user’s purchases on certain partner sites to other people in the user’s social network, it was quickly met with outrage from *Facebook* users.²¹³ In response

211. For example, Sony Online Entertainment has recently granted researchers access to several terabytes of data, representing the entire collected actions of 400,000 players in *Everquest 2* over a four-year period. John Timmer, *Science Gleans 60TB of Behavior Data from Everquest 2 Logs*, ARS TECHNICA (Feb. 15, 2009, 4:00 PM), <http://arstechnica.com/science/news/2009/02/aaas-60tb-of-behavioral-data-the-everquest-2-server-logs.ars>; see also Humphreys, *supra* note 34, at 156–57 (discussing the use of spyware to monitor the activities of participants in massively multiplayer online games).

212. See Jennifer Bresnahan, *Personalization, Privacy, and the First Amendment: A Look at the Law and Policy Behind Electronic Databases*, 5 VA. J.L. & TECH. 8, at *3 (2000); see also Tal Z. Zarsky, *Mine Your Own Business: Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion*, 5 YALE J.L. & TECH. 1, 16 (2002).

213. See McCarthy, *supra* note 96.

to user feedback, the system was later changed to an opt-in system.²¹⁴ Social networks are potentially rich sources of revenue for advertisers, but they also raise difficult questions about the juxtaposition of commercial and personal social relationships.

Tensions over the use of personal information are likely to continue playing out in and around virtual communities for the foreseeable future. Many territorial states have some form of privacy legislation or general law rules on the collection, use, and disclosure of personal information, which may be used to restrain or limit the ability of a provider to unilaterally deal with participant data.²¹⁵ More difficult considerations arise when participants are asked to trade consent to control their data for some internal or external benefits. In some cases, these trade-offs are benign and desirable; others, of course, may be exploitative.²¹⁶ Regulating the disclosure of personal information from virtual communities into other contexts is accordingly a difficult process, but a familiar one.²¹⁷

Using a rule of law framework, however, highlights that regulating collection and disclosure of information may not be sufficient to address the privacy interests of participants. Difficult issues arise when considering the use of private information within a virtual community, not merely its leakage out of the community. Because consenting to some level of collection is usually required in order to participate in a community, limits on disclosure typically mean that monitoring and use of information collected within the community by the provider itself is largely unregulated. Under such a regime, providers have an unfettered ability to monitor the communications and the actions of their participants, a proposition that conflicts with the limitations on governance expected of territorial states.

As participants become more involved in virtual communities over an increasing range of activities, limits on the storage and use of information collected within the community itself are likely to grow in importance. The potential for harmful use of personal information that passes through a social

214. Mark Zuckerberg, *Thoughts on Beacon*, FACEBOOK (Dec. 5, 2007), <http://blog.facebook.com/blog.php?post=7584397130>.

215. See generally DANIEL J. SOLOVE, MARC ROTENBERG & PAUL M. SCHWARTZ, *INFORMATION PRIVACY LAW* (3d ed. 2009) (providing an overview of information privacy laws).

216. Radin argues that even where the trade-off is fully informed, individual waiver may deleteriously alter the character of society for those who do not waive their rights and accordingly prohibit all such waivers. Radin, *supra* note 38, at 151.

217. See Tal Z. Zarsky, *Information Privacy in Virtual Worlds: Identifying Unique Concerns Beyond the Online and Offline Worlds*, 49 N.Y.L. SCH. L. REV. 231, 243 (2004) (discussing privacy interests in virtual worlds).

network is not limited to linking internal profiles to external shopping or browsing information. A *Facebook* member may have a legitimate expectation that her private messages and photos will not be viewed and distributed by people within the organization, for example, even if they are not distributed to third parties. Participants in other communities, particularly virtual worlds, may have an interest in preventing the provider from building and utilizing a comprehensive behavioral profile in order to increase retention rates or to deliver highly targeted and influential marketing campaigns.²¹⁸ Further tensions exist around the powerlessness of participants in controlling the information that is collected about themselves and the internalization of external norms under the perpetual potentiality of surveillance.²¹⁹

If these types of problems are to be addressed, privacy may need to be reconceptualized,²²⁰ because a model that focuses on leakage of information outside of the initial area of collection is unlikely to properly consider issues of use, within the community, of information that is necessarily divulged through participation. These types of concerns will likely become more important in the future as the personal information that is collected within virtual communities continues to grow.²²¹ There are, however, difficult issues to resolve in any conception of privacy that attempts to address these tensions. Most importantly, any such concerns must be carefully balanced against the benefit that participants obtain through enjoying a community that is tailored to their tastes and needs.

4. *Rights of Legal Enforcement*

The rule of law discourse suggests another substantive limit on autonomy that is derived from the ideal of access to justice: a requirement that citizens ought to be able to enforce their rights in the legal system.²²² This is essentially a corollary to the principle that governance should be limited by law; an idea that would have little significance if the citizen is practically prevented from challenging the actions of the government. Applying this principle to online contracts highlights potential problems with contractual

218. *Id.* at 255–56, 259–64.

219. Julie E. Cohen, *Privacy, Visibility, Transparency, and Exposure*, 75 U. CHI. L. REV. 181, 194 (2008); Penney, *supra* note 110, at 233 (discussing privacy issues that arise around actions and information within virtual environments); Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393, 1415 (2000).

220. *See* Cohen, *supra* note 219, at 194–96; Penney, *supra* note 110 (arguing that a simplified model of privacy is required in order to address new concerns about privacy in virtual spaces).

221. Lastowka & Hunter, *supra* note 15, at 72 n.386.

222. Raz, *supra* note 86, at 200–02.

terms that purport to exclude legal enforcement.²²³ This category includes terms providing “that the recipient would have no right of legal action or remedy under any circumstances,” and other “gray area” terms, such as requirements to submit to binding arbitration, exclusion of class actions, undertakings to pay attorney’s fees, and “severe curtailment of remedies,” such as “clauses limiting the remedy for a victorious plaintiff to whatever the recipient paid for a service.”²²⁴

Clauses in this broad category are relatively common in virtual community contracts. The *Second Life* Terms of Service, for example, previously required that any plaintiffs submit to binding arbitration in Linden Lab’s home state.²²⁵ Such a clause can be very effective at limiting legal redress for participants who allege that they have been wronged because arbitration is often expensive, travel to the provider’s jurisdiction may be prohibitive, and arbitrators tend to determine cases in favor of the large corporate actors.²²⁶ The district court in *Bragg* refused to uphold the binding arbitration clause in Linden’s favor, holding that it was procedurally and substantively unconscionable.²²⁷ In coming to this conclusion, Judge Robreno held that “[i]n effect, the TOS provide Linden with a variety of one-sided remedies to resolve disputes, while forcing its customers to arbitrate any disputes with Linden.”²²⁸

The decision in *Bragg* follows that of the U.S. district court in *Comb v. Paypal*, which held that a compulsory arbitration clause in the Paypal Terms of Service was unconscionable because of “a lack of mutuality in the User Agreement and the practical effects of the arbitration clause with respect to consolidation of claims, the costs of arbitration, and venue.”²²⁹ These cases

223. Radin, *supra* note 38, at 150.

224. *Id.* at 150–51.

225. *See* *Bragg v. Linden Research, Inc.*, 487 F. Supp. 2d 593, 606–07 (E.D. Pa. 2007). The updated Terms of Service provide for optional binding non-appearance-based arbitration. *See* Linden Lab, SECOND LIFE TERMS OF SERVICE, § 12 (Oct. 6, 2010), <http://secondlife.com/corporate/tos.php>.

226. *See, e.g.*, Robert Berner & Brian Grow, *Banks vs. Consumers (Guess Who Wins)*, BUSINESS WEEK, June 5, 2008, available at http://www.businessweek.com/magazine/content/08_24/b4088072611398.htm (explaining that individuals who agree to credit card terms of agreement unknowingly submit to the arbitration clauses that make it difficult to prevail against the large corporations); Alex Chasick, *Mandatory Binding Arbitration Still Sucks*, CONSUMERIST (June 9, 2008, 6:18 PM), <http://consumerist.com/5014412/mandatory-binding-arbitration-still-sucks> (highlighting claims that the vast majority of arbitrations between corporations and consumers are resolved in the corporation’s favor).

227. *Bragg*, 487 F. Supp. 2d at 611.

228. *Id.* at 608.

229. *Comb v. Paypal, Inc.*, 218 F. Supp. 2d 1165, 1173 (N.D. Cal. 2002).

suggest that courts may be increasingly willing to refuse to uphold terms that limit participants' rights of legal enforcement.²³⁰

5. *Summary of Substantive Values and a More General Application*

The stronger forms of the arguments for cyberspace self-governance suggest that the role of the state in imposing substantive limits on autonomy should be minimal to non-existent. These arguments are generally premised on the fact that individuals who disagree with the norms within a given virtual community have the ability to leave the community, a power which is much more difficult to exercise in the corporeal world. With this logic, since values are subjective, it makes little sense for the territorial state to limit the scope of autonomy and consensual participation in virtual communities.

This logic is faulty for a number of reasons. Primarily, as argued above, the deterministic assumptions that norms of virtual communities will necessarily be better than those of territorial states are fundamentally flawed. Irrespective of those assumptions, however, the territorial state continues to have some responsibility to protect its citizens and limit their autonomy, whether they are interacting with other citizens or with foreigners, online or off. Accordingly, the substantive values that a territorial state believes are important are likely to influence the boundaries of acceptable self-governance, at least for citizens of that territorial state and to the extent that any such limits can be effective.

It is not possible to provide any definitive answers as to which values should be read into virtual community governance structures. The answer will always depend upon the community context, the level of harm that participants are exposed to, and the beneficial effects, if any, of allowing the community to determine its own substantive values. The exact content and bounds of any such limits will always be highly contextual.

The rule of law analysis helps to highlight some of the more pressing tensions that surround private governance in virtual communities. The sets of values canvassed here: discrimination, speech, property, privacy, and rights of legal enforcement, are merely indicative of a much larger set of the issues that societies are continuously debating as new technologies bring changing social practices. It would be a mistake to treat any of these values as having a universal application, but this set provides a first pass that may give courts reason to pause and more closely consider the legitimacy of a contractual

230. See Ryan Kriegshauser, *The Shot Heard Around the Virtual Worlds: The Emergence and Future of Unconscionability in Agreements Relating to Property in Virtual Worlds*, 76 UMKC L. REV. 1077, 1094–1107 (2008).

framework that purports to disclaim them. The effectiveness of substantive limits on autonomy is of course widely varied, but it no longer seems plausible to claim that cyberspace is immune from the exercise of power by territorial states. This rule of law analysis supports the conclusion that territorial states have a legitimate interest in restraining the autonomy of their citizens, whether that autonomy is mediated through cyberspace or not.

IV. FORMAL LEGALITY

Because of the difficulty in articulating universally applicable substantive rights, many modern liberal rule of law theorists developed models centering on formal legality in legitimate governance instead.²³¹ This conception of the rule of law requires “that laws be declared publicly in clear terms in advance, be applied equally, and be interpreted and applied with certainty and reliability”²³² in order that the law “be capable of guiding the behavior of its subjects.”²³³ It follows that “[a]ll laws should be prospective, open, and clear” and that “[l]aws should be relatively stable.”²³⁴ These principles, stated in a number of different ways, form the standard liberal understanding of the rule of law.²³⁵ The emphasis on the law’s ability to guide the behavior of its subjects leads to two somewhat separable themes in this conception of the rule of law: an aspiration towards clarity and predictability in legal rules and, to a lesser extent, a set of due process requirements in the application of those rules.

A. PREDICTABILITY

An important component of formal legality is the ideal that laws ought to be sufficiently predictable to allow citizens to structure their lives with some degree of certainty. The rule of law “makes it possible to foresee with fair certainty how the authority will use its coercive powers in given circumstances to plan one’s individual affairs on the basis of this knowledge.”²³⁶ To enhance predictability and the liberty of legal subjects, scholars who advocate formal legality emphasize the importance of a system

231. TAMANAHA, *supra* note 98, at 119.

232. *Id.* at 34 (discussing “legal liberty”).

233. Raz, *supra* note 86, at 198 (emphasis removed).

234. *Id.* at 198–99 (emphasis removed).

235. See generally FULLER, *supra* note 86, at 38–39 (explaining failure in a legal system predominantly by reference to the clarity and regularity of law); HAYEK, *supra* note 86, at 80 (arguing that the rule of law, “[s]tripped of all technicalities,” is concerned with enabling individuals to plan their affairs).

236. HAYEK, *supra* note 86, at 80.

that provides clear, prospective rules that are well-promulgated, reasonably constant, and consistently enforced.²³⁷

When comparing the practice of governance within virtual communities against the requirements of formal legality in the rule of law, it becomes clear that private governance does not currently live up to the ideals of encouraging predictability and guiding behavior.²³⁸ This may, of course, be perfectly desirable; one can imagine that some games, for example, may be much more interesting if the rules are not completely predictable.²³⁹ Alternatively, a lack of predictability in the interests of community solidarity in relatively homogeneous communities could be acceptable. For example, a small, tight-knit community with shared understandings of appropriate behavior may not need formally articulated rules or restraints on the power of the administrator to eject members deemed to be disruptive or unwanted. In some communities, however, particularly those that foster a more diverse population and are relatively open-ended, a perceived lack of predictability may be harmful to the interests of participants and imposing limits on private governance may be justified.

1. *Clear Rules*

The requirement that rules be clearly expressed and promulgated is familiar in the liberal rule of law discourse, where the emphasis is on the ability of law to guide behavior and the ability of citizens to plan their lives.²⁴⁰ This discourse immediately highlights that the rules in virtual communities are often unclear, obscure, and difficult to understand. The contractual terms of service and end user license agreement (“EULA”) documents are usually written in dense legalese and are usually presented in a form that discourages reading.²⁴¹

237. Raz, *supra* note 86, at 198–200.

238. See Risch, *supra* note 105, at 19.

239. See Aki Järvinen, *Introducing Applied Ludology: Hands-on Methods for Game Studies*, Presentation at Situated Play, Proceedings of the DiGRA 2007 Conference: International Conference of the Digital Games Research Association, 134, 141–42 (Sept. 27, 2007) (transcript available at <http://www.digra.org/dl/db/07313.07490.pdf>) (arguing that “the emotion of suspense is a fundamental emotion of player experiences, because it is a compound emotion where the emotions of hope, fear, and uncertainty come together”).

240. FULLER, *supra* note 86, at 38–40; HAYEK, *supra* note 86, at 80; RAWLS, *supra* note 87, at 238; Raz, *supra* note 86, at 200–02; see also TAMANAHA, *supra* note 98, at 93–94.

241. Clapperton & Coronas, *supra* note 37, at 9; Fred Von Lohmann, *Machinima: Copyright and Contract in a New Medium*, Presentation at the Computer Games, Law, Regulation, and Policy Symposium (Feb. 14, 2008).

To the extent that Montesquieu is correct in saying that “[l]iberty is the right to do everything the law permits,”²⁴² virtual communities do not rate highly on an imaginary scale of liberty. Some communities may create additional terms of conduct to govern internal behavior. These terms of conduct are often more clearly enumerated than the purely contractual terms of service, but even these are often unclear and indeterminate.²⁴³ Where these codes are sufficiently clear and effective, they may be more useful in structuring a participant’s behavior within the community than the contractual terms, and may therefore more adequately satisfy the ideals of the rule of law and Montesquieu’s conception of liberty.

This leaves the question, however, of what to make of the obscure terms that form part of the formal contract but are not clearly understood by the community. There is at least an argument that the more onerous or surprising of these should not be upheld,²⁴⁴ which would force providers to make an effort to ensure that participants are aware of and understand the key rules. This type of contractual approach may not address any problems with the substantive content of EULAs and terms of service, but it is likely to at least enhance the rule of law ideal that rules be sufficiently clear and promulgated.

2. *Changing Rules*

Another problematic component of virtual community governance is the rate at which legal rules can change and the lack of responsibility that providers have to compensate any participants who may be adversely affected by rule changes. Many providers purport to have the right to modify the terms of service at any time, often without notice to the participants.²⁴⁵ Changes in these legal rules are rarely highlighted to the participant, who may

242. CHARLES DE SECONDAT MONTESQUIEU, *THE SPIRIT OF THE LAWS* 155 (Anne M. Cohler, Basia Carolyn Miller & Harold Samuel Stone eds. & trans., 1989); *see also* TAMANAHA, *supra* note 98, at 52 (explaining the importance of the rule of law as a protection from tyranny for liberal legality).

243. *See* Risch, *supra* note 105, at 31.

244. *See* KARL N. LLEWELLYN, *THE COMMON LAW TRADITION: DECIDING APPEALS* 370 (1960).

245. Andrew Jankowich, in a study of the license agreements of virtual communities, found that “[o]f the agreements surveyed in this study, 75.00% reserved to proprietors the right to modify the agreements at their discretion and 39.58% allowed proprietors to modify documents without notice to the participants who are the other less powerful party.” Jankowich, *supra* note 40, at 47.

have substantial difficulty in identifying the changes and their legal effect.²⁴⁶ This suggests that the mechanism of changing rules should be investigated, requiring, for example, that providers make clear statements about the effects of any changes and highlight modified sections in the dense legal agreements in order to enable participants to identify and understand rule changes.²⁴⁷

Apart from the difficulty in identifying changes, rule changes can have significant effects on the entitlements of participants within the virtual community. An interesting example comes from the ban on gambling within *Second Life* in July 2007.²⁴⁸ For some time, a number of participants in *Second Life* were able to profit from establishing in-world casinos, where players could gamble Linden Dollars in unregulated gaming machines. Linden Dollars, as mentioned below, are fluidly convertible with U.S. dollars, but are stated by Linden Lab to be a “limited license” right, not a currency.²⁴⁹ After some interest by the U.S. Federal Bureau of Investigation on the practice of gambling in *Second Life*,²⁵⁰ Linden introduced a rule change that prohibited any gambling outright.²⁵¹

The immediate effect of the ban was that participants who had invested in the creation of casinos were forced to close down, losing future revenues upon which they may have been relying.²⁵² Many participants complained

246. *See id.* (arguing that the lack of clarity in rule changes “seems designed to encourage participants to be responsible for their role under EULAw while discouraging them from being aware of the extent of those responsibilities”).

247. Fairfield argues that:

[T]his kind of coercive information forcing rarely helps in the context of mass-market contracts. Consumers never read the new and improved contracts that courts labour over. Requiring consumers to read lengthy contracts . . . is not a solution, it is part of the problem. The resulting transaction costs would kill many of the mass-market deals that, in the aggregate, provide an enormous benefit to society. Thus, the old judicial standby of adopting information-forcing rules that require consumers to read contracts is inadequate.

Fairfield, *supra* note 55, at 468–69.

248. Robin Linden, *Wagering in Second Life: New Policy*, SECOND LIFE BLOGS (July 25, 2007), <https://blogs.secondlife.com/community/features/blog/2007/07/26/wagering-in-second-life-new-policy>.

249. *Terms of Service*, SECOND LIFE, § 5.1 (Oct. 6, 2010), <http://secondlife.com/corporate/tos.php>.

250. *Virtual Feds Visit Second Life Casinos*, CNN.COM (Apr. 4, 2007, 9:50 AM), <http://web.archive.org/web/20070408124303/http://www.cnn.com/2007/TECH/internet/04/04/secondlife.gambling.reut/index.html> (Internet Archive copy).

251. Linden, *supra* note 248.

252. *See* Thomas Claburn, *Second Life Gambling Ban Gets Mixed Reaction*, INFORMATIONWEEK (July 26, 2007, 5:00 PM), <http://www.informationweek.com/news/internet/showArticle.jhtml?articleID=201201441>.

about the rule change, arguing that while unregulated online gambling was not permissible in the United States, there were casino operators and players who were not situated in the United States.²⁵³ Linden responded to this claim by stating that “[t]his policy applies to all use of *Second Life*. It isn’t intended to describe what is or isn’t legal for any particular resident or in any particular place. It describes what Linden Lab believes is appropriate to maintain its business requirements and to operate *Second Life*.”²⁵⁴

The longer term effects of the ban were more widely felt. Unregulated banks had become popular in *Second Life* as a result of the growing virtual economy, some of which were offering returns of between thirty and sixty percent.²⁵⁵ When Linden banned gambling, the casino operators, who were making thousands of USD equivalent Linden Dollars in profit every month, quickly sought to redeem their stored Linden Dollars, and a run on the virtual banks ensued.²⁵⁶ The biggest bank, Ginko Financial, collapsed, taking with it several thousands of U.S. investment dollars.²⁵⁷ This eventually prompted Linden Lab to introduce another rule change, banning virtual banks by prohibiting the payment of interest in-world by anyone not registered as a regulated bank by a territorial government.²⁵⁸

This example shows that rule changes can have significant effects. Certainly, gambling within *Second Life* was likely to be illegal under U.S. law and that of several other jurisdictions.²⁵⁹ Further, the unregulated banking industry within *Second Life* appeared to be completely unsustainable and

253. See Christine Hurt, *From Virtual Tax to Virtual Gambling*, CONGLOMERATE (Apr. 9, 2007), http://www.theconglomerate.org/2007/04/from_virtual_ta.html.

254. Linden, *supra* note 248.

255. David Bester, *A Virtual Crash: The Rise and Fall of Ginko Financial*, THINK MAGAZINE (Jan. 2009), available at <http://www.algorithmics.com/think/January09/Algo-THINK0109-VC-Bester.pdf>.

256. *Id.*

257. See Pixeleen Mistral, *Ginko Financial's End-Game*, ALPHAVILLE HERALD (Aug. 6, 2007), <http://alphavilleherald.com/2007/00/ginko-financial-2.html>. Media reports indicated that the total amount lost to Ginko Financial was in the vicinity of \$750,000, but this may be highly inflated: “[T]his figure doesn’t reflect actual losses. It likely includes fictitious interest to be paid out over time, and employee salaries. The average individual loss to depositors was probably in the hundreds or in some cases the low thousands.” Bester, *supra* note 255 (quoting Benjamin Duranske).

258. Kend Linden, *New Policy Regarding In-World “Banks,”* SECOND LIFE BLOGS (Jan. 8, 2008, 6:43 PM), <https://blogs.secondlife.com/community/features/blog/2008/01/08/new-policy-regarding-in-world-banks>.

259. See Unlawful Internet Gambling Enforcement Act, 31 U.S.C. §§ 5361–5367 (2006); see also Susan W. Brenner, *Fantasy Crime: The Role of Criminal Law in Virtual Worlds*, 11 VAND. J. ENT. & TECH. L. 1, 54 (2008).

resembled Ponzi schemes more than legitimate banking institutes.²⁶⁰ Both of these rule changes were likely justified as protecting Linden Lab and *Second Life* participants. The changes, however, did create real financial losses for people who were encouraged to invest in *Second Life* because of the high potential returns and the lack of prohibitions on gambling or financial markets.

In this case, it is likely that the ban on gambling was not an illegitimate evolution of *Second Life* norms:

The ban was not a frequent change; it was not as if Second Life banned entire lines of business and then reinstated them on a regular basis. The contract amendment was not arbitrary; gambling is illegal in many jurisdictions. The rule had no ex post facto effect; no one was penalized for past gambling. Additionally, the change was not targeted; it was a general rule with general application. So long as Second Life made no affirmative promises that gambling would be legal, the contractual law against gambling was no different from any legislative ban on real-world gambling, in accordance with the rule of law.²⁶¹

This analysis highlights, however, that where a change is not legitimate, imposing a requirement on the providers of virtual communities that just compensation be paid when entitlements are destroyed is not inconceivable, particularly where virtual currency is fluidly convertible into real currencies. The competing tensions are the provider's ability to make and change internal rules that evolve over time and to suit new circumstances or to comply with external requirements,²⁶² against the participants' interests in having some measure of security in their virtual assets.

3. *Emergent Behavior and Uncertain Rules*

As part of the emphasis on predictability, liberal rule of law theorists strongly disfavor ex post facto laws.²⁶³ A law that is not clearly expressed at a

260. Benjamin Duranske, a prominent commentator on *Second Life* legal news, claimed that he was "now completely certain that Ginko is paying its obligations to previous depositors with new depositors' money rather than investing that money. As such, over two years of speculation about whether Ginko is a Ponzi scheme is over—it undeniably is." Benjamin Duranske, *Law Journal Says Ginko Financial Probable Ponzi; Yield Down 60% in 16 Months*, VIRTUALLY BLIND (Feb. 23, 2007), <http://virtuallyblind.com/2007/02/23/business-law-journal-ginko/>; see also Mark Cassidy, *Virtual Bank, Real Scam?*, ILL. BUS. L.J. (Feb. 12, 2007, 5:51 PM), <http://www.law.uiuc.edu/bljournal/post/2007/02/12/virtual-Bank-real-scam.aspx>.

261. Risch, *supra* note 105, at 29.

262. See Balkin, *supra* note 28, at 2051; Bartle, *supra* note 50, at 33.

263. FULLER, *supra* note 86, at 51–62; RAWLS, *supra* note 87, at 238.

time before a citizen takes an action is unable to guide that person's behavior as "[o]ne cannot be guided by a retroactive law."²⁶⁴ This raises some immediate concerns in virtual communities, where apart from being subject to change, the rules are often enforced on an ad-hoc or retroactive basis, particularly when a new exploit is discovered, for example.²⁶⁵ When unanticipated emergent behavior results in undesirable consequences, providers may attempt to punish the participants who make use of a newly discovered bug or exploit in the platform. This is done in a way that lessens certainty. Participants may have difficulty differentiating between behavior that is rewarded with material advantage or fame within the community and behavior that will be deemed against the rules and punished after the fact.²⁶⁶

Some examples may be useful here. In the *Bragg* case, Bragg allegedly took advantage of a bug in Linden's auction management software to purchase land that had not been advertised for sale, with the lack of competition allowing him to purchase the land significantly under market value.²⁶⁷ Linden responded by terminating his account, alleging that he had taken advantage of an exploit.²⁶⁸ Alternatively, take the case of a guild in *World of Warcraft* that was accidentally given a developer item which gave them unparalleled power in the virtual world.²⁶⁹ When they used the item to beat the hardest challenges in the game, they were swiftly punished for exploitation, by permanent cancellation of their accounts.²⁷⁰

It may be that the participants in these examples should have known that their behavior would likely be punished. On the other hand, however, it is not always simple to identify wrongdoing. Significant gains are often achieved by members of virtual communities who are able to push the

264. Raz, *supra* note 86, at 198.

265. See MIA CONSALVO, CHEATING: GAINING ADVANTAGE IN VIDEOGAMES 114–16, 142–44 (2007).

266. TAYLOR, *supra* note 34, at 51 (“[M]any actions deemed ‘griefing’ or ‘exploiting’ exist on the boundary lines of the game—often in spaces in which the rule set is not clearly defined or the system itself is ambiguous.”); Humphreys, *supra* note 32, at 91 (reporting that “[t]rouble seemed to arise around the finer points of when play is actually cheating and when it is just clever, expert play from someone who knows the game inside out”).

267. *Bragg v. Linden Research, Inc.*, 487 F. Supp. 2d 593, 597 (E.D. Pa. 2007).

268. *Id.*

269. See Andres Guadamuz, *Avatars Behaving Badly*, TECHNOLLAMA (June 3, 2009), <http://www.technollama.co.uk/avatars-behaving-badly>.

270. For a similar account of developers punishing participants who were mistakenly given powers, see Sal Humphreys, *Massively Multiplayer Online Games Productive Players and Their Disruptions to Conventional Media Practices* 152–54 (2005) (unpublished Ph.D. dissertation, Queensland University of Technology), available at <http://eprints.qut.edu.au/16119/>.

boundaries and find innovative new ways of doing things.²⁷¹ There is clearly room for disagreement as to whether certain forms of emergent behavior are or ought to be prohibited. A participant who is punished for behavior that she believed to be within the scope of the rules may legitimately feel aggrieved by a provider's determination that it was not.

The prevalence of unanticipated consequences to technical changes and the propensity of participants to exploit them suggests that providers may need a certain degree of flexibility in the application and enforcement of rules in order to maintain a cohesive community. The ability to punish retroactively may be necessary in the interests of maintaining order, particularly where the exploitative behavior clearly contravenes community expectations, if not explicit rules. Enforcing retrospective rules would not necessarily contravene the ideals of the rule of law in such situations:

In the pragmatic view, a rule will be public whenever strong social agreement exists in practice, regardless of whether a legislature or a court has spoken. Similarly, if a rule exists normatively even without specific legislative enactment (as, for example, would a rule against intentional homicide), then later legislative confirmation would not necessarily mean that it would be unfair retroactive application to punish earlier transgressions. Moreover, where the line of evolution of legal interpretation is clearly foreseeable, it would not be unfair to hold people to what they can see is the emerging interpretation.²⁷²

This reasoning suggests that some leeway is required in order to allow providers to react to emergent behavior. In other cases, however, where participants are acting in accordance with both the stated rules and community standards, retroactive changes to the rules that significantly impact their interests could give rise to an obligation of compensation. The gray areas, where behavior is neither clearly within or outside of community standards, are much more difficult to satisfactorily determine. In some communities, it will be best to defer to the findings of the provider in order to maintain social cohesion; in others, particularly where the community is less cohesive or more open-ended, it may be best to take the more liberal view and allow all behavior that is not explicitly prohibited. As always, the context is important; the needs of particular communities will be different in every case. There is, however, a real tension between the need for flexibility and the serious threat posed by inconsistent application of the rules.

271. See CONSALVO, *supra* note 265, at 122–23 (explaining various motivations for cheating in games).

272. Margaret Jane Radin, *Reconsidering the Rule of Law*, 69 B.U. L. REV. 781, 815 (1989).

4. *Inconsistent Application and Discretionary Enforcement*

Perhaps the most troubling aspect of virtual community governance is that the rules on the books, the EULAs and terms of service, sometimes bear almost no resemblance to the rules in force in the community. Virtual community contracts are typically drafted in a very risk-averse manner, reserving for the provider almost total power to deal with members of the community. This often includes broad prohibitions on behavior that is commonplace within the community.²⁷³ In many cases, the provider is not interested in enforcing these contracts as written but will use them as a tool against particular participants as it sees fit.²⁷⁴ Essentially, these contracts are designed to reserve a wide range of discretionary powers for the provider, which is a concept that directly contradicts the values of formal legality in the rule of law that are generally understood to require that “similar cases be treated similarly.”²⁷⁵

Resolving the tension between the need for flexibility and the need to avoid the worst effects of inconsistent application of discretionary rules is a difficult task that speaks to the core of the tension between formal and substantive conceptions of justice.²⁷⁶ In moving away from purely positive accounts of law and responding to the need to allow, but simultaneously constrain the discretionary exercise of governance powers, the next set of values of the rule of law embrace requirements of fairness, equality, and transparency as measures of legitimacy in decision making.

273. See, e.g., Fairfield, *supra* note 55, at 462.

274. Risch, *supra* note 105, at 45 (“What providers generally want is a strict set of rules that they can enforce at will against a few users—a position directly contrary to the rule of law.”).

275. RAWLS, *supra* note 87, at 237.

276. Hayek, for example, strongly argued against the exercise of discretion in pursuit of substantive equality as threatening the impartiality and generality requirements of the rule of law. See FRIEDRICH A. VON HAYEK, LAW, LEGISLATION AND LIBERTY: A NEW STATEMENT OF THE LIBERAL PRINCIPLES OF JUSTICE AND POLITICAL ECONOMY 62–64 (1982). However, Allan contends:

Hayek’s account of the rule of law may justly be criticized for adopting an interpretation of equality—in the sense of generality or impartiality—that leaves no scope for legitimate political debate and action. By excluding redistributive economic aims and outlawing governmental powers of economic management, Hayek’s theory of constitutional freedom strips politics of the role it must play if the citizen is to be in any real sense an architect (together with others) of the scheme of justice he is expected to serve and endorse.

ALLAN, *supra* note 85, at 15.

B. PROCEDURAL FAIRNESS

This next set of formal rule of law values includes requirements of procedural fairness and “the availability of a fair hearing within the judicial process.”²⁷⁷ The rule of law requires an independent judiciary, the observation of principles of natural justice, judicial review over legislative and administrative power, easy access to courts, and limits on the discretion of the police.²⁷⁸ The rule of law encompasses “the regular, impartial, and in this sense fair administration of law,”²⁷⁹ and requires some form of due process: that is, a process reasonably designed to ascertain the truth, in ways consistent with the other ends of the legal system, as to whether a violation has taken place and under what circumstances.²⁸⁰ Only through conducting orderly trials and hearings with defined rules of evidence could the legal system “preserve the integrity of the judicial process.”²⁸¹

This second part of the ideal of formal legality focuses on the procedure through which legal norms are enforced. A key component of this aspect of the rule of law requires that laws are enforced fairly and that there are guarantees of fair hearings and due process available to those adversely affected.²⁸² Here again, private governance in virtual communities is potentially problematic. Providers are generally used to wielding absolute power, and the determination of when participants have broken the rules and what punishments are to be inflicted are very rarely subject to accountable procedural safeguards.

Bragg once again provides a good case study of potential procedural limits on a provider’s exercise of power. Essentially, *Bragg* was alleged to have broken the rules by exploiting a loophole and purchasing virtual land significantly under market value, and Linden Lab took action by cancelling his account and confiscating not only the contested land, but all his other virtual *Second Life* assets.²⁸³ *Bragg* disputed both the allegation that he had broken the rules and the penalty that was applied. Linden’s position, as the creator, enforcer, and adjudicator of the rules, makes it difficult for *Bragg* to trust that Linden’s decision was arrived at fairly and makes it altogether

277. TAMANAHA, *supra* note 98, at 119.

278. Raz, *supra* note 86, at 200–02.

279. RAWLS, *supra* note 87, at 235.

280. *Id.* at 239.

281. *Id.* at 238.

282. ALLAN, *supra* note 85, at 121 (“Conformity to [precepts of due process and equal justice] ensures a genuine—substantive—equality of all before a law that serves a coherent (if capacious and adaptable) conception of the common good.”).

283. *Bragg v. Linden Research, Inc.*, 487 F. Supp. 2d 593, 597 (E.D. Pa. 2007).

impossible for Bragg to appeal within the system of review for either the finding of guilt or the penalty imposed.

Another example comes from *World of Warcraft* in 2006, when Blizzard banned a large number of players who were running the game under a GNU/Linux operating system.²⁸⁴ Blizzard's anti-cheating software mistakenly identified these players as cheaters, and they were accordingly banned for using unauthorized third party software. The lack of due process had harmful effects on these players:

The players found the process involved in getting their accounts reinstated very opaque. They were sent form letter responses to their appeals to customer service. No indication was given that an investigation was underway and there was no way to know whether any of their complaints were being addressed. The lack of transparency and the realization that there would not necessarily be any "justice" was a source of great concern.²⁸⁵

The accounts of the affected players were eventually reinstated, along with an apology and a compensatory credit.²⁸⁶ This as an example of a desirable resolution and a satisfactory review policy. The tensions that this example highlights, however, are the damaging effects on players of a lack of due process—not just the result of being banned, but the uncertainty, the frustration of not being able to appeal the decision, and the damage to the participant's reputation and integrity that accompanies a false accusation.²⁸⁷

These examples raise an interesting set of questions. In order to provide a useful platform and create a harmonious community, the provider generally requires some discretion in the ability to create and enforce internal rules. In order for the exercise of discretion to be considered legitimate, however, the lack of procedural fairness, perceived equality, and transparency that often negatively characterizes the private exercise of power must be addressed.

Conceivably, virtual communities could create governance and oversight mechanisms that ensure that decisions to enforce the rules and punish participants are justly enforced. There will likely still be problems, however, where either these procedures do not exist or where they do not instill

284. See Anonymous, *Over 50% Cedega WoW Accounts Banned*, LINUX-GAMERS.NET (Nov. 18, 2006, 10:45 AM), <http://web.archive.org/web/20061123130644/http://www.linux-gamers.net/modules/news/article.php?storyid=1852> (Internet Archive copy).

285. Humphreys, *supra* note 34, at 157.

286. See Ty, *Blizzard Unbans Linux World of Warcraft Players*, LINUX LOOKUP (Nov. 22, 2006, 10:00 AM), http://www.linuxlookup.com/2006/nov/22/blizzard_unbans_linux_world_of_warcraft_players.

287. Humphreys, *supra* note 34, at 157.

sufficient confidence to reassure participants that the result is just. The question raised by the *Bragg* example remains: To what extent should the provider's discretion in enforcing the rules be externally reviewable?

A blanket rule that all administrative decisions are judicially reviewable would likely introduce much more overhead than is warranted, resulting in a system where the development and operation of innovative virtual communities is unduly disincentivized.²⁸⁸ Each additional measure of public oversight adds some overhead to the process, some drag to community governance. It is important to achieve a sensible balance between the desire to protect participants and the desire to encourage the development and growth of virtual communities.

It may not be necessary or desirable to bring in the whole of public administrative review processes into virtual governance decisions. Perhaps some of the ideals of administrative review could be used in the adjudication of contract law in these circumstances. It may be possible to limit broad discretionary powers in virtual community contracts or to impose restrictions on the exercise of those powers. Conceivably, if courts are able to find that virtual community contracts have been improperly terminated due to a lack of procedural fairness, then virtual communities will be prompted to implement internal procedures that engender the trust of the community. Obviously there are communities that will have no such need for procedural fairness, such as games where arbitrary action forms part of the entertainment value.²⁸⁹ But for other communities, a court may be able to evaluate with some sensitivity whether the procedures for imposing punishments or terminating subscriptions are carried out within the reasonable expectations of participants. In communities where both legitimacy and flexibility is important, it is only through introducing requirements of fairness and equality that states can ensure that discretion is legitimately exercised.

If the load on courts proves too great, establishing specialized tribunals to review these types of contractual governance issues could be an option. It seems likely, however, that only exceptional cases will continue to make it to the legal system. Accordingly, where there is significant procedural integrity in the exercise of a discretionary power in a virtual community contract, courts should probably defer to the provider's judgment.²⁹⁰ In cases where a

288. See Richard Bartle, *The Point of No Return*, TERRA NOVA (Apr. 4, 2008), http://terranova.blogs.com/terra_nova/2008/04/the-point-of-no.html.

289. See Greg Lastowka, *Rules of Play*, 4 GAMES AND CULTURE 379, 390 (2009).

290. For example, Grimmelmann argues:

significant lack of procedural fairness can be shown, however, courts may be justified in holding that the contractual power was not properly exercised, potentially relying on such limiting doctrines as unconscionability,²⁹¹ waiver,²⁹² good faith,²⁹³ or estoppel.²⁹⁴ If the contractual terms are rendered unenforceable in such exceptional cases with sufficient certainty to encourage providers to adopt reasonable safeguards on internal governance, a significant positive effect on the bulk of internal decision making may be achieved by establishing meaningful external bounds to providers' executive discretion.

V. THE ROLE OF CONSENT AND DEMOCRACY

Some conceptions of the rule of law predicate legitimacy on the consent of the governed, expressed primarily through the democratic process. In this way, consent provides substantive limits in a pluralistic system where universal values can no longer be explicitly justified.²⁹⁵ Fundamentally, “the

Although plaintiff Marc Bragg's allegations that Linden expropriated his land were explosive, Linden answered them with credible evidence that Bragg had taken unfair advantage of a bug in the land transaction system. That fact alone makes Linden's suspension of his account sensible. The case settled, but had it reached a decision on the merits, the law should have treated Linden's response as presumptively legitimate.

Grimmelmann, *supra* note 59; see also ALLAN, *supra* note 85, at 16 (arguing that courts, in practice, generally give substantial deference to the discretionary exercise of reviewable powers “in recognition of their specialist knowledge and expertise”).

291. See, e.g., Bragg v. Linden Research, Inc., 487 F. Supp. 2d 593, 597 (E.D. Pa. 2007) (holding that Linden Lab's binding arbitration clause was procedurally and substantially unconscionable).

292. See, e.g., Erez Reuveni, *On Virtual Worlds: Copyright and Contract Law at the Dawn of the Virtual Age*, 82 IND. L.J. 261, 299–300 (2007) (arguing that waiver at common law may be applicable to virtual world contracts where developers do not consistently or uniformly enforce contractual terms).

293. See U.C.C. § 1-304 (1977) (obligation of good faith); see also Meehan, *supra* note 209, ¶¶ 57–60.

294. See RESTATEMENT (SECOND) OF CONTRACTS § 90 (1981); see also Steven J. Horowitz, Bragg v. Linden's *Second Life: A Primer in Virtual World Justice*, 34 OHIO N.U. L. REV. 223, 236 (2008) (arguing that Bragg could potentially rely on an argument in estoppel to prevent Linden Lab from reneging on its assertion that *Second Life* residents own their virtual land); Kurt Hunt, *This Land is Not Your Land: Second Life, CopyBot, and the Looming Question of Virtual Property Rights*, 9 TEX. REV. ENT. & SPORTS L. 141, 155–56 (2007) (arguing that inducing participants to treat in-world currency as real money may lead to an enforceable modification to the EULA by promissory estoppel in communities like *Project Entropia* and *Second Life*); David P. Sheldon, *Claiming Ownership, but Getting Owned: Contractual Limitations on Asserting Property Interests in Virtual Goods*, 54 UCLA L. REV. 751, 779–82 (2007) (discussing the possibility of a successful promissory estoppel claim in virtual worlds).

295. TAMANAHA, *supra* note 98, at 99–100.

modern legal order can draw its legitimacy only from the idea of self-determination: citizens should always be able to understand themselves also as authors of the law to which they are subject as addressees.”²⁹⁶

Consent may be the single most important aspect of legitimacy in the governance of virtual communities. Cyberlaw theory suggests that the main benefit of the autonomy of virtual communities is the ability of participants to come together in spaces whose norms differ from those of other communities.²⁹⁷ At its libertarian extreme, this ideal holds that through consensual participation in a boundless array of potential communities, each community’s rules will more closely match the preferences of its participants than any default set of rules could. In less strong conceptions, there is still a clear recognition that the promising potential of cyberspaces is their malleability, through which individuals and communities can consensually determine their own norms and create their own meaning.

The forms through which consent can be expressed differ for any given community. Some communities, like Wikipedia for example, explicitly integrate democratic processes, complete with the massive bureaucratic overhead that such processes entail.²⁹⁸ Others, like the Internet Engineering Task Force, rely on “rough consensus” and active participation.²⁹⁹ Still other communities, like *A Tale in the Desert*, *Facebook*, and *EVE Online* have attempted to involve their participants in the generation of constitutional rules and ongoing community governance.³⁰⁰ For many other communities, maintaining ongoing consent is an intricate exercise in customer relations.³⁰¹ For still more, consent is expressed by ongoing participation in the community where the rules are dictated by the provider, a hard line “take it or leave it” approach.³⁰²

296. HABERMAS, *supra* note 88, at 449, *quoted in* TAMANAHA, *supra* note 98, at 99.

297. *See generally* Johnson & Post, *Law and Borders*, *supra* note 3.

298. *See* Malte Ziewitz, *Order Without Law*, Presentation at the Games Convention Online Conference Leipzig (Aug. 1, 2009).

299. Froomkin, *supra* note 43, at 794, 799–801.

300. *See generally* Timothy Burke, *Play of State: Sovereignty and Governance in MMOGs* 11–13 (Aug. 2004) (unpublished manuscript), *available at* <http://www.swarthmore.edu/SocSci/tburke1/The%20MMOG%20State.pdf> (discussing models of state within virtual worlds).

301. *See, e.g.*, Banks, *supra* note 186 (describing the importance of maintaining engagement amongst diverse player groups).

302. For example, the Something Awful community forums’ moderators “pride [themselves] on running one of the most entertaining and troll-free forums on the internet” by “charging a \$10 fee to filter out folks not serious about adhering to the rules, and banning those who manage to slip through and break them.” *See Forum Rules*, SOMETHING AWFUL (Jan. 1, 2006), <http://www.somethingawful.com/d/forum-rules/forum-rules.php?page=1>.

Whatever the form consent takes, its existence will almost always change the evaluation of legitimacy of community governance. Where consent does not exist, there is little theoretical reason to allow the default rules of society to be suspended or modified. Where informed consent does exist, then concerns about predictability or substantive fairness are likely to be greatly alleviated. There must be room for participants who consensually choose to participate in communities whose rules may seem strange or arbitrary.³⁰³ A good example is *EVE Online*, where internal norms include the concept that “fraud is fun.”³⁰⁴ *EVE*’s participants understand, if not at the point of creating an account then certainly before they become heavily invested in the game, that they may be defrauded by other participants at any time. This consensual understanding is the primary reason that the large-scale frauds perpetrated by *EVE*’s participants should not be understood as either theft or fraud; the loss of thousands of hours of invested time through the deceit of another is fully understood to be within the rules of participation.³⁰⁵ Fraud cannot exist because consent nullifies the action.

Some difficulties appear when consensual internal norms conflict with external social values, particularly those which are expressed as partially or completely inalienable.³⁰⁶ Where consensual rules conflict with external values, territorial states continue to have an interest in limiting autonomy. Territorial states will often limit the internal norms that are socially repugnant or that have deleterious effects on people outside of the community. Territorial states routinely limit the scope of consent in issues of discrimination, for example, or in content matters such as the sexualized depiction of underage persons. In this context, police and policymakers have begun to grapple with the apparently consensual practice of teenagers sharing

The rules of participation in the forum are vigorously but subjectively enforced and continued participation is generally understood to be at the discretion of the administrators.
Id.

303. Bartle, *supra* note 50.

304. Fairfield, *supra* note 55, at 460–61 (arguing that “the scope of acceptable behavior is not ultimately determined by the EULA. Whether ‘fraud is fun’ in a community ultimately depends on the views of a particular community. That, in turn, depends on the norms worked out between community members”).

305. *Id.*

306. See generally Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1111 (1972) (describing inalienability rules designed to prevent inefficient outcomes from significant negative externalities); Radin, *supra* note 158 (explaining partial restrictions on commodification and alienability of interests).

sexual photos of themselves with other teens.³⁰⁷ Concerns about sexual play and the exposure and exploitation of children in virtual worlds are also increasingly prominent as territorial states begin to consider what type of behavior is permissible and when regulation is necessary.³⁰⁸

The emphasis on consent in this conception of the rule of law also illustrates a key tension between the internal norms of a community and the contractual terms of service. The contractual documents that purport to govern virtual communities are somewhat problematic in that they are rarely designed to encourage readability and understanding.³⁰⁹ Moreover, they often conflict with the social norms within the community, usually because of discretionary enforcement, but also because norms within the community are continuously evolving through participation, whereas the written terms are unilaterally set in advance by the provider. As the community cultivates a separate understanding of the norms than is set out in the contractual documents, real questions of consent arise when the provider attempts to enforce the conflicting contractual provisions.

The *EVE* example can be contrasted with other cases in which consent to a purported change to default social rules is clearly not manifested. The *Bragg* case once again provides a useful example; there is a clear social norm within *Second Life* that participants own their virtual property and currency, one cultivated and encouraged by Linden Lab in its advertising materials and public statements.³¹⁰ The fine print in the Terms of Service, however, purports to disclaim any enforceable interests participants may have in virtual goods or currency.³¹¹ If the contractual terms are literally enforced, they will override the consensual social practices within the community. The proposition that Linden Lab was able to modify the default rules of property ownership unilaterally, in direct opposition to the understanding of the

307. See, e.g., Nancy Rommelmann, *Anatomy of a Child Pornographer*, REASON (July 2009), available at <http://reason.com/archives/2009/06/04/anatomy-of-a-child-pornographer> (discussing the ramifications of “sexting”—the exchange of explicit images or videos amongst teens).

308. See, e.g., FED. TRADE COMM’N, VIRTUAL WORLDS AND KIDS: MAPPING THE RISKS (2009), available at <http://www.ftc.gov/bcp/edu/pubs/consumer/alerts/alt038.shtm>; R. Bloomfield & B. Duranske, *Protecting Children in Virtual Worlds Without Undermining Their Economic, Educational, and Social Benefits*, 66 WASH. & LEE L. REV. 1175 (2009); J. A. T. Fairfield, *Virtual Parentalism*, 66 WASH. & LEE L. REV. 1215 (2009); J. M. Shaughnessy, *Protecting Virtual Playgrounds: Introduction*, 66 WASH. & LEE L. REV. 995 (2009); R. F. Wilson, *Sex Play in Virtual Worlds*, 66 WASH. & LEE L. REV. 1127 (2009).

309. Dale Clapperton, *Electronic Contracts: A Law unto Themselves?*, 130 MEDIA INT’L AUSTL., INCORPORATING CULTURE & POL’Y 102, 103 (2009).

310. See *Bragg v. Linden Research, Inc.*, 487 F. Supp. 2d 593, 595–96 (E.D. Pa. 2007).

311. *SECOND LIFE*, *supra* note 249.

community, violates the ideal of governance limited by law. Instead, it resembles governance by fiat, where the technical rules that govern interaction are determined solely by the provider, in almost-absolute discretion, and bear almost no resemblance to those understood and accepted by the community.

David Post argues strongly that rules imposed by external states are less legitimate than the rules developed by virtual communities themselves. State-imposed rules completely abandon “any notion that governments derive . . . their just power from the consent of the governed, or that the individuals to whom law is applied have the right to participate in formulating those laws.”³¹² The imposition of rules by the territorial state stifles the ability of virtual communities to develop “as true communities, with shared norms and customs and expectations characteristic of each and continually being created and re-created by the members within each.”³¹³ If the sovereignty of virtual communities is not recognized, then “no matter what steps they take to set up a fair and reasonable system for resolving virtual world disputes in accordance with newly created virtual world law, their efforts will come to nothing because they can’t create ‘real law,’ ” and users will “be stuck with the chaotic nonsense” of law imposed by various territorial jurisdictions.³¹⁴ Being able to fall back on enforceable state law risks making virtual law “play-law.” “If everyone believes that ‘real law’ from ‘real sovereigns’ is the only law that matters (or can ever matter),”³¹⁵ then “who will undertake the hard work required to set up a legal system if it’s just play-law?”³¹⁶

There are two main readings of Post’s argument. The strongest is a proposition that the rules of a virtual community can be the only legitimate source of law for participants in that community. This proposition is not particularly helpful. To suggest that internal norms will not adequately develop in the shadow of the territorial state seems to be a suspect assumption. After all, virtual communities rely upon the enforcement of territorial contract and property law, and “property and contract presuppose limits and enforcement shaped by a sovereign authority.”³¹⁷ Cyberspace self-governance and state rules “form a mesh of rules,”³¹⁸ where state rules

312. Post, *supra* note 122, at 910.

313. *Id.* at 912.

314. *Id.* at 913.

315. *Id.* (emphasis removed).

316. *Id.*

317. Radin & Wagner, *supra* note 10, at 1296.

318. JEANNE PIA MIFSUD BONNICI, SELF-REGULATION IN CYBERSPACE 199 (2008).

support, maintain, and oversee self-regulation practices.³¹⁹ This suspect assumption is unnecessary as there seems to be no reason to accept either full self-rule or total state control.³²⁰

A less forceful reading of Post's argument, however, agrees with this Article's conception of the role of consent in legitimate governance. If the internal norms that a community develops are not respected, then the ability of the community to govern itself may be harmed.³²¹ In cases where the internal rules conflict with external values, such harm may be a necessary limit to self-governance. In other cases, however, legal results that conflict with internal norms for no justifiable reason must be treated with suspicion. The example from *Second Life* seems to reflect this concern: while the community organizes itself around principles of ownership interests in land and currency, the spectre of immanent revocation by Linden is likely to seriously limit any consensual governance processes.

A further example may be found in the enforcement of bans on real money trades ("RMT"). Many virtual worlds ban the sale of virtual property for corporeal profit. Many such worlds, however, simultaneously introduce game mechanics that encourage RMT.³²² In cases where a ban on RMT is not actually enforced within the community, it may make sense not to allow its enforcement in territorial courts.³²³ For example, Sony Online Entertainment prohibits RMT between participants and explicitly disclaims any liability for destruction of the value of in-world property but provides some servers with an officially sanctioned trading hub where it is able to tax trades.³²⁴ An argument that subscribers own no value in their virtual property, based upon

319. Jeanne Bonnici argues:

The advantages of the customised regulation of self-regulation cannot be achieved however without the constant support of states and state legislation. . . . [T]here is a continuing relevance of national legal orders. States are especially indispensable in providing a general framework of legislation and legal mechanisms that ground self-regulation. It is also important that states continue acting as "watchdog" on the regulatory actions of the groups. Oversight by states is indispensable for the fair running of the customised rules. States should continue to assist in the development and maintenance of the self-regulation rules, including by continuing financial assistance.

Id. at 213.

320. Cohen, *supra* note 4, at 224.

321. Post, *supra* note 122, at 913.

322. Juho Hamari & Vili Lehdonvirta, *Game Design as Marketing: How Game Mechanics Create Demand for Virtual Goods*, 5 INT'L J. BUS. SCI. & APPLIED MGMT. 14 (2010).

323. Balkin, *supra* note 28, at 78.

324. *Station.com—Terms of Service*, SONY ONLINE ENTMT, § VII(F)(2), <http://www.station.sony.com/en/termsofservice.vm#n7> (last visited Aug. 27, 2009).

a technical reading of the EULA, may be suspect if there turns out to be a general community expectation that both avatars and property are fluidly exchangeable for real world currency.³²⁵

One of the key features of rule of law limits on governance seems to be that they exist because people believe they exist.³²⁶ This circular recognition may actually prove quite useful in evaluating appropriate regulatory responses to governance issues. In the *Second Life* example, the primary justification for enforcing the property rights of *Second Life* residents is that the residents believe they have them. The provider, by supporting and encouraging the belief that its power is limited, can be expected or compelled to uphold those expectations. In another environment, where the participants do not believe that property rights exist, then the provider or other participants have no obligations to respect the possessive rights of participants to their virtual assets. Essentially, this approach prioritizes the role of real consent in substantive governance, which distinguishes internal norms that ought to be upheld from those that should not.

Recourse to external standards to enforce disputes in virtual communities is potentially damaging to the development of internal dispute resolution mechanisms and internal governance.³²⁷ At least to the extent that internal norms do not conflict with external values, then, it may also be desirable to avoid overriding consensual internal governance with a strict literal interpretation of the contractual documents.³²⁸ Staying with a property-based example, this would mean that a contractual term that purported to remove any claim that participants may have to their virtual assets in a community where the internal norms support an entitlement to assets that the provider

325. A whitepaper commissioned for Sony Online Entertainment reveals that in its first year of operation, the StationExchange trading hub collected \$1.87 million in player transactions, providing a revenue to Sony of \$274,000. See Noah Robischon, *Station Exchange: Year One*, GAMASUTRA (Jan. 19, 2007), http://www.scribd.com/doc/23941/SOE-Station_Exchange-White-Paper-1-19; Michael Zenke, *SOE's Station Exchange—The Results of a Year of Trading Gamasutra*, GAMASUTRA (Feb. 7, 2007), http://www.gamasutra.com/features/20070207/zenke_pfv.htm.

326. H.L.A. HART, *THE CONCEPT OF LAW* 56 (2d ed. 1961) (arguing that rules become binding either because the community as a whole generally accepts them or because they are legitimately made by those who have the authority to make rules); see TAMANAHA, *supra* note 98, at 119; cf. Ronald Dworkin, *The Model of Rules*, 35 U. CHI. L. REV. 14, 20–21 (1967) (discussing and critiquing H.L.A. Hart's positivism).

327. Post, *supra* note 122, at 913.

328. See Fairfield, *supra* note 55, at 463 (arguing that the internal norms of the community should be taken into consideration when interpreting the contractual terms of service).

has not effectively negated should not be upheld.³²⁹ This approach accords well with rule of law ideals of legitimacy in governance, as well as the dominant justification for encouraging cyberspace self-rule in the first place: the proposition that better rules can be generated through consensual participation in virtual communities. Essentially, if a provider wishes to enforce certain contractual rules, it must ensure that those rules are understood and accepted by the community.

The single largest difficulty with a consent model of virtual community governance is evaluating consent in fact. Communities are not all homogeneous and determining whether a community has consented to any given norm is an impossible task. Any such evaluation is most likely to proceed on an assumption of consent, which is a factual determination of whether or not the hypothetical reasonable person, joining and participating in the community, could be deemed to consent to the rule in question. Whilst clearly not a perfect model of consensual governance, this approximation at least provides an avenue for territorial courts to examine the internal social norms of the community in relation to both external values and contractual terms. Consent provides a useful indication of the internal legitimacy of community rules that can then be used as a normative guide as to whether the territorial state ought to support a particular contractual interpretation or not.

VI. CONCLUSION

Governance within virtual communities occurs at the intersection of constraints from the market, the law, technology, internal community standards, and external social values. There is a trend in cyberlaw theory, however, that attempts to reduce the legitimacy of private governance to the drawing of borders. From the act of crossing over into cyberspace to the emphasis on private contract and property rights, these borders tend to delegitimize government intervention in the practice of governance in virtual communities. These conceptions of self-governance rely on assumptions about the technology, the market, and the communities that isolate participation from the remainder of society.

329. Duranske further argues that:

If a company wishes to profit by selling currency and land, and outright encourages users to make their real-life living in the virtual space, it cannot reasonably protest that the fine print says it is 'only a game' when faced with users who expect to extract that stored value or expect policies that genuinely protect the assets they have purchased.

See DURANSKE, *supra* note 200, at 113.

Regulatory approaches that rely on deterministic projections of any of these forces are unlikely to provide satisfactory outcomes. The importance placed upon autonomy in much of cyberlaw theory risks overlooking the importance of restraints on the exercise of power within virtual communities. The tensions that permeate these communities are governance tensions and should be addressed as governance tensions. If governance in virtual communities is to be regulated through private law, then it is desirable to analyze the continued suitability of private law through constitutional discourses which are receptive to the potential threats posed by private governance.

The values of the rule of law and the rights of citizens are continuously protected by the evolution of the private common law.³³⁰ The myriad legal determinations regarding how power can be exercised by members of society substantially construct the rights and interests of all citizens. So too, in virtual communities, the boundaries of private law doctrines mediate the relationships between participants and providers, as they do in disputes between participants. The rule of law, as a discourse that emphasizes the legitimacy of governance and appropriate limits on the exercise of power, provides a useful framework as a first step to reconceptualizing and evaluating these tensions in communities at the intersection of the real and the virtual, the social and the economic, and the public and the private.

330. DICEY, *supra* note 1, at 187–88.

WHY LICENSE AGREEMENTS DO NOT CONTROL COPY OWNERSHIP: FIRST SALES AND ESSENTIAL COPIES

Brian W. Carver[†]

I.	INTRODUCTION	1888
II.	THE FIRST SALE DOCTRINE (§ 109) AND § 117	1891
	A. BACKGROUND.....	1891
	B. THE PROBLEM POSED BY <i>MDY INDUSTRIES LLC V. BLIZZARD ENTERTAINMENT, INC.</i>	1892
	C. THE DISTINCTION BETWEEN THE INTANGIBLE COPYRIGHT AND THE TANGIBLE COPY	1896
III.	WHEN DOES TITLE TO A COPY PASS TO THE TRANSFEREE?	1898
	A. THE RESERVATION OF TITLE APPROACH	1899
	1. <i>The “Magic Words” Cases: MAI Systems Corp. v. Peak Computer, Inc. and Microsoft Corp. v. Harmony Computers & Electronics, Inc.</i>	1899
	2. <i>Wall Data Inc. v. L.A. County Sheriff’s Department and S.O.S. Inc. v. Payday, Inc.</i>	1902
	B. THE AGREEMENT CONTROLS APPROACH	1905
	1. <i>DSC Communications Corp. v. Pulse Communications, Inc.</i>	1906
	2. <i>Novell, Inc. v. CPU Distributing, Inc. and Novell, Inc. v. Unicom Sales, Inc.</i>	1908
	3. <i>MDY Industries LLC v. Blizzard Entertainment, Inc.</i>	1911
	4. <i>Vernor v. Autodesk, Inc. (9th Circuit panel)</i>	1912
	C. THE UNIFORM COMMERCIAL CODE CONTROLS APPROACH	1914
	D. THE ECONOMIC REALITIES APPROACH.....	1915
	1. <i>United States v. Wise</i>	1916
	2. <i>Microsoft Corp. v. DAK Industries</i>	1916
	3. <i>Krause v. Titleserv, Inc.: The “Incidents of Ownership”</i>	1918
	E. THE PERPETUAL POSSESSION APPROACH.....	1920
	1. <i>UMG Recordings, Inc. v. Augusto</i>	1920

© 2010 Brian W. Carver.

[†] Brian W. Carver, Assistant Professor, University of California, Berkeley, School of Information. J.D., University of California, Berkeley School of Law (Boalt Hall).

2.	Vernor v. Autodesk, Inc. (<i>district court</i>)	1923
IV.	THE ADOBE OPINIONS AND INTERPRETIVE DIFFICULTY	1925
V.	CHARACTERISTICS OF A CORRECT APPROACH TO DETERMINING COPY OWNERSHIP	1930
A.	MUST BE LOGICALLY CORRECT	1930
1.	<i>Must Not Equivocate with Respect to the Words "License" or "Software"</i>	1931
2.	<i>Must Recognize the Possibility of Ownership of a Copy Independent from Ownership of the Copyright</i>	1935
3.	<i>Must Not Look to Factors that are Wholly Orthogonal to Resolving the Issue of Copy Ownership</i>	1937
a)	Copyright Permissions are Irrelevant to a Determination of Copy Ownership	1937
b)	Particularly Problematic Permissions	1939
B.	MUST RESPECT PRECEDENT	1939
1.	<i>Must Respect Supreme Court Precedent and Should Seek to Harmonize Circuit Court Precedents</i>	1940
2.	<i>Should Heed the Warning of the Supreme Court's Recent Quanta Computer Decision</i>	1942
C.	MUST RESPECT CONGRESSIONAL CHOICES.....	1944
D.	MUST TAKE INTO ACCOUNT THAT THE COPYRIGHT ACT EXPLICITLY ENUMERATES THOSE METHODS OF DISTRIBUTION WITHIN THE EXCLUSIVE RIGHTS OF COPYRIGHT OWNERS AND A DISTRIBUTION IN WHICH TITLE TO THE COPY IS RESERVED IS NOT AMONG THEM.....	1945
1.	<i>Must Recognize the Limiting Role Played by §§ 107 Through 122 and Examine Whether Contrary Contractual Provisions are Federally Preempted</i>	1949
2.	<i>Should Preserve as Far as Possible the National Uniformity Copyright Law Seeks</i>	1951
VI.	HOW TO ADDRESS MDY INDUSTRIES LLC V. BLIZZARD ENTERTAINMENT, INC.	1953
VII.	CONCLUSION	1954
I.	INTRODUCTION	

This Article argues for an analytic approach to determine ownership of a tangible copy of a copyrighted work. Courts have been surprisingly divided on this apparently simple question and have employed several distinct and conflicting approaches, sometimes within the same Circuit.

That courts resolve the issue of copy ownership predictably is important for at least five reasons. First, only those who are “owners” of a copy are entitled to exercise the important rights of 17 U.S.C. § 109 and § 117,¹ the first sale rights, and the right to make essential-step copies and adaptations of computer programs, respectively. Both sections play a critical role in efforts by Congress to create a system of copyright that achieves a balance between the exclusive rights given to a copyright owner and the rights reserved for the user of a copyrighted work. Without a consistent approach to determine copy ownership, users of copyrighted works cannot exercise rights under § 109 and § 117 without uncertainty and a fear of liability.

Second, as illustrated by the district court opinion in *MDY Industries LLC v. Blizzard Entertainment, Inc.*,² the question of copy ownership can be determinative of whether a third party distributor of interoperable software is liable for secondary copyright infringement. If the approach to copy ownership taken by the *MDY Industries* district court is adopted by more courts, copyright owners’ monopolies may be limited only by the imaginativeness of their end-user license agreements (“EULAs”). Thus, relying on a misguided approach to copy ownership enables copyright misuse of sweeping proportions.³

Third, the current practice in software licensing of asserting that the distributor “licenses but does not sell” its software is logically incoherent. Software distributors make this assertion in order to preserve the argument that the software users—even those with a right to perpetual possession of the tangible copy—are not “owners” of the copy and thus cannot exercise rights under § 109 and § 117. If software distributors knew that they could not succeed in such claims, those distributors that had a true business need to prevent either downstream sales or § 117 copies would be incentivized to adopt a rental or leasing model. The result would be greater clarity and, potentially, economic benefits for both the distributor and the users.

Fourth, exhaustion principles such as the first sale doctrine embodied in § 109 have promoted access to knowledge, preservation of culture, and resistance to censorship. In addition, these principles have provided the basis for secondary markets that have increased beneficial competition, have served to curtail inefficient and unfair price discrimination, and have long served to decrease the information cost of ascertaining restrictions on

1. 17 U.S.C. § 109(a) (2006) (limiting the rights of § 109 to “the owner of a particular copy . . . lawfully made”); 17 U.S.C. § 117(a) (2006) (limiting the rights of § 117 to “the owner of a copy of a computer program”).

2. No. CV-06-2555-PHX-DGC, 2008 WL 2757357 (D. Ariz. July 14, 2008).

3. See generally *Lasercomb Am., Inc. v. Reynolds*, 911 F.2d 970, 974–77 (4th Cir. 1990).

tangible goods. Consequently, exhaustion has reconciled copyright law with property law policies that disfavor restraints on alienation.⁴ But since the first sale rights of § 109 are only available to owners of a copy, uncertainty and inconsistency in the adjudication of copy ownership necessarily creates uncertainty and inconsistency in the adjudication of first sale rights. As a result, these beneficial effects are at risk of being diminished or lost.

Fifth, the questions addressed in determining copy ownership are often related to a larger debate concerning whether mass consumer contracts should trump the policies embodied in statutory schemes enacted by Congress. Public policy rationales favoring freedom to contract collide with the realities of non-negotiated boilerplate EULAs that purport to restrict statutory rights. Users have grown accustomed to these rights, and Congress intended that they have them as part of its larger effort to craft a balanced Copyright Act. Inconsistently resolving the issue of copy ownership thus muddles an already difficult question.

This Article proceeds in five parts. Part II provides background information on § 109 and § 117 and gives context for the Article's arguments regarding licensing, detailed in subsequent Parts. Part III reviews the various approaches courts and commentators have adopted regarding the determination of copy ownership. Part IV presents a case study of three Adobe cases decided in the same circuit. Part V proposes three essential features to a proper approach for determining copy ownership: logical consistency, respect of precedent, and respect for congressional choices. It also criticizes the other approaches taken by courts and argues that the Perpetual Possession approach (i.e., the right to perpetual possession of a copy is the primary, if not the dispositive factor in determining copy ownership) has all three essential features. Part VI applies the Perpetual Possession approach to the facts of *MDY Industries* in order to demonstrate why the district court's holding on copy ownership was erroneous.⁵

4. See Brief of Amici Curiae American Library Ass'n et al. in Support of Plaintiff & Affirmance, *Vernor v. Autodesk, Inc.*, 621 F.3d 1102 (9th Cir. 2009) (No. 09-35969), available at http://www.eff.org/files/filenode/vernor_v_autodes/VernorAmicus.pdf; Brief of Amici Curiae of Consumers Union (CU), Electronic Frontier Foundation (EFF), & Public Knowledge in Support of the Petitioner, *Quanta Computer, Inc. v. LG Electronics, Inc.*, 553 U.S. 617 (2008) (No. 06-937), available at http://www.eff.org/files/filenode/quanta_v_lg/quanta_amicus.pdf.

5. The Ninth Circuit reversed the district court's holding on secondary copyright liability on other grounds, discussed *infra* Section II.B, while reaching the same erroneous result on the copy ownership issue. See *MDY Indus. LLC v. Blizzard Entm't, Inc.*, 629 F.3d 928 (9th Cir. 2010).

II. THE FIRST SALE DOCTRINE (§ 109) AND § 117

A. BACKGROUND

Copyright law seeks a balance between the exclusive rights given to a copyright owner and the rights reserved for the user of a copyrighted work.⁶ Section 106 enumerates those exclusive rights of reproduction, modification, distribution, public display, and public performance, but it first states that those exclusive rights are “[s]ubject to” the limitations and exceptions found in §§ 107 through 122, such as fair use and first sale.⁷ Thus, these limitations and exceptions do not encroach upon the rightful domain of the copyright owner but instead serve to define the very scope of the exclusive rights of copyright in the first instance. Sections 107 through 122 reflect public rights created through the operation of law, not permissible acts subject to the authority of a copyright owner. That is, no one needs a license to make a fair use; everyone already has a right of fair use.

The public’s expectations regarding a balanced copyright system integrally involve the first sale doctrine expressed in § 109. Even before the doctrine was enshrined in the copyright statute, it was recognized by the Supreme Court of the United States in *Bobbs-Merrill Co. v. Straus*, in which the Court rejected a book publisher’s attempt to restrict resale of a book by printing a notice in the front of the book that purported to prohibit sales for less than one dollar.⁸

The balance struck has thus long been that a copyright owner is entitled to reap her reward on the first sale of a copyrighted work, but that after such a sale, downstream owners of copies are free to sell or otherwise dispose of the possession of that copy however they see fit without needing any permission from the copyright owner. As a consequence, we have libraries, we have second-hand booksellers, and we can loan copies of books to friends, all without the requirement of a cumbersome request for permission from the copyright owner.

The Third Circuit noted in 1991 how the first sale doctrine came into conflict with software sales as home computing took off in the 1980s.⁹ The court explained that under the first sale doctrine, one could purchase a copy of a computer program and then lease or lend it to another without infringing the copyright on the program. Software producers feared that

6. *Stewart v. Abend*, 495 U.S. 207, 228 (1990); *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429 (1984).

7. 17 U.S.C. § 106 (2006).

8. 210 U.S. 339 (1908).

9. *See Step-Saver Data Sys. v. Wyse Tech.*, 939 F.2d 91, 96 n.7 (3d Cir. 1991).

software rental stores comparable to video rental stores would become commonplace, and then, because of the ease of copying software, consumers would simply rent and duplicate software rather than purchase their own copy of the program.¹⁰ Congress addressed this concern by enacting the Computer Software Rental Amendments Act of 1990,¹¹ which amended § 109 to forbid the unauthorized rental, lease, or lending of copies of computer programs for the purposes of direct or indirect commercial advantage except by non-profit libraries and non-profit educational institutions.¹²

However, Congress was not merely responsive to the concerns of the software industry but also to the needs of users. It enacted § 117 to enable the owners of copies of copyrighted computer programs to make certain types of copies and adaptations of that program essential to its use¹³:

[n]otwithstanding the provisions of section 106, it is not infringement for the owner of a copy of a computer program to make or authorize the making of another copy or adaptation of that computer program provided[] that such a new copy or adaptation is created as an essential step in the utilization of the computer program¹⁴

Thus, just as one does not need a license to make a fair use, an owner of a copy of a computer program does not need a license to make “essential step” copies. Every owner of a copy already has that right under § 117.

B. THE PROBLEM POSED BY *MDY INDUSTRIES LLC V. BLIZZARD ENTERTAINMENT, INC.*

The district court decision in *MDY Industries* tellingly illustrates the harm that can result from allowing copyright owners to leverage their copyright in an effort to control matters outside the scope of their exclusive rights under the Copyright Act.

Blizzard had to tell a very complex story to explain why MDY should be held liable for contributory copyright infringement. Blizzard distributes copies of *World of Warcraft* (“WoW”), a multiplayer online game, at most software and video game retailers for a one-time fee.¹⁵ Players also pay a

10. *Id.*

11. Computer Software Rental Amendments Act of 1990, Pub. L. No. 101-650, tit. VIII, 104 Stat. 5134, 5134–37.

12. 17 U.S.C. § 109(b) (2006).

13. Bayh-Dole Act, Pub. L. No. 96-517, § 10(b), 94 Stat. 3015, 3028–29 (1980) (amending 17 U.S.C. § 117 to substantially its present form).

14. 17 U.S.C. § 117(a)(1) (2006).

15. *World of Warcraft: Wrath of the Lich King*, was available from <http://www.bestbuy.com> for \$39.99 on July 10, 2009. Note that Blizzard’s “Quick Start

monthly fee for online access.¹⁶ MDY sold “Glider,” a software program called a “bot” that automated play of WoW. Glider automated repetitive, time-intensive parts of WoW, allowing a player’s in-game character to more quickly reach the highest level, which then provides such characters with access to advanced game content that is not otherwise available. Many players have several characters at once, and Glider allowed them to develop these additional characters without replaying certain time-consuming portions of the game.¹⁷ Glider was not a cracking or hacking tool and it did nothing to enable making copies of WoW that was not already possible. Blizzard instead complained that it was entitled to design its game in a way that prevented “skip[ping] to the end”¹⁸ and that numerous WoW players had complained to them about the presence of Glider users in the game.¹⁹

For most software, including WoW, a user’s computer makes a copy of at least portions of the software into its Random Access Memory (“RAM”) whenever the program is used. RAM copies have been held to be infringements of the copyright holder’s exclusive right to make copies.²⁰ However, RAM copies made from a copy owned by the user are permitted under § 117; for non-owners, copies incidental to use might be permitted under a provision in the EULA (if such a provision is present).

Guide” describing a downloadable ten-day free trial says, “You may upgrade your account at any time to a full version by *purchasing* a retail copy of the game at most software and video game retailers.” Blizzard Entm’t, Inc., *Quick Start Guide*, WORLD OF WARCRAFT, <https://signup.worldofwarcraft.com/trial/qs5.html> (last visited Jun. 30, 2009) (emphasis added) (on file with author).

16. Available in most countries are a month-to-month package at \$14.99 per month, a three-month plan at \$13.99 per month, and a six-month plan at \$12.99 per month. Blizzard Entm’t, Inc., *General FAQ*, WORLD OF WARCRAFT, <http://www.worldofwarcraft.com/info/faq/general.html> (last visited Nov. 23, 2010). The subscription fees for the three-month plan and the six-month plan must be paid in full at the beginning of the period. *See id.*

17. *See* First Brief on Cross Appeal at 5–7, MDY Indus. LLC v. Blizzard Entm’t, Inc., 629 F.3d 928 (9th Cir. 2010) (Nos. 09-15932 & 09-16044).

18. Response Brief of Appellees & Opening Brief of Cross-Appellants at 64, *MDY Indus. LLC*, 629 F.3d 928 (Nos. 09-15932 & 09-16044).

19. *See* Blizzard Entertainment, Inc. & Vivendi Games, Inc. Motion for Summary Judgment & Memorandum of Points & Authorities in Support at 20–21, MDY Indus. LLC v. Blizzard Entm’t, Inc., No. 06-02555 (D. Ariz. Mar. 21, 2008) [hereinafter Motion for Summary Judgment] (listing examples of user complaints).

20. MAI Sys. Corp. v. Peak Computer, Inc., 991 F.2d 511, 517–19 (9th Cir. 1993). *But cf.* Cartoon Network, LP v. CSC Holdings, Inc., 536 F.3d 121, 127–30 (2d Cir. 2008) (declining to read *MAI Systems* as holding that, as a matter of law, loading a program into a form of RAM always results in copying); Aaron Perzanowski, *Fixing RAM Copies*, 104 NW. U. L. REV. 1067 (2010) (suggesting that, in addition to durational persistence, courts should consider the extent to which RAM instantiations serve as functional substitutes for traditional copies before classifying those instantiations as copies that implicate the reproduction right).

WoW is governed by a EULA and a Terms of Use (“TOU”) that players must agree to before playing. These agreements now forbid, among other things, the use of bots.²¹ Blizzard alleged that users of WoW were not owners of their copies. Thus, Blizzard argued that the incidentally created RAM copies were permissible only by the terms of the EULA and the TOU and not under § 117. Following this argument, when users launched WoW using Glider, they exceeded the license in the EULA and TOU and created infringing copies of the game client software in their computers’ RAM.²² Thus, in Blizzard’s view, its customers, not MDY, were the *direct* infringers.

Blizzard alleged that MDY was liable for contributory copyright infringement because MDY materially contributed to this direct infringement in RAM by Glider users.²³ MDY allegedly did so by developing and selling Glider with the knowledge that Glider users would create infringing RAM copies.²⁴ Blizzard also alleged that MDY was liable for vicarious copyright infringement because MDY had the ability to stop the Glider-caused infringing activity and derived a financial benefit from that activity.²⁵

The district court’s decision turned in part on whether Blizzard’s end users were owners of their copies of the client software. The district court held that users of WoW were “licensees of the copies of the game client software” rather than owners of the copies and were not entitled to a § 117 defense because of this.²⁶ Without a § 117 right to make essential-step copies, the district court found that end users needed a license to make RAM copies of the software while playing it.²⁷ Since Glider users were violating the EULA and TOU by using a bot, the court, essentially adopting Blizzard’s view, held that any use of WoW by a Glider user that created a RAM copy (i.e., essentially all such uses) was outside the scope of the EULA/TOU and hence was infringement.²⁸ Having established then that Glider users were *direct* infringers, the court concluded without commentary that MDY was

21. Arguably the restriction violated is one on using unapproved third-party interfaces to the games. Furthermore, MDY repeatedly pointed out in its briefs that the agreements did not forbid bots when MDY began its business. MDY has argued that Blizzard should not be able to unilaterally put an end to MDY’s business by changing its agreements with its end users. *See* Motion for Summary Judgment, *supra* note 19, at 41.

22. *See id.* at 5–7.

23. *Id.* at 8–9.

24. *See id.*

25. *Id.* at 9.

26. MDY Indus. LLC v. Blizzard Entm’t, Inc., No. CV-06-2555-PHX-DGC, 2008 WL 2757357, at *9 (D. Ariz. July 14, 2008). As discussed *infra* Section V.A.1, I believe the inherent ambiguity in “licensees of the copies” leads to confusion.

27. *See id.* at *8–9.

28. *See id.* at *8–10.

liable for contributory copyright infringement, presumably because MDY materially contributed to the supposed direct infringement by Glider users.²⁹

While there are many details of the district court's decision that also deserve scrutiny, the chief problem with the result can now be made clear: if copyright owners can, through carefully worded EULAs, both retain title to copies of their copyrighted works and forbid the use of those copies in conjunction with third-party software on pain of license termination, then copyright owners can control those external markets with the stroke of a pen.

A hypothetical may make the situation clearer: Under the district court's rationale, Microsoft could add a term to its Word EULA forbidding the use of third-party software that integrates into Word. It would thereby make a secondary infringer out of Thomson Reuters, the makers of the Endnote bibliographic software that integrates into and interoperates with Word.³⁰ Microsoft could do this out of spite, to extract fees from Thomson, or to favor its own stand-alone bibliographic software—were they to make any—or even for more benign reasons such as preventing the use of third-party software that might make Word more prone to crash. Microsoft Word may be the market-leading word processing software, but it is not anti-trust grounds that would make such a licensing technique objectionable. This competition via license would be wrong merely because the copyright owner was misusing its copyright in order to control independent interoperable software made by third parties. That is precisely what Blizzard has done here, and it constitutes copyright misuse.³¹

While the Ninth Circuit looked primarily to other grounds for reversing the district court's decision on secondary copyright liability in *MDY Industries*,³² the holding that purchasers of copies of WoW were not owners of those copies for purposes of § 117 was flawed. A contrary holding on that issue alone would have prevented the harm just described.

29. *See id.* at *10.

30. *See* ENDNOTE, <http://www.endnote.com> (last visited Sept. 17, 2010).

31. *Lasercomb Am., Inc. v. Reynolds*, 911 F.2d 970, 979 (4th Cir. 1990) (“The misuse arises from Lasercomb’s attempt to use its copyright in a particular expression, the Interact software, to control competition in an area outside the copyright, i.e., the idea of computer-assisted die manufacture, regardless of whether such conduct amounts to an antitrust violation.”).

32. *MDY Indus. LLC v. Blizzard Entm’t, Inc.*, 629 F.3d 928 (9th Cir. 2010).

C. THE DISTINCTION BETWEEN THE INTANGIBLE COPYRIGHT AND THE TANGIBLE COPY

Part III, which reviews the various approaches taken by the courts to the question of copy ownership, criticizes the courts' equivocal uses of the words "license" and "software." Such equivocation is a chief reason many courts reach the wrong result on the question of copy ownership. When reading these court decisions, one must always ask, "license to what?"

Software distributors use the word "license" to describe a permanent transfer of a copy of a copyrighted work in which the transferor purportedly retains full title to and ownership of the transferred copy.³³ No other type of transfer of a tangible good is comparable.³⁴ A sale or gift involves the passing of title to the recipient. A rental, lease, or lending involves only a temporary transfer of rightful possession of the thing rented, leased, or loaned, with an expectation that the period of rightful possession will cease and the good transferred must be returned.³⁵ But the software distributor's "license"

33. See *infra* Section V.A.1.

34. There is a notion of a "license" with respect to real property, but software distributors use the term with respect to chattel. See 53 C.J.S. *Licenses* § 133 (2010). Section 133 states:

A real property license is simply the authority to enter the land of another and perform a specified act or series of acts without obtaining any permanent interest in the land. It is a permit or privilege to do what otherwise would be a trespass, a tort, or otherwise unlawful.

Id. A "bailment" is created when the owner of a chattel gives another person possession of the chattel without granting that person title to it. RESTATEMENT (THIRD) OF TRUSTS § 5 cmt. f (2003). But a bailment also provides the bailor with the right to terminate the bailment at any time and to demand that the bailee deliver the property on demand. *United States v. Alcaraz-Garcia*, 79 F.3d 769, 775 (9th Cir. 1996). No software distributor has suggested that—without cause—it could simply demand return of consumer software at any time as it could if the relationship were truly one of bailment. See *Sturm v. Boker*, 150 U.S. 312, 329–30 (1893). In *Sturm v. Boker*, the Court found:

The recognized distinction between bailment and sale is that, when the identical article is to be returned in the same or in some altered form, the contract is one of bailment, and the title to the property is not changed. On the other hand, when there is no obligation to return the specific article, and the receiver is at liberty to return another thing of value, he becomes a debtor to make the return, and the title to the property is changed. The transaction is a sale.

Id.

35. U.C.C. § 2A-103(1)(j) (2003) ("Lease" means a transfer of the right to possession and use of goods for a term in return for consideration . . ."); BLACK'S LAW DICTIONARY 970 (9th ed. 2009) ("lease, *n.* . . . 5. A contract by which the rightful possessor of personal property conveys the right to use that property in exchange for consideration."); BLACK'S LAW DICTIONARY, *supra*, at 985 ("lend, *vb.* 1. To allow the temporary use of (something), sometimes in exchange for compensation, on condition that the thing or its equivalent be

typically involves no such expectation of return. It is a wholly unprecedented form of transferring a tangible object—a fiction invented by software distributors.³⁶

The choice to call this permanent transfer of a tangible good a “license” when that word already had an established usage³⁷ has been the source of enormous confusion. No one disputes that copyright owners may license their *exclusive rights* under copyright. Furthermore, no one disputes that end users of software granted such copyright licenses are rightly called “licensees” when the word “license” is used to refer to the granting of permissions with respect to a copyright owner’s exclusive rights found in § 106. But this Article does dispute that copyright owners may “license” *tangible copies* of copyrighted works. Individuals who have a right to perpetual possession of a copy of a copyrighted work should not be called “licensees” in this sense, where the word “license” is used to refer to a permanent transfer of a copy of a copyrighted work in which the transferor purportedly retains full title to and ownership of the transferred copy. There simply is no such thing in copyright law. *One may license exclusive rights, not tangible goods.*

returned.”); BLACK’S LAW DICTIONARY, *supra*, at 1410 (“rent, *n.* 1. Consideration paid, usu. periodically, for the use or occupancy of property (esp. real property).”).

36. It is comparable to attempts to impose servitudes on chattel, but even in those cases the parties typically agreed that a sale and some form of ownership, even if burdened by the servitude, had transferred. *See* Zechariah Chafee, Jr., *Equitable Servitudes on Chattels*, 41 HARV. L. REV. 945 (1928); Molly Shaffer Van Houweling, *The New Servitudes*, 96 GEO. L.J. 885 (2008). The software distributors’ “license” appears to be a scheme designed to avoid the concerns long expressed by courts and scholars about chattel servitudes.

37. *See* 53 C.J.S. *Licenses* § 1 (2010) (“A license is a right or permission granted by some competent authority to carry on a business or do an act which, without such license, would be illegal.”). *Black’s Law Dictionary* defines “license” as:

A permission, usu. revocable, to commit some act that would otherwise be unlawful; esp., an agreement (not amounting to a lease or profit à prendre) that it is lawful for the licensee to enter the licensor’s land to do some act that would otherwise be illegal, such as hunting game.

BLACK’S LAW DICTIONARY, *supra* note 35, at 1002.

III. WHEN DOES TITLE TO A COPY PASS TO THE TRANSFEREE?

Courts and commentators have adopted various approaches regarding when title to a copy of a copyrighted work passes to the transferee. This Part focuses on an exposition of the approaches, but critiques of these are reserved for the next Part.³⁸

The various approaches used by courts are difficult to categorize precisely and often reflect a mixture of several distinct approaches. The approaches³⁹ can be roughly distinguished as follows:

On one extreme is the Reservation of Title approach. Using this approach, courts focus almost exclusively on whether the copyright holder *purports* to reserve title in the copies through an accompanying agreement. For these courts, a mere recitation by the copyright holder of a few magic words is sufficient to make virtually all other features of the transaction irrelevant.

Slightly less extreme is the Agreement Controls approach. Under this approach, courts may look to a reservation of title first, or as one of several factors, in determining ownership of a copy. However, unlike under the Reservation of Title approach, which typically looks to a single clause in the agreement, courts employing an Agreement Controls approach examine the agreement more broadly to determine ownership of copies.

Some courts use the UCC Controls approach. Since the question confronting courts concerns the passing of title to a tangible good, these courts use the Uniform Commercial Code's ("U.C.C.") Passing of Title rules to resolve the question.

Other courts use the Economic Realities approach. These courts explicitly reject that mere words recited by the agreement govern the

38. This issue has received significant scholarly attention as well. In addition to those articles cited elsewhere herein, see, for example, Lothar Determann & Aaron Xavier Fellmeth, *Don't Judge a Sale by Its License: Software Transfers Under the First Sale Doctrine in the United States and the European Community*, 36 U.S.F. L. REV. 1 (2001); Nancy S. Kim, *The Software Licensing Dilemma*, 2008 BYU L. REV. 1103; Joseph P. Liu, *Owning Digital Copies: Copyright Law and the Incidents of Copy Ownership*, 42 WM. & MARY L. REV. 1245 (2001); Christian H. Nadan, *Software Licensing in the 21st Century: Are Software "Licenses" Really Sales, and How Will the Software Industry Respond?*, 32 AIPLA Q.J. 555 (2004); R. Anthony Reese, *The First Sale Doctrine in the Era of Digital Networks*, 44 B.C. L. REV. 577 (2003); David A. Rice, *Digital Information as Property and Product: U.C.C. Article 2B*, 22 U. DAYTON L. REV. 621 (1997); David A. Rice, *Licensing the Use of Computer Program Copies and the Copyright Act First Sale Doctrine*, 30 JURIMETRICS J. 157 (1990); John A. Rothchild, *The Incredible Shrinking First-Sale Rule: Are Software Resale Limits Lawful?*, 57 RUTGERS L. REV. 1 (2004).

39. The names of the different approaches are my own inventions.

determination of ownership of a copy. Instead, they look to the substance of the transaction and consider its “economic realities.” They consider factors such as the terms of the agreement not related to copyright permissions and facts about the transaction or its consequences not discussed in the agreement at all.

Finally, two district courts have used the Perpetual Possession approach, a more theoretically satisfying application of the reasoning underlying the Economic Realities approach. Under this approach, the right to perpetual possession of a copy is determinative. These courts clarified the law not by ignoring precedent, but by following it more faithfully and by understanding what was important in earlier cases addressing this issue. The Perpetual Possession approach satisfies the three essential requirements for a proper approach to determining copy ownership: logical consistency, respect of precedent, and respect for congressional choices.

A. THE RESERVATION OF TITLE APPROACH

“Far too often, courts merely accept plaintiff’s description of the transaction as a license.” William Patry.⁴⁰

Courts that employ the Reservation of Title approach make it easy for the copyright holder to prevent transfer of ownership of a copy. These courts have looked to the agreement between the parties for guidance, but they have been satisfied with the slimmest of facts favoring the copyright holder. In some cases, it has been enough for copyright holders to simply make a representation to the court that they “license their software” and do not “sell” it. This extreme sub-variety of the Reservation of Title approach can be dubbed the “Magic Words” approach, because for these courts the magic words “we license not sell” ends the inquiry. Other courts following the Reservation of Title approach delve slightly deeper and look for specific language in the agreement reserving title in the copies to the copyright holder. If such language is found, these courts place all or almost all of the weight on this factor and find no transfer of title.

1. *The “Magic Words” Cases*: MAI Systems Corp. v. Peak Computer, Inc. and Microsoft Corp. v. Harmony Computers & Electronics, Inc.

A cursory, unsupported footnote, consisting of a single declarative sentence, in the Ninth Circuit’s *MAI Systems Corp. v. Peak Computer* opinion has done more damage to the appropriate development of the law with

40. 4 WILLIAM F. PATRY, PATRY ON COPYRIGHT § 13:25 (2010).

respect to transfer of title of copies than perhaps anything else. The court's entire analysis of the issue of copy ownership occurred in the following sentence: "Since MAI licensed its software, the Peak customers do not qualify as 'owners' of the software and are not eligible for protection under § 117."⁴¹ The court's conclusory pronouncement on this topic has been widely criticized by both leading commentators⁴² and the Court of Appeals for the Federal Circuit,⁴³ a fact recognized by the Ninth Circuit itself in a subsequent opinion.⁴⁴ The main fault articulated is that the *MAI* court's statement fails to pay heed to 17 U.S.C. § 202.⁴⁵ The *MAI* court conflated the intangible copyright with the tangible copy, and it thus failed to recognize that a licensee of one of the rights of copyright might also be an owner of a particular copy.⁴⁶

Notably, the Ninth Circuit cited to no statute or precedent for this supposed dichotomy between licensees and owners of a copy. Nevertheless, the damage has been extensive because this opinion represented one of the earliest instances of a circuit court addressing the question in the software context,⁴⁷ and thus it has been cited and followed by subsequent courts.⁴⁸

41. See *MAI Sys. Corp. v. Peak Computer, Inc.*, 991 F.2d 511, 518 n.5 (9th Cir. 1993). The court rendered this decision without the benefit of oral argument. *Id.* at 511.

42. See, e.g., 2 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 8.08[B][1][c] (1997).

43. See *DSC Commc'ns Corp. v. Pulse Commc'ns, Inc.*, 170 F.3d 1354, 1360 (Fed. Cir. 1999).

44. *Wall Data Inc. v. L.A. Cnty. Sheriff's Dep't*, 447 F.3d 769, 785 n.9 (9th Cir. 2006).

45. 17 U.S.C. § 202 (2006) ("Ownership of a copyright, or of any of the exclusive rights under a copyright, is distinct from ownership of any material object in which the work is embodied.").

46. See *DSC Commc'ns*, 170 F.3d at 1360; NIMMER, *supra* note 42, at 8.08[B][1][c]. Although some might group another MAI case, *Advanced Computer Services of Michigan, Inc. v. MAI Systems Corp.*, 845 F. Supp. 356, 367 (E.D. Va. 1994), with the one discussed in the text of this Section, I follow Nimmer in concluding that it is so wide off the mark that it does not even cite the most crucial phrase—"of a copy"—in § 117, and so I do not discuss that case further.

47. The issue was previously addressed in the software context by a circuit court in *S.O.S., Inc. v. Payday, Inc.*, 886 F.2d 1081, 1088 (9th Cir. 1989), discussed *infra* Section III.A.2.

48. See, e.g., *Wall Data*, 447 F.3d at 784–85 ("[U]nder *MAI*, if a software developer retains ownership of every copy of software, and merely licenses the use of those copies, § 117 does not apply."). There is another case, *Triad Systems Corp. v. Southeastern Express Co.*, 64 F.3d 1330 (9th Cir. 1995), which I do not discuss at length because the court merely (and erroneously) assumed that there is an insuperable distinction between licensees and owners of copies. *Id.* at 1333 ("In 1986, however, Triad began licensing rather than selling its software . . ."). It is thus not accurate to say that *Triad* "followed" *MAI* with respect to the issue of copy ownership, although the courts committed similar errors.

In the year after the *MAI* decision, the court in *Microsoft Corp. v. Harmony Computers & Electronics, Inc.*⁴⁹ furthered the unfortunate trend of short statements that fail to carefully distinguish between copyrights and copies. The *Harmony* court stated, “[e]ntering a license agreement is not a ‘sale’ for purposes of the first sale doctrine.”⁵⁰ However, the court failed to recognize that selling a copy and entering into a license agreement are not mutually exclusive: whether there is a sale for purposes of the first sale doctrine involves ownership of a particular copy, something that is completely consistent with entering a license agreement.

The court’s statement is particularly unfortunate because it appears that the defendant was selling counterfeit copies⁵¹ and, since such copies were not “lawfully made,”⁵² the first sale doctrine would not have applied for that more fundamental reason.⁵³ Still, evidence of the counterfeiting might not have changed the court’s grounds for rejecting the first sale defense, as the court implied that it would have been satisfied with mere declarations from Microsoft that it “license[d] and d[id] not sell” its software.⁵⁴

49. 846 F. Supp. 208 (E.D.N.Y. 1994).

50. *Id.* at 213 (E.D.N.Y. 1994); *see also* *Microsoft Corp. v. ATS Computers*, No. 93-1273, 1993 WL 13580935, at *5–6 (S.D. Cal. Oct. 29, 1993) (making some statements suggestive of a “magic words” approach, but ultimately deciding the issue based on the defendant’s failure to offer any evidence that a sale occurred).

51. *Harmony*, 846 F. Supp. at 212.

52. 17 U.S.C. § 109(a) (2006).

53. The court was ruling on a preliminary injunction within the first weeks of the case and had not yet held an evidentiary hearing to make a finding on the genuineness of the products sold by defendants. *Harmony*, 846 F. Supp. at 212. Microsoft made an ex parte application for seizure of the allegedly counterfeit product on January 12, 1994, and the court’s ruling here was issued February 7, 1994, just twenty-six days later. *Id.* at 209, 212.

54. *Id.* at 213. The court stated:

Plaintiff’s counsel declares that Microsoft only licenses and does not sell its Products. Entering a license agreement is not a “sale” for purposes of the first sale doctrine. Moreover, the only chain of distribution that Microsoft authorizes is one in which all possessors of Microsoft Products have only a license to use, rather than actual ownership of the Products.

Id. (citing *ISC-Bunker Ramo Corp. v. Altech, Inc.*, 765 F. Supp. 1310, 1311 (N.D. Ill. 1990)). The court also looked to the fact that Microsoft “‘established a course of conduct . . . consistent with an intention to retain all the rights associated with the grant of copyright’ of the Microsoft Products.” *Id.* (quoting *Am. Int’l Pictures, Inc. v. Foreman*, 576 F.2d 661, 665 (5th Cir. 1978)). By focusing here on the rights of copyright, the court conflated the intangible copyright with the tangible copy. Also, the case cited to, *ISC-Bunker*, contained no analysis of the copy ownership issue and simply concluded, “given the substantial evidence that ISC only licensed and did not sell its copyrighted software, the first sale doctrine has no application to Altech as a matter of law.” *ISC-Bunker*, 765 F. Supp. at 1331. Exactly what the “substantial evidence” was, is unclear from that court’s discussion. Another case that simply followed *Harmony* without analysis is *Microsoft Corp. v. Software Wholesale Club, Inc.*, 129 F.

2. *Wall Data Inc. v. L.A. County Sheriff's Department and S.O.S. Inc. v. Payday, Inc.*

Two additional Ninth Circuit opinions appear to have adopted the Reservation of Title approach in an uncritical manner: *Wall Data Inc. v. L.A. County Sheriff's Department*⁵⁵ and *S.O.S. Inc. v. Payday, Inc.*⁵⁶ In a passage where the *Wall Data* court had been discussing *MAI*, the court wrote, “[g]enerally, if the copyright owner makes it clear that she or he is granting only a license to the copy of software and imposes significant restrictions on the purchaser’s ability to redistribute or transfer that copy, the purchaser is considered a licensee, not an owner, of the software.”⁵⁷ The *Wall Data* court cited to no authority beyond *MAI*, which itself cited to no statute or precedent in support of its conclusory statement addressing ownership of copies.⁵⁸ Thus, the approach just articulated has no obvious precedential basis.

However, *Wall Data* lacks vitality as a copy ownership case because the court went out of its way to indicate that “a more fundamental reason” for its decision was the fact that the Sheriff’s Department’s actions would not constitute an “essential step” as required by § 117.⁵⁹ Thus, whether the Department owned its copies or not was not the fundamental basis of the

Supp. 2d 995, 1002 (S.D. Tex. 2000) (“If the copyright owner licenses, rather than sells, the copyrighted work, the first-sale doctrine may not apply.”) (citing *Harmony*, 846 F. Supp. at 212–14). Another case of interest is *Davidson & Assocs. v. Internet Gateway*, 334 F. Supp. 2d 1164, 1177 (E.D. Mo. 2004), *aff’d*, 422 F.3d 630 (8th Cir. 2005). In *Davidson*, the court found:

[T]he EULAs and TOU are enforceable under the UCC. First, the defendants did not purchase the Blizzard software, rather they purchased a license for the software. A sale consists in the passing of title from the seller to the buyer. When defendants purchased the games, they bought a license to use the software, but did not buy the software The EULAs and TOU in this case explicitly state that title and ownership of the games and Battle.net remain with Blizzard. Defendants do not produce sufficient evidence demonstrating that title and ownership of the games passed to them.

Id. at 1177–78 (citations omitted). This court also conflated the intangible copyright with the tangible copy, equivocating on the use of the word “software.” Reserving title to “the games” or to “Battle.net” is not the same thing as reserving title to the tangible medium in which the game is embodied.

55. 447 F.3d 769 (9th Cir. 2006).

56. 886 F.2d 1081 (9th Cir. 1989).

57. *Wall Data*, 447 F.3d at 785. That the court qualified this statement with “generally” suggests one of many reasons not to take the *Wall Data* court as stating a universally-applicable test for determining copy ownership.

58. *Wall Data* does cite to *Adobe Systems, Inc. v. One Stop Micro, Inc.*, 84 F. Supp. 2d 1086 (N.D. Cal. 2000), in a footnote in order to quote some of the declarations of members of the software industry, but this is neither statute nor controlling precedent. *Wall Data*, 447 F.3d at 787 n.9.

59. 17 U.S.C. § 117 (2006).

Court's decision.⁶⁰ For that reason, *Wall Data* should not be treated as a controlling precedent on the issue of under what circumstances title to a copy passes to a transferee, and it should not guide district courts squarely facing that question.⁶¹

Perhaps more importantly, as is explained later in this Section, the controlling precedent in the Ninth Circuit on this issue is found in an opinion never cited in *Wall Data*.⁶² Many Circuits have adopted a "rule of orderliness" that when two panel opinions are like "ships passing in the night" and a subsequent panel reaches a contrary result without mentioning the earlier case, "the later opinion is a nullity; any other rule would invite judicial chaos."⁶³ Therefore, for failure even to consider, much less follow, controlling precedent, *Wall Data* should be treated as a nullity with respect to determining the circumstances in which title to a copy passes to a transferee.

In *S.O.S. Inc. v. Payday, Inc.*, the Ninth Circuit appeared to be satisfied with any statement that might reserve title in the copy. In this case, the transferee merely rented the computer on which the disputed copy of software resided. The Ninth Circuit wrote,

60. See *Wall Data*, 447 F.3d at 785 n.9.

61. The result might be different if the two rationales in *Wall Data* were "alternate" holdings both equally supporting its decision, but the court itself chose to describe the "essential step" holding as "more fundamental." *Id.* The limitation of § 117(a)(1) applies to (i) owners of a copy that make another copy, (ii) created as an essential step, (iii) that is used in no other manner. Since the statute treats all these requirements equally, the court should be interpreted to have rested its ruling solely on the "essential step" issue because there is nothing inherently or statutorily "more fundamental" about that requirement. Thus, for the court to use such language should be taken to indicate its determination of the basis of its ruling. Further, as already indicated, on the "owner" issue the *Wall Data* court merely cited to *MAI*, which provided a single footnote which cited to nothing, and so deserves less deference than the earlier *DAK Industries* or *United States v. Wise*, 550 F.2d 1180 (9th Cir. 1997), decisions which provided far more guidance on how to make the ownership determination. It is unfortunate that the *Wall Data* court failed to cite or discuss those two cases, as then it could have helped to clarify this troublesome area of law.

62. See *Wise*, 550 F.2d 1180, discussed *infra* Section III.D.1.

63. *Grabowski v. Jackson Cnty. Pub. Defenders Office*, 47 F.3d 1386, 1400 n.4 (5th Cir. 1995) (Smith, J., concurring in part and dissenting in part). The full quote follows:

[O]ur rule of orderliness comes into play when two panels become "ships passing in the night." A subsequent panel may be unaware of an earlier holding and, consequently, may reach a contrary result. No *interpretation* is involved, as the later panel makes no mention of the earlier case. In such an instance, we can easily say that the later opinion is a nullity; any other rule would invite judicial chaos.

Id. (emphasis in original); see also *Recursion Software, Inc. v. Interactive Intelligence, Inc.*, 425 F. Supp. 2d 756, 776–77 (N.D. Tex. 2006) (discussing the Fifth Circuit's rule of orderliness).

[t]he literal language of the parties' contract provides that S.O.S. retains "*all* rights of ownership." (Emphasis added.) This language plainly encompasses not only copyright ownership, but also ownership of any copies of the software. Payday has not demonstrated that it acquired any more than the right to possess a copy of the software for the purpose of producing "product" for its customers.⁶⁴

However, the court here also repeatedly noted that Payday was to lease the computers on which the software at issue ran, and that these computers would even reside in a third-party accountant's office, and thus the parties initially may not have even contemplated that Payday would ever possess a copy of the software.⁶⁵ Thus, while the language used by the court here appears to be a strong statement of a Reservation of Title approach, the rationale behind the court's decision may be more nuanced. When evaluating the contract as a whole, the court saw a leasing arrangement that applied not only to the computer hardware, but to the individual copies of the software as well.

The approach advocated in this Article does not seek to disturb the fact that a true lease should be treated as a lease that does not give rise to § 109 or § 117 rights. A true lease is found in the *S.O.S.* case, and while the court reached the correct result, its rationale and explanation make this case less useful in guiding courts trying to determine whether an agreement denominated a "license" actually provides for the sale of copies.⁶⁶ The court's suggestion here that merely stating that one reserves "all rights of ownership" without even specifying whether one intends to reserve one's copyrights or one's rights in the tangible thing in which one's copyrighted work might be embodied, borders on a "magic words" sort of approach. But the court's explanation that Payday was engaged in the "rental of computer time" suggests that the court concluded that the transaction was really a rental or lease.⁶⁷

In summary, the Reservation of Title approach largely looks to the agreement accompanying the copy, focusing especially on whether the copyright holder makes any claim to reserve title in the copies delivered to the transferee. Too often courts adopting this approach rely solely on the word "license" as if the mere incantation of that word puts an end to the inquiry. Other courts rightly dig deeper.

64. *S.O.S. Inc. v. Payday, Inc.*, 886 F.2d 1081, 1088 (9th Cir. 1989).

65. *See id.* at 1088 n.9.

66. The *S.O.S.* court's failure to cite *Wise* also weakened its discussion of the sorts of factors one should consider when determining title to a copy.

67. *See S.O.S.*, 886 F.2d at 1088 n.9.

B. THE AGREEMENT CONTROLS APPROACH

“The terms of the contract control.” Raymond T. Nimmer (2006).⁶⁸

Some courts and commentators have explicitly or impliedly adopted a view that the agreement that accompanies a copy of a copyrighted work completely controls whether title to the tangible thing passes to the recipient or remains with the copyright holder. On this view, one looks almost exclusively to the four corners of the agreement to determine if the possessor of a copy of a copyrighted work is also the owner of that copy.

For example, Raymond T. Nimmer has written, “since Section 117 does not delineate when a person becomes an owner of a copy of a computer program, whether ownership is transferred will be determined in light of and controlled by the agreement of the parties. The terms of the contract control.”⁶⁹ He buttressed this view with the further proviso that where an agreement is silent on copy ownership, “courts should look to the context of the transaction, including any established industry practice of licensing rather than selling copies.”⁷⁰

68. RAYMOND T. NIMMER, *THE LAW OF COMPUTER TECHNOLOGY* § 1:115 (2006).

69. *Id.*

70. *Id.* Raymond T. Nimmer’s view has apparently developed over the years, as several courts quote earlier versions of his text for what appear to be different views. Nimmer has written:

Ownership of a copy should be determined based on the actual character, rather than the label, of the transaction by which the user obtained possession. Merely labeling a transaction as a lease or license does not control. If a transaction involves a single payment giving the buyer an unlimited period in which it has a right to possession, the transaction is a sale. In this situation, the buyer owns the copy regardless of the label the parties use for the contract. Course of dealing and trade usage may be relevant, since they establish the expectations and intent of the parties. The pertinent issue is whether, as in a lease, the user may be required to return the copy to the vendor after the expiration of a particular period. If not, the transaction conveyed not only possession, but also transferred ownership of the copy.

Applied Info. Mgmt. v. Icart, 976 F. Supp. 149, 154 (E.D.N.Y. 1997) (citing RAYMOND NIMMER, *THE LAW OF COMPUTER TECHNOLOGY* § 1.18[1], at 1-103 (1992)). A similar sentiment is echoed in the *DSC Commc’ns* opinion. *DSC Commc’ns Corp. v. Pulse Commc’ns, Inc.*, 170 F.3d 1354, 1362 (Fed. Cir. 1999) (“One commentator has argued that when a copy of a software program is transferred for a single payment and for an unlimited term, the transferee should be considered an ‘owner’ of the copy of the software program regardless of other restrictions on his use of the software.” (citing RAYMOND T. NIMMER, *THE LAW OF COMPUTER TECHNOLOGY* § 1.24[1], at 1-143–44 (3d ed. 1997))). Such language no longer appears in the 2006 or 2009 editions, and as a “looseleaf” treatise whose pages are updated and often discarded, it is difficult to trace the development of Prof.

When courts adopt this view, it often takes the form of evaluating the terms of the agreement to determine what restrictions are placed on the copyright permissions granted to the recipient. Part V explains why such an approach is misguided. First, I review some examples of this approach in particular cases.

1. DSC Communications Corp. v. Pulse Communications, Inc.

In *DSC Communications Corp. v. Pulse Communications, Inc.*,⁷¹ the Federal Circuit wrote,

[p]lainly, a party who purchases copies of software from the copyright owner can hold a license under a copyright while still being an “owner” of a copy of the copyrighted software for purposes of section 117. We therefore do not adopt the Ninth Circuit’s characterization of all licensees as non-owners. Nonetheless, the *MAI* case is instructive, because the agreement between MAI and Peak, like the agreements at issue in this case, imposed more severe restrictions on Peak’s rights with respect to the software than would be imposed on a party who owned copies of software subject only to the rights of the copyright holder under the Copyright Act. And for that reason, it was proper to hold that Peak was not an “owner” of copies of the copyrighted software for purposes of section 117.⁷²

Raymond T. Nimmer’s evolving view. However, the 2006 and 2009 editions contain numerous statements that suggest he now advocates a Reservation of Title approach. *See* NIMMER, *supra* note 68, § 1:115, at 1-302 (“[I]f the terms of a license provide that ownership of the copy and the copyright remain in the copyright owner, the provisions preclude transfer of title to the copy of the licensee.”); RAYMOND T. NIMMER, *THE LAW OF COMPUTER TECHNOLOGY* § 1.110, at 1-268–69 (4th ed. 2009). Prof. Raymond T. Nimmer was retained as Autodesk’s expert in the *Vernor* case, and the district court there also noted that his views on this issue appear to have changed over time. *See Vernor v. Autodesk, Inc.*, No. C07-1189RAJ, 2009 WL 3187613, at *1–2, 9–11 (W.D. Wash. Sept. 30, 2009) (“While Autodesk asks the court to defer to Mr. Nimmer’s views, it does not acknowledge that those views have proven malleable.”).

71. 170 F.3d 1354.

72. *Id.* at 1359 (applying Fourth Circuit law, but finding none on point); *see also* Adobe Sys. v. One Stop Micro, Inc., 84 F. Supp. 2d 1086, 1091 (N.D. Cal. 2000) (“These numerous restrictions imposed by Adobe indicate a license rather than a sale because they undeniably interfere with the reseller’s ability to further distribute the software.”). In *One Stop Micro*, the court held:

[B]ased upon the undisputed evidence submitted by Adobe regarding the intent of the parties in entering into the agreement, trade usage, the unique nature of distributing software, as well as the express restrictive language of the contract, the OCRA is a licensing agreement. Thus, contrary to One Stop’s assertions, the OCRA does not represent a first sale between the reseller and Adobe.

84 F. Supp. 2d at 1092; *see also* Adobe Sys. v. Stargate Software, Inc., 216 F. Supp. 2d 1051, 1055 (N.D. Cal. 2002) (“The determination of ownership in turn is based primarily on an

Here, the Federal Circuit did not conflate the intangible copyright with the tangible copy and appears to have been focused on the correct issue. The court misstepped, however, when it allowed consideration of a restriction on a permission related to a right of copyright to guide its determination of ownership of a copy. The court wrote,

[T]he DSC-Ameritech agreement provides that Ameritech shall “not provide, disclose or make the Software or any portions or aspects thereof available to any person except its employees on a ‘need to know’ basis without the prior written consent of [DSC]” Such a restriction is plainly at odds with the section 109 right to transfer owned copies of software to third parties.⁷³

The Federal Circuit noted that the agreement purported to prohibit the end users of the software “from using the software on hardware other than that provided by DSC.”⁷⁴ The Supreme Court⁷⁵ and other courts, however, have expressed concern about the restraint of trade and efforts to extend the copyright (or patent) monopoly beyond the covered work to unrelated goods.⁷⁶ It is perplexing that the Federal Circuit failed to consider that problem here, especially because the court was bound to apply Fourth Circuit law, where a leading copyright misuse case, *Lasercomb*, was decided.⁷⁷ Had the

examination of the OCRA, the agreement between Adobe and its distributors.”); *Icart*, 976 F. Supp. at 155 (finding that where no terms defined the transfer of a copy and where payment was divided into three different payment periods, ownership of a copy was an issue of fact precluding summary judgment). Note that Raymond T. Nimmer was Adobe’s expert in *One Stop Micro*. 84 F. Supp. 2d at 1091. These cases are discussed herein as exemplars of the Reservation of Title approach, but these statements show that the approaches are more of a continuum than distinct categories, and that in particular, the Reservation of Title approach could be thought of as simply a more formalistic variety of the Agreement Controls approach in which the reservation of title clause in the agreement is given dispositive weight.

73. *DSC Commc’ns*, 170 F.3d at 1361. Looking to restrictions on subsequent distribution to determine copy ownership is misguided in several ways. See *infra* Part V.

74. *DSC Commc’ns*, 170 F.3d at 1361.

75. See *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 441 (1984) (finding it extraordinary to suggest that the Copyright Act conferred upon all copyright owners collectively, the exclusive right to distribute VTRs simply because they may be used to infringe copyrights, and finding that to construe copyright law as Universal urged would have enlarged the scope of Universal’s statutory monopolies to encompass control over an article of commerce that was not the subject of copyright protection); *Motion Picture Patents Co. v. Universal Film Mfg. Co.*, 243 U.S. 502 (1917); *Bauer & Cie v. O’Donnell*, 229 U.S. 1 (1913); *Bobbs-Merrill Co. v. Straus*, 210 U.S. 339 (1908).

76. See *Alcatel USA, Inc. v. DGI Techs., Inc.*, 166 F.3d 772 (5th Cir. 1999) (affirming the district court’s finding that DSC used its copyrights to indirectly gain commercial control over its non-copyrighted microprocessor cards, which constituted misuse).

77. On the non-patent issues, such as the copyright and state law issues, the Federal Circuit follows the law of the circuit from which the appeal is taken; in this case, the Fourth Circuit. *DSC Commc’ns*, 170 F.3d at 1359 (citing *Glaxo, Inc. v. Novopharm, Ltd.*, 110 F.3d

court been confronted by a Microsoft Windows license agreement that purported to require that the software only be used on computers provided by Microsoft, would the court have enforced the agreement and prevented the entire computer industry from functioning as we know it?⁷⁸ If not, and market power is the distinguishing factor, then the court should have at least noted that such a restriction could only be enforced where not anti-competitive.

2. *Novell, Inc. v. CPU Distributing, Inc. and Novell, Inc. v. Unicom Sales, Inc.*

*Novell, Inc. v. CPU Distributing, Inc.*⁷⁹ provides another example of the Agreement Controls approach. This case involved the resale of Novell's NetWare software by the defendant, who had acquired the software from original equipment manufacturers ("OEMs"). Novell alleged that the OEMs were mere licensed distributors rather than owners of the copies. The court wrote,

To determine whether the transfer of copies of Novell software from Novell to its OEMs is a sale, the Court must construe the OEM Agreement between Novell and each of the OEMs. The OEM Agreement provides that it will be "construed in accordance with the laws of the State of Utah"⁸⁰

Here the opinion provides an example of one of the deficiencies of the Agreement Controls approach. Such an approach appears to require resort to the varying laws of the states. In this case, the court looked to Utah law, but since the choice of law depends on the jurisdiction specified in the agreement, in other cases courts have looked to the canons of contract construction from other states.⁸¹

In this case, though, the particular choice of law did not appear to determine the result. Once the court determined to look to the agreement, the court found "unambiguous sales language" throughout the agreement,

1562, 1572 (Fed. Cir. 1997), and *Atari, Inc. v. JS & A Group, Inc.*, 747 F.2d 1422, 1439–40 (Fed. Cir. 1984) (en banc); see also *Lasercomb Am., Inc. v. Reynolds*, 911 F.2d 970 (4th Cir. 1990).

78. As of this writing one is still free to purchase computer hardware capable of running Microsoft operating systems from Hewlett-Packard, Dell, Acer, Lenovo, Toshiba, Apple, and from numerous components manufacturers.

79. No. H-97-2326, 2000 U.S. Dist. LEXIS 9975 (S.D. Tex. May 4, 2000).

80. *Id.* at *11.

81. *Applied Info. Mgmt., Inc. v. Icart*, 976 F. Supp. 149 (E.D.N.Y. 1997) (applying New York law); *Novell, Inc. v. Network Trade Ctr., Inc.*, 25 F. Supp. 2d 1218, 1229–31 (D. Utah 1997) (applying the U.C.C.).

citing nearly two dozen examples.⁸² Furthermore, there are two features of the court's reasoning that are more characteristic of the Economic Realities approach.⁸³ First, the court seemed swayed by the fact that all risk of loss passed to defendants upon delivery of the copies.⁸⁴ Second, nothing in the contract suggested that Novell was providing only temporary possession of the copies to defendants and expected the copies returned to Novell if unsold. However, given the court's overall focus on allowing the terms of the agreement to control, the opinion is better categorized with the Agreement Controls cases.

This opinion wisely avoided looking to restrictions on the rights of copyright to determine copy ownership. For example, the court considered the term of the agreement that provided that "OEM agrees not to provide Novell Products or any part or copies thereof to any third party without the prior written consent of Novell."⁸⁵ Other courts have viewed this sort of term restricting transfers as nearly decisive.⁸⁶ Those courts have accordingly found that the transferees are not owners of the copies.⁸⁷ But this court recognized the lesson of Supreme Court precedent and wrote, "[a]s was true in the *Bobbs-Merrill* case, this provision may contractually obligate the OEM not to transfer the copies except under the circumstances set forth in the OEM Agreement, but it does not preclude the application of the 'first sale doctrine' as codified in § 109(a)."⁸⁸

In *Novell, Inc. v. Unicom Sales, Inc.*,⁸⁹ the court applied the Agreement Controls approach even though it gave the impression it was looking at something more. The court wrote, "[i]n determining whether a transaction is a sale or a license, the Court reviews the substance of the transaction, rather than simply relying on the plaintiff's characterization of the transaction."⁹⁰

82. See *CPU Distrib.*, 2000 U.S. Dist. LEXIS 9975, at *13-17.

83. See *infra* Section III.D.

84. *CPU Distrib.*, 2000 U.S. Dist. LEXIS 9975, at *16.

85. *Id.* at *17 n.7.

86. E.g., *MDY Indus. LLC v. Blizzard Entm't, Inc.*, No. CV-06-2555-PHX-DGC, 2008 WL 2757357, at *7-10 (D. Ariz. July 14, 2008).

87. E.g., *id.*

88. *CPU Distrib.*, 2000 U.S. Dist. LEXIS 9975, at *17 n.7; see also *Applied Info. Mgmt., Inc. v. Icart*, 976 F. Supp. 149 (E.D.N.Y. 1997). The *Icart* court was also careful not to fall into some of the over-simplifications and logical errors of earlier decisions, and largely relied on the agreement in an effort to determine copy ownership. *Icart*, 976 F. Supp. at 149. But ultimately the court decided that there was a genuine issue of material fact as to whether the copy was owned and this precluded summary judgment. *Id.*

89. No. 03-2785, 2004 U.S. Dist. LEXIS 16861 (N.D. Cal. Aug. 17, 2004).

90. *Id.* at *28-29 (citing *United States v. Wise*, 550 F.2d 1180, 1190 (9th Cir. 1977)).

Unicom Sales involved Novell software acquired under two different agreements. The defendants acquired some copies from Frederick County Public Schools (“FCPS”), who in turn had received the software from Novell under an agreement titled the “School License Agreement” (“SLA”). In examining the substance of that initial transaction, the court wrote that the license agreement

granted FCPS a license to copy the software for use only by its students and employees. The license was for a specific term of one year, and required payment of an annual license fee. When the SLA expired, FCPS was required to return the software to Novell. All of these terms are consistent with a license, rather than a sale.⁹¹

The court thus held that the first sale doctrine did not apply to defendants’ distribution of such software.⁹² The *Unicom Sales* court focused on the terms of the agreement that restricted use to students and employees, charged an annual fee for the software, and required that the users return the software at the end of the term. This focus on contractual terms reflects the court’s use of the Agreement Controls approach.⁹³

The defendants had also acquired Novell software from parties who were termed “Qualified Educational Institutions” under Novell’s Direct Education Order program (“DEO program”).⁹⁴ Unlike the SLA software which involved FCPS itself signing an agreement with Novell, the Qualified Educational Institutions were not required to enter into a separate written license agreement, although a printed “software license” was included in the box in which the software was packaged.⁹⁵

91. *Id.* at *30.

92. *Id.* at *31.

93. To the extent the court here looked at the one-year term, the annual license fee, and the requirement of return as part of a true rental or leasing arrangement, the Perpetual Possession approach, described *infra* Section III.E, would also have no quarrel with such considerations and would reach the same result. *See* *Stuart Weitzman, LLC v. Microcomputer Res., Inc.*, 510 F. Supp. 2d 1098 (S.D. Fla. 2007), *vacated*, 542 F.3d 859 (11th Cir. 2008) (dismissing for lack of subject matter jurisdiction). This case represents another instance where the court appeared willing to look to the agreement, specifically at the restrictions on Stuart Weitzman’s permissions with respect to rights of copyright, in order to determine title to the copies in Stuart Weitzman’s possession. *Id.* However, the court ultimately applied factors consistent with the perpetual possession approach. *Id.* at 1107–09. In *Unicom Sales*, there was no signed agreement between the parties, and so the court turned instead to the factors identified in *Krause v. Titleserv, Inc.*, 402 F.3d 119 (2d Cir. 2005), discussed *infra* Section III.D.3.

94. *Unicom Sales*, 2004 U.S. Dist. LEXIS 16861, at *6, *12 n.3.

95. *Id.* at *31.

At this point in the opinion the court displayed its narrower focus on the agreement more clearly: “the Court will proceed to examine *the terms of Novell’s Software License* to determine whether it describes a sale or a license.”⁹⁶

The court then considered the following:

The agreement grants the right to use the software, for an unlimited period of time, but only within the customer’s organization. If the customer breaches any of the terms, the license is automatically terminated, and the customer must either destroy the software or return it to Novell. The agreement expressly states that any rights not expressly granted are reserved to Novell.⁹⁷

The court notes that *SoftMan* had found the perpetual possession factor to be key, but attempted to distinguish *SoftMan* by emphasizing that “[h]ere, although the Software License is not for a specific time period, it expressly requires return or destruction of the software if the license is terminated.”⁹⁸ The court concluded that “[t]he limited rights in the software, as set forth in the Software License, describe a license in the software, rather than a sale.”⁹⁹

This analysis of the DEO software agreement applies an Agreement Controls approach, emphasizing the term of the agreement that required return in event of breach.

3. MDY Industries LLC v. Blizzard Entertainment, Inc.

In the previously mentioned case of *MDY Industries*, some users of Blizzard’s World of Warcraft multi-player online game used a bot program developed by MDY Industries to automate certain aspects of the game play.¹⁰⁰ The district court addressed whether a purchaser of Blizzard’s World of Warcraft (“WoW”) software was an “owner” of his copy.¹⁰¹

The court wrote that “*Wall Data* provides a two-part test” to determine copy ownership, which the court applied to determine that purchasers of WoW were not owners of their copies of the software.¹⁰² The *Wall Data*

96. *Id.* at *35 (emphasis added). The court later reiterated the need to look to substance, but did so only in order to dismiss the label on the agreement and not in order to focus on the realities of the transaction beyond the agreement’s terms. *Id.* at *37.

97. *Id.* at *35.

98. *Id.* at *36 n.10.

99. *Id.* at *35. Setting up a dichotomy between “a license in the software” and “a sale” comes dangerously close to the logical errors that will be discussed *infra* Section V.A.

100. *MDY Indus. LLC v. Blizzard Entm’t, Inc.*, No. CV-06-2555-PHX-DGC, 2008 WL 2757357, at *8 (D. Ariz. July 14, 2008). For a full summary of the facts, see *supra* Section II.B. The case is also analyzed *infra* Part IV.

101. *Id.*

102. *Id.*

court did not explicitly call its approach a “test” and even qualified its approach with the word “generally.”¹⁰³ Nevertheless, the *MDY Industries* court looked at the following factors in deciding that purchasers of WoW did not own their copies of the software: First, the EULA accompanying the software expressly stated that “[a]ll title, ownership rights, and intellectual property rights in and to the Game and all copies thereof . . . are owned or licensed by Blizzard.”¹⁰⁴

Second, the *MDY Industries* court looked to restrictions on the transfer and use of the software. According to the EULA,

The user may transfer his “rights and obligations” under the EULA only by transferring the original media containing the game client software along with all original packaging and all manuals or other documentation distributed with the software; the user must delete all copies and installations of the software from his computer; and the recipient of the software must agree to the terms of the EULA.¹⁰⁵

Third, the court noted that the TOU placed additional restrictions on the use of the software, restrictions that it felt were at least as severe as the restrictions in *Wall Data*.¹⁰⁶

The Court concluded, therefore, that users of WoW were “licensees of the copies of the game client software” and were not entitled to the § 117 defense because of this.¹⁰⁷

4. *Vernor v. Autodesk, Inc. (9th Circuit panel)*

While the *Vernor* district court’s approach to copy ownership will be heralded in this Article as perhaps the best analysis of the precedents to date, the Ninth Circuit panel that reversed the district court committed nearly every error this Article warns about as it adopted a narrow version of the Agreement Controls approach.

Timothy Vernor sold used comic books, collectibles, and software on eBay, the online auction site.¹⁰⁸ Vernor purchased several used copies of Autodesk, Inc.’s AutoCAD software from Autodesk’s direct customers, and

103. *Wall Data Inc. v. L.A. Cnty. Sheriff’s Dep’t*, 447 F.3d 769, 785 (9th Cir. 2006).

104. *MDY Industries*, 2008 WL 2757357, at *8.

105. *Id.* at *9.

106. *Id.* at *10.

107. *Id.* The inherent ambiguity in “licensees of the copies” leads to confusion. *See infra* Section V.A.1.

108. Appellee’s Brief at 1, *Vernor v. Autodesk, Inc.*, 621 F.3d 1102 (9th Cir. 2010) (No. 09-35969), available at http://www.eff.org/files/filenode/vernor_v_autodes/Vernor%20Appellee%27s%20Br.pdf.

he resold the copies on eBay.¹⁰⁹ Autodesk sent eBay several notices of claimed infringement asserting that Vernor's online resale of its software infringed Autodesk's copyright.¹¹⁰ In response, eBay terminated Vernor's pending sales and eventually shut down his online business, cutting off his primary source of income.¹¹¹

While the district court held that Vernor was an owner of the copies of AutoCAD and therefore entitled to resell them under the first sale doctrine,¹¹² the Ninth Circuit reversed,¹¹³ adopting a new test for determining copy ownership¹¹⁴ to "reconcile[]" earlier conflicting precedents.¹¹⁵

The Ninth Circuit panel declared,

We hold today that a software user is a licensee rather than an owner of a copy where the copyright owner (1) specifies that the user is granted a license; (2) significantly restricts the user's ability to transfer the software; and (3) imposes notable use restrictions.¹¹⁶

The Ninth Circuit's summary of its application of the three-prong test indicates that it regarded the first factor as equivalent to purporting to reserve title to the copy.¹¹⁷ Other courts employing an Agreement Controls approach have examined the other two factors.¹¹⁸ However, those courts employing an Agreement Controls approach have typically examined the entire agreement, including things such as the period of authorized use, whether return is required, the amount of payment and whether it is periodic, and who bears the risk of loss, among other terms. The Ninth Circuit panel instead takes a narrower approach than any other court applying the Agreement Controls approach and focuses on just the three aspects of the Agreement contained in the citation above.

109. *Vernor*, 621 F.3d at 1102.

110. Appellee's Brief, *supra* note 108, at 1.

111. *Id.*

112. *Vernor v. Autodesk, Inc.*, 555 F. Supp. 2d 1164, 1170–71 (W.D. Wash. 2008), *rev'd*, 621 F.3d 1102 (9th Cir. 2010).

113. *Vernor*, 621 F.3d at 1102.

114. *Id.* at 1110–11.

115. *Id.*

116. *Id.* at 1111.

117. *Id.* at 1112 ("Thus, because Autodesk reserved title to Release 14 copies and imposed significant transfer and use restrictions, we conclude that its customers are licensees of their copies of Release 14 rather than owners.")

118. *See supra* Section III.A.

C. THE UNIFORM COMMERCIAL CODE CONTROLS APPROACH

“The retention or reservation of title by a seller of goods notwithstanding shipment or delivery to the buyer under Section 2-401 is limited in effect to a reservation of a ‘security interest.’ ”¹¹⁹

An analogue of the Agreement Controls approach is found in a handful of decisions that are just as formalistic about following the dictates of a given controlling document. But that document is not a particular agreement; it is the U.C.C. itself. This approach tends to arise when courts are more concerned with the goods being transferred and less focused on the copyrighted works that may be embodied in those goods. It also tends to arise when there is no written agreement from which to determine the passing of title, and so some other source of title passage rules are sought. Finally, this approach seems to be more often applied where there is less of a dispute about whether a sale occurred or was attempted, and more of a question of who currently owns the underlying goods.

Section 2-401 of the U.C.C. covers the “Passing of Title” and provides several rules for the passage of title to a good. One of the chief rules is that any retention or reservation by the transferor¹²⁰ of the title in goods shipped or delivered to the transferee is limited in effect to a reservation of a security interest.¹²¹ That is, in a permanent transfer of a tangible good, reservation of full title is simply not permitted under the U.C.C.¹²² Further, where one pays full price in one lump sum, no security interest exists and the transferee is simply deemed to hold title to goods upon full payment.

119. U.C.C. § 1-201(35) (2001).

120. Where the U.C.C. speaks of seller and buyer, this Article will speak of transferor and transferee so as not to bias the discussion, which in many cases is thought of as a question of whether there was a “sale” at all. What the U.C.C. calls a sale, this Article will call a “permanent transfer,” as that accurately describes the nature of most so-called “licenses.” The aim is to analogize from the U.C.C. principles, not necessarily to claim that they apply as written.

121. U.C.C. § 2-401(1) (2001); U.C.C. § 1-201(35) (“The retention or reservation of title by a seller of goods notwithstanding shipment or delivery to the buyer under Section 2-401 is limited in effect to a reservation of a ‘security interest.’ ”); *In re Telemart Enters., Inc.*, 524 F.2d 761, 765 (9th Cir. 1975) (“A reservation of title by the seller is to be regarded as reservation of a security interest.” (citing UCC § 2-401(1))).

122. Autodesk’s brief in answer to Vernor’s petition to the Ninth Circuit for rehearing en banc argued that the “Unless otherwise explicitly agreed” language of section 2-401(2) would permit reservation of title after sale and delivery. Answer to Petition for Rehearing En Banc at 23, *Vernor v. Autodesk, Inc.*, 621 F.3d 1102 (9th Cir. 2010) (No. 09-35969), available at <http://www.scribd.com/doc/43215193>. On the contrary, the U.C.C. contemplates that a buyer and seller might vary the *timing* of title transfer so that it need not always be contemporaneous with physical delivery, but nothing in that section suggests a buyer and seller could agree that title would *never* transfer, even after a fully completed sale and delivery.

Another U.C.C. title-passage rule is that, unless otherwise explicitly agreed, title to the goods passes at the time and place at which the transferor completes performance with reference to the delivery of the goods, despite any reservation of a security interest. Title passes upon delivery at destination only if the contract so requires. Otherwise, it passes upon shipment.¹²³

Applying these title-passage rules, courts have concluded that the title to a rug containing a copyrighted design passed in accordance with § 2-401(2),¹²⁴ that title to books passed to a buyer upon physical delivery of the books per Idaho Code § 28-2-401(2) (Idaho's enactment of the relevant U.C.C. section),¹²⁵ and that title to a computer and software was determined pursuant to both California's enactment of the U.C.C., Cal. U. Comm. Code 2401(1)–(3),¹²⁶ and to the District of Columbia's enactment, D.C. Code Ann. § 28:2-401.¹²⁷

D. THE ECONOMIC REALITIES APPROACH

“Ownership of a copy should be determined based on the actual character, rather than the label, of the transaction by which the user obtained possession. Merely labeling a transaction as a lease or license does not control.” Raymond T. Nimmer (1992).¹²⁸

While the Reservation of Title and the Agreement Controls approaches elevate form over substance, a slight majority of courts have been willing to look beyond the four corners of the agreement and look instead to “the economic realities” of the transaction. The factors considered can include terms of the agreement not related to copyright permissions, as well as facts about the transaction not discussed in the agreement at all.

The leading articulations of the approach come from two Ninth Circuit cases, and an expansion of the approach comes from a Second Circuit case where the court looked beyond formal “title” to determine who was an “owner” under § 117.

123. U.C.C. § 2-401(2) (2001); *United States v. Long*, 706 F.2d 1044, 1050 (9th Cir. 1983) (“UCC § 2-401(2) provides that title passes at the time and place at which the seller completes his performance with reference to physical delivery of the goods.”).

124. *Classic Concepts, Inc. v. Linen Source, Inc.*, Nos. 04-8088 & 04-8457, 2006 U.S. Dist. LEXIS 96767, at *49–50 (C.D. Cal. Apr. 27, 2006).

125. *Old West Realty, Inc. v. Idaho State Tax Comm'n*, 716 P.2d 1318, 1320–21 (Idaho 1986); *see also Skripak v. Comm'r*, 84 T.C. 285, 315 (T.C. 1985) (books); *Middlebrooks v. Comm'r*, 34 T.C.M. (CCH) 1187, at *12–15 (T.C. 1975) (magazines).

126. *Mahru v. Superior Court*, 191 Cal. App. 3d 545, 549 (Ct. App. 1987).

127. *Synergistic Techs., Inc. v. IDB Mobile Commc'ns, Inc.*, 871 F. Supp. 24, 29 (D.D.C. 1994).

128. *Applied Info. Mgmt. v. Icart*, 976 F. Supp. 149, 154 (E.D.N.Y. 1997) (citing RAYMOND NIMMER, *THE LAW OF COMPUTER TECHNOLOGY* § 1.18[1], at 1-103 (1992)).

1. United States v. Wise

Without using the phrase “economic realities,” the Ninth Circuit in *United States v. Wise*¹²⁹ nonetheless gave birth to an approach to determining ownership of a copy that looks to such realities of the transaction. The case involved feature-length motion picture prints that were occasionally provided, under various agreements, to “V.I.P.s” described as “prominent member[s] of the motion picture industry or community” and to television studios.¹³⁰

The court looked beyond the failure of some of the agreements to expressly reserve title in the copyright owner and nonetheless found with respect to one of the transactions that “the general tenor of the entire agreement” was inconsistent with the concept of a sale.¹³¹ The *Wise* court also wrote that since “copyright proprietors frequently transfer rights in their works by complicated agreements which cannot simply be called ‘sales[,]’ in each case, the court must *analyze the arrangement at issue* and decide whether it should be considered a first sale.”¹³²

Analyzing the “arrangement at issue” and looking to the “general tenor of the agreement” led the *Wise* court to find that some transactions involved sales and others did not. One contract the *Wise* court considered allowed the actress Vanessa Redgrave to keep possession of the film print “at all times” for her “personal use and enjoyment,” but purported to prevent her from transferring the print to anyone else.¹³³ The court determined that this contract was a sale despite the purported restrictions on transfer, illustrating that such restrictions on transfer were not dispositive in the *Wise* court’s view.¹³⁴

2. Microsoft Corp. v. DAK Industries

Another example of the Economic Realities approach is *Microsoft Corp. v. DAK Industries*.¹³⁵ Here, the Ninth Circuit addressed whether a transaction was “a lump sum sale of software units to DAK” or merely “a grant of

129. 550 F.2d 1180 (9th Cir. 1977), *cert. denied*, 434 U.S. 929 (1977).

130. *Id.* at 1184.

131. *Id.* at 1191.

132. *Id.* at 1188–89 (emphasis added) (citing *United States v. Bily*, 406 F. Supp. 726, 731 (E.D. Pa. 1975)).

133. *Id.* at 1192.

134. *Id.*

135. 66 F.3d 1091 (9th Cir. 1995); *Datalex Ltd. v. PSA, Inc.*, No. 01-06482, 2003 U.S. Dist. LEXIS 27563, at *6 (C.D. Cal. Jan. 30, 2003) (“‘Economic realities’ of an exchange determine whether a transaction is a sale, a lease, or a license.” (citing *DAK Indus.*, 66 F.3d at 1095)).

permission to use an intellectual property.”¹³⁶ The court wrote that “we must look through [this transaction’s] form to the ‘economic realities of the particular arrangement.’”¹³⁷ Furthermore, “[b]ecause we look to the economic realities of the agreement, the fact that the agreement labels itself a ‘license’ and calls the payments ‘royalties,’ both terms that arguably imply periodic payment for the use rather than sale of technology, does not control our analysis.”¹³⁸

Courts that have addressed the very same question—how to determine if a transaction has resulted in a “sale” despite it bearing a “license” label—have often ignored *DAK*.¹³⁹ These oversights by subsequent courts were probably the result of the different context. In *DAK*, the Ninth Circuit needed to determine whether a transfer of title had occurred, but not for the purpose of applying the first sale doctrine or for evaluating § 117 rights. Instead, the court needed to make this determination in order to make a finding under the bankruptcy code regarding whether the debt related to this transaction arose prepetition or postpetition.¹⁴⁰ This is unfortunate because the court provided extensive guidance for determining when a sale has occurred, looking to salient factors about the transaction.

The court noted the following about the transaction:

- *DAK*’s entire debt to Microsoft arose at once at the outset. It was not incurred in installments over time.
- The amount of *DAK*’s payment was calculated based upon quantity of units, not upon duration of use.
- *DAK* received all of its rights when the term of the agreement commenced, even though payments were made on an installment plan over time, as in the purchase of goods on unsecured credit.
- The timing of the payments was not correlated to *DAK*’s sales of the copies.
- The agreement granted *DAK* a “right to sell” not “permission to use” the copies.

136. *DAK Indus.*, 66 F.3d at 1095.

137. *Id.*

138. *Id.* (quoting *In re Moreggia & Sons, Inc.*, 852 F.2d 1179, 1182 (9th Cir. 1988)).

139. *DAK Industries* was cited in *UMG Recordings, Inc. v. Augusto*, 558 F. Supp. 2d 1055, 1058 (C.D. Cal. 2008), *aff’d*, 628 F.3d 1175 (9th Cir. 2011), and *SoftMan Products Co. v. Adobe Systems Inc.*, 171 F. Supp. 2d 1075 (C.D. Cal. 2001), but not by the numerous other cases within the Ninth Circuit discussed herein.

140. *DAK Indus.*, 66 F.3d at 1095.

- Microsoft did not provide anything at its expense to DAK after the commencement of the agreement.¹⁴¹

The court concluded, “[f]or these reasons, the economic realities of this agreement indicate that it was basically a sale, not a license to use.”¹⁴²

It was also likely important to the court that “[t]he agreement provided that if DAK sold more copies than those paid for by the minimum commitment, DAK would pay Microsoft \$55 for each additional copy sold. However, if DAK sold fewer copies than those paid for by the minimum commitment, Microsoft would not refund any of the commitment.”¹⁴³ Thus, unlike a situation in which Microsoft might license a manufacturer and distributor to make copies and sell them on its behalf, here the entire risk of loss had been transferred to DAK. One can certainly enter into an agreement where one licenses one’s reproduction and distribution rights, such that the same entity can make copies and then sell them on the licensor’s behalf. In such a case, the licensor could rightly argue that the licensee has no ownership interest in the copies and that it reserves the right to repossess any not sold. However, *DAK*’s lesson appears to be that when the licensed distributor pays the copyright owner a fixed amount for a fixed number of copies that cannot be returned if not sold, such a term, perhaps because it transfers the entire risk of loss, transforms a licensing arrangement into a sale of those copies.

3. Krause v. Titleserv, Inc.: *The “Incidents of Ownership”*

In *Krause v. Titleserv, Inc.*,¹⁴⁴ the Second Circuit pushed the Economic Realities approach a step further than any other court had by rejecting the idea that “formal title” was even necessary for purposes of being an “owner” under § 117.¹⁴⁵ The court wrote in an oft-quoted passage,

141. *See id.*

142. *Id.* at 1096.

143. *Id.* at 1093.

144. 402 F.3d 119 (2d Cir. 2005).

145. Many courts considering § 117 discuss the legislative history in which the CONTU report had recommended language for the statute that applied to the “rightful possessor of a copy of a computer program.” Congress changed the term “rightful possessor” to “owner” but did not explain its reason. *Compare* FINAL REPORT OF THE NATIONAL COMMISSION ON NEW TECHNOLOGICAL USES OF COPYRIGHTED WORKS 12 (1978), *with* H.R. REP. NO. 96-1307, pt. 1, at 23 (1980), *reprinted in* 1980 U.S.C.C.A.N. 6460, 6482. Some courts have thought this change signified Congress’ intent to restrict the application of § 117 to those with full “title” to the copy. In contrast, the *Krause* court rightly points out that, [t]he term “rightful possessor” is quite broad. Had that term been used, the authority granted by the statute would benefit a messenger delivering a program, a bailee, or countless others temporarily in lawful possession of

[I]t seems anomalous for a user whose degree of ownership of a copy is so complete that he may lawfully use it and keep it forever, or if so disposed, throw it in the trash, to be nonetheless unauthorized to fix it when it develops a bug, or to make an archival copy as backup security. We conclude for these reasons that formal title in a program copy is not an absolute prerequisite to qualifying for § 117(a)'s affirmative defense. Instead, courts should inquire into whether the party exercises sufficient incidents of ownership over a copy of the program to be sensibly considered the owner of the copy for purposes of § 117(a). The presence or absence of formal title may of course be a factor in this inquiry, but the absence of formal title may be outweighed by evidence that the possessor of the copy enjoys sufficiently broad rights over it to be sensibly considered its owner.¹⁴⁶

In looking at these “incidents of ownership,” the court focused on the following:

Titleserv paid Krause substantial consideration to develop the programs for its sole benefit. Krause customized the software to serve Titleserv's operations. The copies were stored on a server owned by Titleserv. Krause never reserved the right to repossess the copies used by Titleserv and agreed that Titleserv had the right to continue to possess and use the programs forever, regardless whether its relationship with Krause terminated. Titleserv was similarly free to discard or destroy the copies any time it wished.¹⁴⁷

The court concluded, “Titleserv's right, for which it paid substantial sums, to possess and use a copy indefinitely without material restriction, as well as to discard or destroy it at will, gave it sufficient incidents of ownership to make it the owner of the copy for purposes of applying § 117(a).”¹⁴⁸

a copy. Congress easily could have intended to reject so broad a category of beneficiaries without intending a narrow, formalistic definition of ownership dependent on title.

Krause, 402 F.3d at 123. The *Krause* court has the better analysis, as courts have only caused mischief by seeking to divine Congressional intent in a situation where nothing definitive is available.

146. *Krause*, 402 F.3d at 123–24.

147. *Id.* at 124.

148. *Id.* at 124–25.

E. THE PERPETUAL POSSESSION APPROACH

“If a transaction involves a single payment giving the buyer an unlimited period in which it has a right to possession, the transaction is a sale.” Raymond T. Nimmer (1992, 1997).¹⁴⁹

In opinions from courts ostensibly applying either the Economic Realities approach or the Agreement Controls approach, yet another approach arguably drives the results. Two district court decisions, *UMG Recordings, Inc. v. Augusto* and *Vernor v. Autodesk, Inc.* have, intentionally or not, teased this interpretative focus out of the Economic Realities approach to state what this Article calls the Perpetual Possession approach. After explaining this understanding of, or variation on, the Economic Realities approach, Part VI will argue that the Perpetual Possession approach is an excellent development that courts faced with this question should embrace. As the preceding survey of the cases has revealed, courts are in desperate need of a logical approach that respects precedent and congressional choices, and which, in an ideal world, would simply be straightforward for the courts to apply. After this Section provides examples of the Perpetual Possession approach, Part VI argues that the Perpetual Possession approach is such an approach.

1. *UMG Recordings, Inc. v. Augusto*

Troy Augusto made his living selling collectible merchandise on eBay, the internet auction site.¹⁵⁰ UMG’s copyright infringement claim against Augusto was based on his offering “promo CDs” for sale on eBay.¹⁵¹ The twenty-six eBay auctions at issue involved authentic CDs lawfully made and distributed by UMG.¹⁵² Augusto purchased the CDs from retailers in the Los Angeles area or on eBay, but UMG argued that Augusto was not an “owner” of the CDs he bought, because the promo CDs contained one of the two following legends:

149. *DSC Commc’ns Corp. v. Pulse Commc’ns, Inc.*, 170 F.3d 1354, 1362 (Fed. Cir. 1999) (citing RAYMOND T. NIMMER, *THE LAW OF COMPUTER TECHNOLOGY* ¶ 1.24[1], at 1-143–44 (3d ed. 1997)); *Applied Info. Mgmt. v. Icart*, 976 F. Supp. 149, 154 (E.D.N.Y. 1997) (quoting RAYMOND NIMMER, *THE LAW OF COMPUTER TECHNOLOGY* § 1.18[1], at 1-103 (1992)).

150. Appellee’s Brief at 1, *UMG Recordings, Inc. v. Augusto*, 628 F.3d 1175 (9th Cir. 2011) (No. 08-55998), available at [http://www.eff.org/files/filenode/umg_v_augusto/Appellee's Brief.pdf](http://www.eff.org/files/filenode/umg_v_augusto/Appellee's%20Brief.pdf).

151. *Id.* at 2.

152. *UMG Recordings, Inc. v. Augusto*, 558 F. Supp. 2d 1055, 1058 (C.D. Cal. 2008), *aff’d*, 628 F.3d 1175 (9th Cir. 2011).

Promotional Use Only—Not for Sale.

or

This CD is the property of the record company and is licensed for the intended recipient for promotional use only. Acceptance of this CD shall constitute an agreement to comply with the terms of the license. Resale or transfer of possession is not allowed and may be punishable under federal and state laws.¹⁵³

Promo CDs are mailed unsolicited to recipients, typically “music industry insiders who are in a position to provide publicity and exposure” for the CD.¹⁵⁴ As of the time of this litigation, UMG had never made any efforts to retrieve promo CDs from recipients, and nothing on the packaging of promo CDs indicated that they must ever be returned to UMG.¹⁵⁵ UMG did not keep permanent records of who received any particular promo CD, and UMG did not take any steps to mark the promo CDs for later tracking.¹⁵⁶ The district court considered all of these “economic realities” and concluded that “title to the CDs is transferred to the insiders.”¹⁵⁷ An authorized disposition of the CDs having occurred, the district court held that UMG’s publication right was exhausted, freeing both the recipients of the CDs as well as Augusto to sell them. UMG has appealed to the Ninth Circuit.¹⁵⁸

The *Augusto* district court wrote, “[i]n determining whether a transaction is a sale or a license, courts must analyze the ‘economic realities’ of the transaction. [T]he fact that the agreement labels itself a ‘license’ . . . does not control our analysis.”¹⁵⁹

The district court then singled out one factor for special consideration:

The right to perpetual possession is a critical incident of ownership. Accordingly, the distributor of a copyrighted product’s intent to regain possession is strong evidence that the product was licensed,

153. Appellee’s Brief, *supra* note 150, at 13–14.

154. *Augusto*, 558 F. Supp. 2d at 1058.

155. *Id.* at 1061. There is evidence that this norm within the promo CD world is changing to include individually watermarked CDs, streaming URLs, listening parties at corporate headquarters, and extensive license language on promotional CD packaging. Erik Davis, *My Data Crime: The Ticking Time Bomb of the Watermarked Advance CD*, TECHGNOSIS (Sept. 8, 2007), <http://www.techgnosis.com/chunks.php?sec=journal&cat=noting&file=chunkfrom-2007-09-08-2341-0.txt>.

156. *Augusto*, 558 F. Supp. 2d at 1061.

157. *Id.* at 1062.

158. *Id.* at 1058, 1062.

159. *Id.* at 1060 (citing *Microsoft Corp. v. DAK Indus.*, 66 F.3d 1091, 1095 (9th Cir. 1995)).

not sold, to the recipient. The absence of this intent is strong evidence that the product was sold.¹⁶⁰

The *Augusto* district court also found that “perpetual possession” was critical to the Ninth Circuit’s decision in *Wise*, noting that the *Wise* court did not find a sale when the agreements required a film print’s return at the end of a specified period, but did find a sale when the recipient was entitled to keep possession of the film print at all times.¹⁶¹

In applying this key factor of a right to perpetual possession to these facts, the district court found that

[h]ere, UMG gives the Promo CDs to music industry insiders, never to be returned. The recipients are free to keep the Promo CDs forever. Nothing on the packaging of the Promo CDs or in the licensing label requires that the recipient return the Promo CDs to UMG. There are no consequences for the recipient should she lose or destroy the Promo CDs—which UMG allegedly considers its property. UMG does not request that any recipients return the Promo CDs and does not otherwise make any affirmative effort to recover possession of the Promo CDs. Further, it appears that UMG could not take these actions; UMG does not keep permanent records identifying who received which Promo CDs. Accordingly, the music industry insiders’ ability to indefinitely possess the Promo CDs is a strong incident of ownership through a gift or sale.¹⁶²

Finally, the *Augusto* district court noted that the only apparent benefit to UMG to claim that it retains title to the copies is as part of an effort to restrain trade by preventing the transfer of its music. However, the court explains that, “[t]his purpose was rejected 100 years ago by the Supreme Court,” citing the Court’s holding in *Bobbs-Merrill*.¹⁶³

Thus, by following Ninth Circuit precedent, the *Augusto* district court highlighted two factors at the heart of the Economic Realities approach: possession and payment. If the right to possession is permanent and the payment¹⁶⁴ is not correlated to the period of time possessed, then there

160. *Id.* (footnote omitted) (citing *Krause v. Titleserv, Inc.*, 402 F.3d 119, 123 (2d Cir. 2005) (describing a person’s degree of ownership of a copy as complete when he may lawfully use it and keep it forever, or if so disposed, throw it in the trash)).

161. *Id.* at 1060–61.

162. *Id.* (footnote omitted) (citations omitted). The *Augusto* court also relied on *DAK Industries* to identify another important factor: the absence of a recurring benefit to the transferor. *Id.*

163. *Id.* (citations omitted) (citing *Bobbs-Merrill Co. v. Straus*, 210 U.S. 339 (1908) (rejecting a book publisher’s attempt to restrict resale of a book through a label that prohibited sales for less than one dollar)).

164. The *DAK Industries* court also already noted that nothing should prevent the court from finding that title to the copy has passed to the transferee where multiple payments are

are strong indicia that title to the copy has passed to the transferee. Conversely, if the possession is expected to be temporary or the payments are correlated to the period of time possessed, then there are strong indicia that title to the copy remains in the transferor.

The importance of these critical indicia of ownership lurked in the background of the Ninth Circuit's key precedents on ownership of copies for decades,¹⁶⁵ but the *Augusto* district court's opinion placed them in sharper relief and thereby illustrated a way forward for future courts.¹⁶⁶

The *Augusto* district court also did something that is too infrequent in these cases: it considered the relevance of Supreme Court first sale precedents. The courts that have looked to restrictive terms in agreements to reach a result on the question of copy ownership have often made no effort to reconcile enforcement of such provisions with the Supreme Court's refusal to enforce such a restrictive provision in *Bobbs-Merrill*. Even if there are cases where a relevant distinction can be made, courts should explain such a distinction explicitly.

2. *Vernor v. Autodesk, Inc.* (district court)

The district court in *Vernor* had two opportunities to address whether copies of Autodesk's AutoCAD software were sold, first on Autodesk's motion to dismiss and subsequently on cross-motions for summary judgment.¹⁶⁷ *Vernor* had acquired the copies at an office equipment sale conducted by an architecture firm, CTA.¹⁶⁸ In both opinions, the district court found sales that provided *Vernor* with a first sale defense.

not correlated with period of possession, but instead are more akin to a single fee paid on an installment plan. *Augusto* itself also illustrates that there need not be payment at all, as in the case of a gift. Instead, the crucial factor is that the copyright holder has voluntarily and permanently parted with its possession of the copy. *Cf.* 17 U.S.C. § 115(c)(2) (2006) (defining "distributed" in the context of compulsory licenses). An exception to "voluntarily" can even be found in the case of a forced judicial sale.

165. And in the more recent Second Circuit opinion in *Krause*, 402 F.3d 119.

166. There are, however, certain phrasings in the *Augusto* opinion that are not ideal, particularly the use of the word "license" which has invited so many past courts to conflate the intangible copyright with the tangible copy. This court avoided that pitfall, but preferably would have written more precisely about title to the copies. Subsequent courts have already begun to apply the *Augusto* district court's formulation. *See* F.B.T. Prods., LLC v. Aftermath Records, No. 07-3314, 2009 WL 137021, at *4–5 (C.D. Cal. Jan. 20, 2009) *rev'd*, 621 F.3d 958 (9th Cir. 2010).

167. *Vernor v. Autodesk, Inc.* (*Vernor I*), 555 F. Supp. 2d 1164, 1170 (W.D. Wash. 2008) (on motion to dismiss); *Vernor v. Autodesk, Inc.* (*Vernor II*), No. C07-1189RAJ, 2009 WL 3187613, at *11–14 (W.D. Wash. Sept. 30, 2009), *rev'd*, *Vernor v. Autodesk, Inc.* (*Vernor III*), 621 F.3d 1102 (9th Cir. 2010) (on cross-motions for summary judgment).

168. *Vernor II*, 2009 WL 3187613, at *11–14.

The *Vernor* district court reviewed the controlling precedent of *Wise* and noted the admonition found there that the court “must analyze the ‘arrangement at issue and decide whether it should be considered a first sale.’”¹⁶⁹ After a careful review of the facts, rationale, and holdings in *Wise*, the district court wrote:

In comparing the transactions found to be sales in *Wise* with those that were not, the critical factor is whether the transferee kept the copy acquired from the copyright holder. When the film studios required that prints be returned; the court found no sale. When the studios did not require the transferee to return the prints, the court found a sale. Even a complete prohibition on further transfer of the print (as in the Redgrave Contract), or a requirement that the print be salvaged or destroyed, was insufficient to negate a sale where the transferee was not required to return the print.¹⁷⁰

Applying this guidance from *Wise*, the *Vernor* district court wrote:

Taking direction solely from *Wise*, the court concludes that the transfer of AutoCAD packages from Autodesk to CTA was a sale. Like the Redgrave Contract, the Settlement Agreement and License allowed CTA to retain possession of the software copies in exchange for a single up-front payment. Like the Redgrave Contract, the Settlement Agreement and License imposed onerous restrictions on transfer of the AutoCAD copies. Similar to the salvage transactions in *Wise*, the License required CTA to destroy the software in the event that it purchased a software upgrade. Under *Wise*, however, this is a “sale with restrictions on use,” and is a sufficient basis to invoke the first sale doctrine.¹⁷¹

The *Vernor* district court also looked beyond the first sale context to those cases that have addressed the copy ownership issue in relation to § 117, and it considered *MAI*, *Triad*, and *Wall Data*. The *Vernor* district court recognized that these courts were, without citing *Wise*, looking to restrictive terms in the agreements accompanying a transferred copy to determine whether title to the copy passed. The district court interpreted these cases to reflect an approach that would counsel that the more restrictions found or the more draconian the restrictions are, the less likely it is that title to the copy rests with the transferee. And the *Vernor* district court acknowledged that, in a comparison of restrictive terms, the Autodesk agreement would fall on the very restrictive end of the comparison as Autodesk’s software is

169. *Vernor I*, 555 F. Supp. 2d at 1169 (citing *United States v. Wise*, 550 F.2d 1180, 1189 (9th Cir. 1977)).

170. *Id.* at 1170 (footnote omitted).

171. *Id.* at 1170–71 (footnote omitted) (citation omitted).

accompanied by agreements that purport to prohibit the transfer of the software without Autodesk's written permission, among other restrictions.

This placed the district court in the uncomfortable position of having two approaches to addressing the same question, both articulated by the Ninth Circuit at different times, and which, in the present case, led to opposite results. The court tried valiantly to find a means to escape this irreconcilable conflict by considering whether any difference between § 109 and its predecessor statute, § 27 considered in *Wise*, could be used to distinguish the cases. The court also looked to the fact that *Wise* was a criminal case, the difference between motion picture film prints and software, and the differences between § 109 and § 117, to see if any of these could provide a basis for resolving the conflict and could place this case more in line with one of these Ninth Circuit precedents.¹⁷² However, the district court persuasively argued that none of these provided an adequate basis for distinguishing the cases and thus the court was guided by the principle that a subsequent panel of the Ninth Circuit cannot overrule an earlier panel absent intervening Supreme Court precedent or en banc authority. With no such intervening authority, the court was bound to follow *Wise*.¹⁷³

The district court's opinions in *Vernor* are carefully reasoned, thorough, and correct in their results. The court achieved this in part through a close reading of *Wise*, which uncovered the same critical factor that the *Augusto* court had emphasized: the right to perpetual possession.¹⁷⁴

IV. THE ADOBE OPINIONS AND INTERPRETIVE DIFFICULTY

Different interpretive approaches can produce opposite outcomes. The three *Adobe* opinions provide a unique case study illustrating this point: one court used elements of the Reservation of Title and Agreement Controls approaches and held for the software distributor opposing a first sale defense; the other used the Economic Realities approach and held for the litigant asserting a first sale defense. Both courts were in the Ninth Circuit. No intervening mandatory precedent occurred between the three cases, and

172. *Id.* at 1173–74; *Vernor II*, 2009 WL 3187613, at *11–14.

173. *Vernor I*, 555 F. Supp. 2d at 1172; *Vernor II*, 2009 WL 3187613, at *11–14.

174. *Vernor I*, 555 F. Supp. 2d at 1170 (“the critical factor is whether the transferee kept the copy acquired from the copyright holder”); *Vernor II*, 2009 WL 3187613, at *11–14 (“In each instance in which the transferee could, at his election, retain possession of the transferred copy indefinitely, and the copyright holder had no right to regain possession, the court found an ownership transfer.”).

they involved the same plaintiff attempting to enforce very similar licenses, but nonetheless the two courts reached opposite results. Both courts, even when clearly cognizant of the other's differing point of view, were unable to interpret the same precedent consistently with one another. When the same plaintiff, seeking to enforce essentially the same right over the same three year time period gets different results depending on whether it files its case in Northern or Central California, there is a pressing need for clarification.

Two of these cases were brought in the Northern District of California, both before the Honorable James Ware.¹⁷⁵ In both cases, Judge Ware used elements of the Reservation of Title and Agreement Controls approaches. He sought to “give effect to the mutual intent of the parties” by looking “exclusively”¹⁷⁶ at the language of the contract, assuming the language is “clear and explicit.”¹⁷⁷ He also reasoned that in the event of ambiguity, the court could look to custom and usage of words in the trade to interpret the contract.¹⁷⁸

In *One Stop Micro*, the contract, known as the “On Campus Reseller Agreement” (“OCRA”), contained purported restrictions on resellers. For example, the OCRA stated that resellers could not distribute outside their own country, and they were required to take specified steps to ensure that educational purchasers were bona fide members of an educational institution.¹⁷⁹ The court wrote, “[t]hese numerous restrictions imposed by Adobe indicate a license rather than a sale because they undeniably interfere with the reseller’s ability to further distribute the software.”¹⁸⁰

However, finding some sales terminology in the agreement, the court felt the agreement was ambiguous and consulted extrinsic evidence in the form of declarations from individuals familiar with the software industry who largely, and uncritically, reported that industry practice is for software to be “licensed” and not “sold.”¹⁸¹ This allowed the court to conclude

175. *Adobe Sys., Inc. v. Stargate Software Inc.*, 216 F. Supp. 2d 1051 (N.D. Cal. 2002); *Adobe Sys., Inc. v. One Stop Micro, Inc.*, 84 F. Supp. 2d 1086 (N.D. Cal. 2000).

176. *One Stop Micro*, 84 F. Supp. 2d at 1090.

177. *Id.*; see also *Stargate Software*, 216 F. Supp. 2d at 1055–56 (“The determination of ownership in turn is based primarily on an examination of the OCRA, the agreement between Adobe and its distributors. The Court looks to the language, content, and intent of the OCRA, in determining whether its terms affect a sale or license of the software.”).

178. *One Stop Micro*, 84 F. Supp. 2d at 1090 (citing CAL. CIV. PROC. CODE § 1856(c)).

179. *Id.* at 1090 n.2.

180. *Id.* at 1091. Looking to restrictions on rights of distribution in order to determine copy ownership is a specific error that will be addressed *infra* Part V.

181. *Id.* at 1091–92. Most of these declarations conflate the intangible copyright with the tangible copy and so are of no benefit. See *supra* Section II.C; *infra* Part V.

that based upon the undisputed evidence submitted by Adobe regarding the intent of the parties in entering into the agreement, trade usage, the unique nature of distributing software, as well as the express restrictive language of the contract, the OCRA is a licensing agreement. Thus, contrary to One Stop's assertions, the OCRA does not represent a first sale between the reseller and Adobe.¹⁸²

In *Stargate Software*, Judge Ware drew repeated comparisons to his earlier decision in *One Stop Micro* and found that, “[s]imilar to the OCRA at issue in *One Stop*, the OCRA in this case contains multiple restrictions that limit the reseller’s ability to distribute Adobe’s software.”¹⁸³ However, in the intervening year, the Honorable Dean D. Pregerson of the Central District of California had decided yet another Adobe case, *SoftMan Products Co.*, and had found that the transaction there had resulted in a sale of the copy giving rise to a first sale defense.¹⁸⁴

*SoftMan Products Co.*¹⁸⁵ applied the Economic Realities approach to find a sale. The court wrote, “[i]t is well-settled that in determining whether a transaction is a sale, a lease, or a license, courts look to the economic realities of the exchange.”¹⁸⁶ The court cited *DAK* and indicated that the mere “license” label did not control its analysis. The court, citing § 202, also recognized that the question before it was not one of ownership of the copyright, but of “ownership of individual pieces of Adobe software.”¹⁸⁷

When examining the transaction’s economic realities, the court highlighted the following:

[T]he purchaser commonly obtains a single copy of the software, with documentation, for a single price, which the purchaser pays at the time of the transaction, and which constitutes the entire payment for the “license.” The license runs for an indefinite term without provisions for renewal. In light of these indicia, many courts and commentators conclude that a “shrinkwrap license” transaction is a sale of goods rather than a license.¹⁸⁸

182. *One Stop Micro*, 84 F. Supp. 2d at 1092.

183. *Adobe Sys., Inc. v. Stargate Software Inc.*, 216 F. Supp. 2d 1051, 1057 (N.D. Cal. 2002).

184. *SoftMan Prods. Co. v. Adobe Sys. Inc.*, 171 F. Supp. 2d 1075 (C.D. Cal. 2001).

185. *Id.* at 1075.

186. *Id.* at 1084.

187. *Id.*

188. *Id.* at 1085.

The court also looked to whether risk of loss transferred as an indicator that title had passed.¹⁸⁹

Ultimately the court distinguished *Harmony* and explicitly declined to adopt the analysis of *One Stop Micro*, finding that “a single payment for a perpetual transfer of possession is, in reality, a sale of personal property and therefore transfers ownership of that property, the copy of the software.”¹⁹⁰

In an effort to distinguish *Stargate Software* from *SoftMan Products*, Judge Ware first noted the unique ease with which software can be copied, making it especially vulnerable to unauthorized copying. He then wrote,¹⁹¹

Lastly, as a matter of general principle . . . the Parties should be free to negotiate and/or set a price for the product being exchanged, as well as set the terms by which the product is exchanged. Fundamental to any free society is the liberty of its members to formulate contracts in accordance with the terms that they agree and consent to mutually execute. “The right

189. *Id.* (“The distributors pay full value for the merchandise and accept the risk that the software may be damaged or lost. The distributors also accept the risk that they will be unable to resell the product.”); *see also supra* Section III.D.

190. *SoftMan Prods.*, 171 F. Supp. 2d at 1086. Note the use of the word “perpetual” here, which is an intimation of the Perpetual Possession approach. The court also rejected the holdings of *Harmony* and *One Stop Micro* with respect to the applicability of agreements never assented to, because here, SoftMan, as a reseller, never installed the software and thus never assented to any Adobe EULA or other licensing agreement. *Id.* at 1087–88. The *SoftMan Products* court noted:

In *One Stop*, the court stated that although One Stop was not a signatory to an Adobe licensing agreement, it was nevertheless subject to the restrictions of those agreements. 84 F. Supp. 2d at 1092. The court found that by obtaining Adobe software from a party to an Adobe licensing agreement, One Stop was bound by any restrictions imposed by that agreement. *Id.* at 1093. In *Harmony*, the court found that “to the extent that defendants bought their Microsoft Products from authorized Microsoft licensees, they were subject to the same licensing restrictions under which those licensees operated.” [Microsoft Corp. v. Harmony Computers & Elecs., Inc., 846 F. Supp. 208, 213 (E.D.N.Y. 1994)] The Court declines to adopt the analysis of these cases.

Id.

191. The paragraphs prior to the one quoted also attempt to distinguish *SoftMan Products*, but fail. The court first notes that *SoftMan Products* involved unbundling of a collection of software in apparent violation of a EULA, whereas *Stargate Software* involved the simple re-distribution of Adobe software, as is. *Adobe Sys., Inc. v. Stargate Software Inc.*, 216 F. Supp. 2d 1051, 1058 (N.D. Cal. 2002). While this is a cognizable distinction, the court fails to explain how that distinction would entail that SoftMan Products’ transaction resulted in a sale, while Stargate Software’s was not a sale. The court then notes that when “ ‘payment’ is made for a particular copy of software, the payment is being made for the value of the objective code that is burned on the CD-ROM,” *id.* at 1059, but fails to explain how this differs from the situation in *SoftMan Products*, which also involved software burned on CD-ROMs whose true economic value was derived from that which was embodied within it. *See id.* at 1058–59.

to contract freely with the expectation that the contract shall endure according to its terms is as fundamental to the society as the right to write and to speak without restraint.” While exceptions are made in the case of unfair or exploitive contracts, or where an inequitable end results as a result of the agreement, commercial parties are generally free to contract as they desire.¹⁹²

Arguably, this passage exemplifies the Reservation of Title approach, because courts applying it appear motivated by a combination of software exceptionalism and a fervent belief in a nearly unfettered freedom to contract.¹⁹³ Adobe’s efforts to control the sale of its software to educational customers was an effort to engage in price discrimination, something that the *Stargate Software* court saw no reason to obstruct. However, in interpreting the first sale doctrine, the Supreme Court of the United States has written, “whether or not we think it would be wise policy to provide statutory protection for such price discrimination is not a matter that is relevant to our duty to interpret the text of the Copyright Act.”¹⁹⁴ Unfortunately, the Copyright Act does not explicitly resolve how courts should determine who is an owner of a copy entitled to § 109 and § 117 rights, and these three *Adobe* cases vividly illustrate that the approach taken to the resolution of that question can be determinative of the result, making it all the more important that courts employ a consistent and logical approach that respects precedent and Congressional choices. As is discussed in Part III, *supra*, the Reservation of Title approach over-relies on the terms of the agreement, without giving

192. *Id.* at 1059 (citations omitted). None of the court’s citations were to controlling precedent.

193. It is, however, disconcerting that this policy argument, based upon differences between software and other copyrighted works, should come from the judiciary rather than the legislature. If Congress wishes to recognize a special protection for software distributors it has proven able to do so. *See* Computer Software Rental Amendments Act of 1990, Pub. L. No. 101-650, 104 Stat. 5134 (codified at 17 U.S.C. § 109(b) (2006)); *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 430–31 (1984). In *Sony Corp. of America*, the Court noted:

From its beginning, the law of copyright has developed in response to significant changes in technology. Indeed, it was the invention of a new form of copying equipment—the printing press—that gave rise to the original need for copyright protection. Repeatedly, as new developments have occurred in this country, it has been the Congress that has fashioned the new rules that new technology made necessary.

464 U.S. at 430–31 (footnotes omitted). In the absence of such a congressionally authorized distinction, the courts should interpret the relevant copyright statutes as written.

194. *Quality King Distribs. v. L’Anza Research Int’l, Inc.*, 523 U.S. 135, 153 (1998).

adequate consideration to controlling Supreme Court precedents that require a focus on other factors.¹⁹⁵

It is useful to reflect on the features we should expect from an interpretative approach to the question of copy ownership, as it will help to confirm the wisdom of the Perpetual Possession approach adopted in *Wise* and followed in *Augusto* and *Vernor*. Thus, the next Part argues for the essential and desirable features of such an approach.

V. CHARACTERISTICS OF A CORRECT APPROACH TO DETERMINING COPY OWNERSHIP

Courts are in desperate need of a logical approach to the question of copy ownership that is logically coherent, respects precedent, and respects congressional choices. The Perpetual Possession approach is such an approach.

A. MUST BE LOGICALLY CORRECT

Two logical errors run rampant in the “license versus sale” jurisprudence, but both arise from the same source: a failure to distinguish between the intangible copyright and the tangible copy. First, the word “license” is used in at least two distinct ways, which makes courts susceptible to inadvertent equivocation. Second, mere status as the licensee of a right of copyright is erroneously presumed to be incompatible with also being the owner of a copy. Any approach to determining ownership of a copy must avoid these logical errors. A third error that is at least misguided, if not illogical, is that courts often look to factors that are wholly independent of the question of copy ownership. All three errors must be avoided in a proper approach to determining ownership of a copy of a copyrighted work.

195. I agree with William Patry’s comments on *One Stop Micro*. Patry writes, [T]he court misinterpreted contractual restrictions as indicating the existence of a license agreement for section 109 purposes. Restrictions on the terms of a sale do not by themselves mean a sale has not occurred. If such restrictions are violated, the contract may have been breached, but a sale took place nevertheless and thus no infringement action is available. For example, nothing in plaintiff’s agreements gave it a right to call back the copies in the event of a breach. Adobe is, therefore, wrongly decided on the copyright claim. Far too often, courts merely accept plaintiff’s description of the transaction as a license.

PATRY, *supra* note 40, § 13:25 (footnotes omitted).

1. *Must Not Equivocate with Respect to the Words "License" or "Software"*

When "license" is used as a noun in the copyright context, it means something like, "a grant by the holder of a copyright to another of any of the rights embodied in the copyright short of an assignment of all rights" as in, "The agreement contained a license to reproduce 20 copies of the photograph."

When "license" is used as a verb it typically means "to give permission or consent" as in, "The author licensed her distribution right to the nation's largest distributor."

These uses of the word relate *only* to the intangible copyright.

The word "license" is also, unfortunately, used in conjunction with tangible things. First, as a noun it is often used synonymously with the terms "agreement" or "contract" when that underlying agreement contains grants of copyright permissions, as in, "Did she sign the license?" This usage does not seem to lead to much confusion and it will not be addressed further here.

There is another use of the word "license," particularly in the software context, that is most confusing. Software distributors often say, "We only license our software. We do not sell it." This is a difficult sentence to parse because of the layers of ambiguity involved. However, by reading the relevant case law,¹⁹⁶ one comes to understand that the intended definition is not just "to give permission or consent" with respect to some right of copyright. It is instead used in a way that means something more like the following:

to transfer to another permanent possession of a tangible object in which a copyrighted work is embodied, without transferring title to the tangible object.¹⁹⁷

It would be useful to have a different term to indicate this unique use of "license." Something like "no title to the copy" license would perhaps convey the intended meaning, but would be exceedingly cumbersome. For purposes of clarity in this Part, uses of "license" in the sense of "no title to the copy" license will be italicized (*license*).

Usage of the word *license* has caused rampant confusion. Before considering some examples of this confusion, it is worthwhile to provide some historical context on the development of this usage of the term *license*.

196. See *supra* Part III.

197. This is my own attempt at a definition of "license" as the software distributors use it.

The Third Circuit explained, in an opinion from 1991, *Step-Saver Data System v. Wyse Technology*, that:

When these form licenses were first developed for software, it was, in large part, to avoid the federal copyright law first sale doctrine. . . . [The court describes software rental companies.] The first sale doctrine, though, stood as a substantial barrier to successful suit against these software rental companies, even under a theory of contributory infringement. By characterizing the original transaction between the software producer and the software rental company as a *license*, rather than a sale, and by making the *license* personal and non-transferable, software producers hoped to avoid the reach of the first sale doctrine and to establish a basis in state contract law for suing the software rental companies directly. Questions remained, however, as to whether the use of state contract law to avoid the first sale doctrine would be preempted either by the federal copyright statute (statutory preemption) or by the exclusive constitutional grant of authority over copyright issues to the federal government (constitutional preemption). . . . Congress recognized the problem, and, in 1990, amended the first sale doctrine as it applies to computer programs and phonorecords. . . . This amendment renders the need to characterize the original transaction as a *license* largely anachronistic.¹⁹⁸

But the usage, even if anachronistic, has persisted, in part because software distributors wanted more than to defeat the use of the first sale doctrine by software rental companies. Even after Congress responded to that concern, software distributors were unwilling to give up the *licensing* fiction because it provides a means to other desirable ends such as price discrimination, controlling ancillary markets, and preventing competition in related goods.¹⁹⁹

The merits of permitting copyright owners these additional benefits are not the focus of this Article. The concern is with how the ambiguous use of the word “license” has created a land mine for courts that end up speaking imprecisely or, in the worst case scenarios, reaching erroneous conclusions.

The *Microsoft Corp. v. Software Wholesale Club, Inc.* opinion provides one example of the imprecise usage of “license.” The court wrote:

198. 939 F.2d 91, 96 n.7 (3d Cir. 1991) (emphasis added) (citations omitted).

199. As an example of price discrimination, most of the Novell and Adobe cases are about educational versions of software offered at a price discounted from the regular retail price. See, e.g., *Novell, Inc. v. Unicom Sales, Inc.*, No. C-03-2785 MMC, 2004 WL 1839117, at *10 (N.D. Cal. Aug. 17, 2004); *Adobe Sys. Inc. v. One Stop Micro, Inc.*, 84 F. Supp. 2d 1086, 1088 (N.D. Cal. 2000). For anti-competitive examples, Blizzard’s terms are designed to prevent third parties from offering competing match-making services for online play and attempt to prevent third-parties from developing independent software that interoperates with Blizzard’s software. See *Davidson & Assocs. v. Jung*, 422 F.3d 630 (8th Cir. 2005). Thus, the terms aim to give Blizzard sole control of those lucrative markets.

However, a party that *licenses* its products rather than selling them may avoid the application of the first-sale doctrine. *See, e.g., Harmony Computers & Elecs.*, (the fact that Microsoft *licenses* rather than sells its products likely precludes application of the first-sale doctrine); *Novell, Inc.*, 2000 U.S. Dist. LEXIS 9975, at *7–18 (the first-sale defense applied, but only because Novell sold, rather than *licensed*, its software product).²⁰⁰

What does the phrase “licenses its products” mean here? Both “licenses” and “products” could have two meanings.²⁰¹ If “licenses” refers to a granting of permission with respect to some right of copyright, then it has fallen into the error of ignoring 17 U.S.C. § 202 by failing to recognize the possibility of ownership of a copy independent from ownership of the copyright. But, if “licenses” instead means *licenses*—that is, a permanent transfer of possession without a transfer of title to the copy—then one has presumed the answer to the question being asked. In determining whether someone is an owner of a copy, it is not much use to say that those who are not owners of a copy do not have the rights of owners of a copy. This is a tautology. What was wanted was a feature of the transaction that would distinguish the owners from the non-owners, without reference to the label applied by the copyright holder.²⁰²

The *MAI* footnote is a paradigmatic example of the logical fallacy of equivocation: “Since MAI licensed its software, the Peak customers do not qualify as ‘owners’ of the software and are not eligible for protection under § 117.”²⁰³

The word “software” is used in two different senses within the footnote, making its conclusion a non sequitur. In the second half of the footnote, when the court speaks of owners of “the software” and the applicability of

200. *Microsoft Corp. v. Software Wholesale Club, Inc.*, 129 F. Supp. 2d 995, 1007 (S.D. Tex. 2000) (emphasis added) (citation omitted).

201. Particularly in the software context, we sometimes think of the intangible copyrighted work as the “product.” At other times we might refer to the tangible copy on which the work is embodied as the “product.” The same ambiguity arises with the word “software” making a “license to use the software” equally, if not more, ambiguous.

202. It might be a logically coherent position to accept that the copyright holder simply can declare that title to the copy is not passing and that no other feature of the transaction matters. However, it would remain unclear where the authority to make such declarations would come from. Even where servitudes are permitted to run with property, public policies impose limits that constrain the would-be creator of such servitudes. *See Chafee, supra* note 36, at 986–87. It would also be a view with surprising consequences, as there would be some non-owners whose bundle of substantive rights would be identical to those of an owner. Only the handful of courts that embraced the Reservation of Title approach would possibly accept such an argument. However, not everything that is logically coherent is the law or good policy.

203. *MAI Sys. Corp. v. Peak Computer*, 991 F.2d 511, 518 n.5 (9th Cir. 1993).

17 U.S.C. § 117, the court can only mean “owners of the copy of the software,” for § 117 applies only to “the owner of a copy of a computer program.”²⁰⁴ However, in the first half of the footnote, when the court states that MAI licensed “its software,” the court can only mean that MAI licensed “its copyright in its software.” This stems from the fact that a “license” is simply a permission granted with respect to an exclusive right, and it is therefore nonsensical to speak of “licensing” a tangible object.²⁰⁵ Thus, the footnote should have read, “Since MAI licensed [its copyright in] its software, the Peak customers do not qualify as ‘owners’ of [the copy of] the software and are not eligible for protection under § 117.” But this reading makes explicit that the *MAI* court has ignored the warning of 17 U.S.C. § 202, that the exclusive rights of copyright are an entirely separate matter from property rights in any material object.

Thus, there are at least two errors of equivocation that courts must guard against when making a determination of copy ownership. First, the sense in which the word “license” is being used should always be made clear because software distributors use that word to describe a concept far removed from the licensing of copyrights. They instead use it to refer to a wholly unique—and I will argue, nonsensical—form of permanently transferring a tangible good while retaining title to it. Second, the word “software” and similar words such as “product” must be used in ways that make plain whether one is referring to the intangible copyrighted work or a particular tangible embodiment of that copyrighted work in the form of a copy. Virtually all of the courts utilizing a “magic words” or Reservation of Title approach make one of these errors.

One key to a proper resolution of the “license versus sale” question is a recognition that the two concepts are not mutually exclusive. Consequently, the entire framing of the question as one of “license versus sale” presents a false dichotomy that should be avoided. In a sale, a copy is transferred, the recipient owns the copy, and the recipient is therefore entitled to the benefits of § 109 and § 117. In a license, one must first determine what sense of “license” is intended. If a copyright license is intended, that is, merely a granting of a permission to reproduce, to create derivatives, to distribute, or to publicly perform or display a copyrighted work, then a license of any of these rights can co-occur with a sale of an underlying copy just as easily as with a lease, lending, or rental of the copy. Rather than polar opposites, the concepts of “license” and “sale” are completely independent of one another.

204. 17 U.S.C. § 117 (2006).

205. I further develop this argument in Section V.A.2, *infra*.

If instead *license* is intended, that is, a permanent transfer of a copy of a copyrighted work in which the transferor retains ownership of the copy, then the concepts are mutually exclusive. However, the problem remains that the entire notion of a *license* is nonsensical. In no other context is anyone even tempted to entertain the view that a permanent transfer of a tangible thing could occur without ownership of the thing passing as well.²⁰⁶

2. *Must Recognize the Possibility of Ownership of a Copy Independent from Ownership of the Copyright*

17 U.S.C. § 202 provides that

Ownership of a copyright, or of any of the exclusive rights under a copyright, is distinct from ownership of any material object in which the work is embodied. Transfer of ownership of any material object, including the copy or phonorecord in which the work is first fixed, does not of itself convey any rights in the copyrighted work embodied in the object; nor, in the absence of an agreement, does transfer of ownership of a copyright or of any exclusive rights under a copyright convey property rights in any material object.²⁰⁷

206. It is essential that software distributors cling to this fiction of retaining title to the copy even after permanently abandoning the copy to the possession of a customer, because the language of § 109 and § 117 speak in terms of an owner of a copy. If software distributors were to concede that their customers owned their copies but simply owned them subject to the restrictions enumerated in the accompanying license agreement, then the entire aim of evading the exceptions imposed by Congress would be thwarted. Understood in this light, this is not a debate about the enforceability of post-sale restrictions, because the claim software distributors are forced to make is that no *sale* has occurred at all. The difficulty for software distributors is that they engage in permanent transfers of copies and then seek in vain for a notion of a permanent transfer of possession that is neither a sale nor a gift and implies no transfer of title. I am unaware of such a notion at common law, in the laws governing chattel, or within common sense. Rather, a right to perpetual possession is the hallmark of our notion of an owner. For “owner” to be a distinct concept from that of a bailee or lessee it must contain this right, because a lessee has a temporary, not a perpetual, right of possession, and a bailee has merely lawful possession (and a duty of care), with no right to maintain possession against the bailor’s demand. *See supra* note 34.

207. 17 U.S.C. § 202 (2006). The predecessor statute, 17 U.S.C. § 27 (1958), read similarly:

The copyright is distinct from the property in the material object copyrighted, and the sale or conveyance, by gift or otherwise, of the material object shall not of itself constitute a transfer of the copyright, nor shall the assignment of the copyright constitute a transfer of the title to the material object; but nothing in this title shall be deemed to forbid, prevent or restrict the transfer of any copy of a copyrighted work the possession of which has been lawfully obtained.

Id. Section 41 of the 1909 Act was identical, with the substitution of “this Act” for “this title.” *See* Copyright Act of 1909, Pub. L. No. 60-349, § 41, 35 Stat. 1075, 1084 (1909).

That ownership of the tangible thing is a separate issue from the rights under copyright that one might have with respect to that tangible thing is also a centuries-old principle in U.S. law. The Supreme Court wrote in 1882,

In *Stephens v. Cady*,²⁰⁸ and again in *Stevens v. Gladding*,²⁰⁹ the point decided was that, by a sale of the copperplate engraving of a map on execution from a State court against the owner of the copyright, the purchaser acquired no right to strike off and sell copies of the map.

Mr. Justice Nelson, in delivering judgment in *Stephens v. Cady*, said:

The copperplate engraving, like any other tangible personal property, is the subject of seizure and sale on execution, and the title passes to the purchaser the same as if made at a private sale. But the incorporeal right, secured by the statute to the author, to multiply copies of the map by the use of the plate, being intangible, and resting altogether in grant, is not the subject of seizure or sale by means of this process²¹⁰

This principle has been reiterated many times throughout the years by many diverse courts.²¹¹ Both the Federal Copyright Statute and the Supreme

208. 55 U.S. 528 (1853).

209. 58 U.S. 447 (1855).

210. *Ager v. Murray*, 105 U.S. 126, 129–30 (1882) (citations omitted).

211. *See* *United States v. Wise*, 550 F.2d 1180, 1187 n.9 (9th Cir. 1977) (“[T]he copyright is distinct from the property which is copyrighted, and the sale of one does not constitute a transfer of the other.”); *Local Trademarks, Inc. v. Price*, 170 F.2d 715, 718 (5th Cir. 1948) (“A copyright is an intangible, incorporeal right in the nature of a privilege or franchise and is independent of any material substance such as the manuscript or plate used for printing. It is entirely disconnected therefrom.”); *Harms v. Cohen*, 279 F. 276, 281 (E.D. Pa. 1922) (“[A] copyright is an intangible thing, and it is separate and distinct from the material object copyrighted”); *Werckmeister v. Am. Lithographic Co.*, 142 F. 827, 830 (C.C.D.N.Y. 1905). The *Werckmeister* court held:

The author of a painting, when it is finished, before publication, owns a material piece of personal property, consisting of the canvas and the paint upon it. He also owns an incorporeal right connected with it; that is, the right to make a copy of it. These two kinds of property, although growing out of the same intellectual production, are in their nature essentially and inherently distinct. The law has always recognized that they are distinct. . . . [T]he law has always recognized that a common-law copyright, before a general publication, is a distinct property from the thing to which the copyright applies. One man may be the owner of the thing, and another of the copyright in the thing.

142 F. at 830; *see also* *Patterson v. J. S. Ogilvie Pub. Co.*, 119 F. 451, 453 (C.C.S.D.N.Y. 1902) (holding that plates sold by sheriff became the property of the purchaser, but gave him no right to publish the copyrighted work which could be printed from them); *Werckmeister v. Springer Lithographing Co.*, 63 F. 808, 812 (C.C.S.D.N.Y. 1894) (holding that author of

Court have recognized that one could own a tangible thing that embodies a copyrighted work, yet might have no right under copyright to copy, modify, or distribute that copyrighted work.

The Agreement Controls approach, which looks to restrictions on the rights of copyright in an effort to determine title to a copy, must therefore be erroneous. Like Cady and Gladding in the nineteenth century, today one can have permission with respect to *none* of the rights of copyright and still be an owner of a copy.²¹²

3. *Must Not Look to Factors that are Wholly Orthogonal to Resolving the Issue of Copy Ownership*
 - a) Copyright Permissions are Irrelevant to a Determination of Copy Ownership

One can have permission to reproduce, permission to make derivatives, permission to distribute, and permission to publicly display or perform and yet still not be an owner of the underlying copies. If that is so, and we acknowledge that one could lack any or all of those permissions and not be an owner of a copy, then one's status with respect to these permissions is necessarily independent of one's status as an owner of a copy. Consequently, when one seeks to determine copy ownership, one should not waste time on these irrelevant and potentially misleading questions. One should instead

painting may sell his painting and retain the right to its copyright); *Hughes Tool Co. v. Fawcett Pubs., Inc.*, 315 A.2d 577, 579–80 (Del. 1974). The *Hughes Tool* court found: 17 U.S.C. § 27 indicates that a statutory copyright “is distinct from the property in the material object copyrighted.” . . . [A] common-law copyright, like its statutory successor, is distinct from that which is copyrighted. It is not a material substance, but is an incorporeal right in the nature of a franchise or privilege of publication.

315 A.2d at 579–80.

212. *Cady* and *Gladding* cannot be distinguished from the software context solely because they involved copperplates from which one would make copies, rather than the copies themselves. The cases following these, cited *supra* note 211, often involved the very embodiment of the work, such as the painting in *Springer Lithographing Co.* Additionally, the nature of engravings is such that the copperplates are not treated differently than the copies they produce. *Springer Lithographing Co.*, 63 F. at 809. For example, one could not create duplicate copperplates without permission and avoid liability for infringement of the reproduction right. This principle also reveals a key error of the Federal Circuit's decision in *DSC Communicationns Corp.* because one could say that Cady and Gladding suffered from rights that were severely limited “in ways that are inconsistent with the rights normally enjoyed by owners” but that had no bearing on whether each held title to the tangible embodiment of the copyrighted work. *DSC Commc'ns Corp. v. Pulse Commc'ns, Inc.*, 170 F.3d 1354, 1361 (Fed. Cir. 1999).

look for a factor that is dispositive in most or many cases. Consider the following hypotheticals:

Suppose Alice is an amateur photographer hoping to make some money from her hobby. Bob operates a film processing and picture-framing shop in which he also showcases the works of local artists, selling the works to the public on the artists' behalf. If Alice brings film in to have it processed and pays for some requested copies (and nothing more), then it is clear from the nature of such transactions that Bob has Alice's permission to make reproductions from her negatives to create the prints she requests. However, Bob neither owns the copies nor has any other right under copyright. For example, he may not make a derivative work or distribute the prints to others.

Suppose further that one of the services Bob offers is to create an oil painting on canvas from any photograph. Alice might provide one of her photographs to Bob and ask for and pay for such an oil painting, and it would be clear that Bob had permission to make this derivative work. It would also be clear that Bob would not own the copy he created, or have permission to make additional reproductions or distributions.

Bob might also make an arrangement with Alice to frame prints of some of her best work and hang it in his shop for potential sale, charging her a flat price for the framing and splitting the proceeds, if any, from sales of the framed prints. One can imagine such an agreement in which Bob has permission to distribute the work to the public, but in which Alice retains ownership of the framed print if it fails to sell. Bob still has no permission to make reproductions or derivatives.

In each of these cases Bob might have or lack any of the permissions of copyright, and in each case, that fact has no bearing on whether or not Bob owns the underlying copies. When Alice pays for prints or an oil painting, it is understood that she intends to return to take possession of the prints or painting. When she sells framed prints through his shop, she will either return to re-claim possession (retaining ownership) of the framed print in the case of no sale, or will sell the prints outright to Bob for potential resale. The dispositive feature in every scenario is whether or not Bob has a right to perpetually possess the copies. When he lacks that right, he is not the owner of the copies, and if he has that right, then he is.

Courts only create confusion when they look to whether an individual has or lacks certain permissions with respect to the underlying copyright in an effort to determine whether the individual is an owner of the tangible copy. These permissions may vary widely across cases, but they are irrelevant since copy ownership status is determined by an entirely different feature of the transaction.

b) Particularly Problematic Permissions

Permissions regarding subsequent transfer or distribution are particularly misleading to a determination of copy ownership for purposes of applying the first sale doctrine. Similarly problematic are permissions regarding reproduction and derivative works rights to a determination of copy ownership for purposes of applying § 117 rights. Both types of analysis involve a methodological circular logic problem because rights under § 109 and § 117 do not require the permission of the copyright owner. If you are an owner of a copy with § 109 rights, then it does not matter what the copyright owner says about distribution as her permission was not needed anyway. If you are an owner of a copy with § 117 rights, then it does not matter what the copyright owner says about reproducing essential-step copies or preparing adaptations because her permission was not needed anyway. One can also easily imagine lease agreements where one has these permissions regarding transfer, but lacks ownership of the copies. Imagine a service from a DVD rental retailer in which after viewing a rented DVD, you can go to the retailer's website and offer the video to a neighbor for renting. If your neighbor consents, they take over the obligation for any rental fees and for returning the DVD on time and you deliver it to your neighbor. In such a system you would have permission to transfer the copy to someone else, but both you and your neighbor would lack ownership of that copy. Again, one's status with respect to these permissions is wholly independent of one's status as a copy owner, so this feature is not dispositive and simply irrelevant to the question of copy ownership. That the Ninth Circuit panel in *Vernor* chose this irrelevant factor to be the second of its three factors for determining copy ownership has actually made the utterly confused state of the law worse.

Courts taking the Agreement Controls approach that look to such restrictions on copyright permissions, such as the *Wall Data* court, make this error. Looking at these features of the transaction tell one nothing dispositive about copy ownership.

B. MUST RESPECT PRECEDENT

The cases that consider ownership of copies seem particularly prone to ignoring relevant precedents. Courts typically do not explicitly refuse to follow an earlier case's approach, but often ignore those precedents through a failure to cite or discuss them. This may be explained by the wide range of cases in which ownership of a copy can be relevant, such as first sale cases, § 117 cases, and even a case primarily addressing a question of bankruptcy

law.²¹³ If the courts search only for those precedents addressing the statute they seek to interpret, they may miss those cases decided under other statutes that address the same underlying issue of ownership of a copy. As the *Vernor* district court held, however, there is no reason to believe that “owner of a particular copy” means one thing in § 109 and “owner of a copy” means something different in § 117.²¹⁴ Thus, whatever the reason for earlier failings, it should be the objective of any interpretive approach to follow controlling precedents.

In particular, the courts frequently fail to ensure that their holdings are consistent with the existing Supreme Court precedents on this issue, particularly the case widely regarded as establishing the first sale doctrine, *Bobbs-Merrill Co. v. Strauss*.

1. *Must Respect Supreme Court Precedent and Should Seek to Harmonize Circuit Court Precedents*

In *Bobbs-Merrill Co. v. Strauss*, the Supreme Court held that while the Copyright Act, then in effect, provided copyright holders with a reproduction and publication right:

[O]ne who has sold a copyrighted article, without restriction, has parted with all right to control the sale of it. The purchaser of a book, once sold by authority of the owner of the copyright, may sell it again, although he could not publish a new edition of it.²¹⁵

The Bobbs-Merrill Company had printed a notice in a book it published which read:

The price of this book at retail is one dollar net. No dealer is licensed to sell it at a less price, and a sale at a less price will be treated as an infringement of the copyright.²¹⁶

This notice shares many characteristics of the terms found in software licensing agreements that courts are asked to interpret over 100 years later. This fact did not escape the district court in *Novell, Inc. v. CPU Distributing, Inc.*, which wrote,

The Court concludes that whether the EULA is otherwise enforceable or not, it is similar to the notice included in the books published by Bobbs-

213. *See* Microsoft Corp. v. DAK Indus. Inc., 66 F.3d 1091 (9th Cir. 1995).

214. *See* Vernor v. Autodesk, Inc., 555 F. Supp. 2d 1164, 1173 n.8 (W.D. Wash. 2008); Vernor v. Autodesk, Inc., No. C07-1189RAJ, 2009 WL 3187613, at *11–14 (W.D. Wash. Sept. 30, 2009), *rev'd*, 621 F.3d 1102 (9th Cir. 2010).

215. 210 U.S. 339, 350 (1908).

216. *Id.* at 341.

Merrill and, as such, does not preclude a finding that a first sale occurs when Novell delivers the copies of its software to the OEMs.²¹⁷

This is the proper analysis. Given that the Supreme Court has held that Bobbs-Merrill's restrictions were unenforceable, there should at least be an investigation of whether software licensing terms that seek to evade the first sale doctrine are suspect as well. Certainly such terms should not be deferred to as key evidence for the proposition that title to the copy has not passed to the transferee.²¹⁸

Additionally, the courts considering transfers of title should, where possible, not create circuit conflicts. There are a series of "waste paper" and salvage cases that held that even where a copyright holder sold copies of his work to another solely for purposes of destruction, such transactions were nonetheless first sales that could not give rise to copyright infringement when the salvage company sold the copies in breach of the agreement.²¹⁹

217. No. H-97-2326, 2000 U.S. Dist. LEXIS 9975, at *8 n.4 (S.D. Tex. May 4, 2000).

218. The *Vernor* panel opinion is unsatisfying in this respect. By permitting software distributors to control all subsequent uses of their distributed software, the panel suggests that Bobbs-Merrill needed only to write a better license agreement in order to enforce post-distribution restrictions on its books. See *Vernor*, 621 F.3d at 1111, 1114. Such an interpretation ignores the extent to which the Supreme Court's holding was animated by concerns over restrictions on the alienability of chattel and by a desire to restrict copyright owners to the narrowly-drawn rights enumerated by the statute. See *Bobbs-Merrill Co.*, 210 U.S. at 350–51. Furthermore, distinguishing between a "notice" in the front of a book and a "license agreement" that accompanies software would be to make a distinction without finding a relevant difference.

219. See *United States v. Wise*, 550 F.2d 1180, 1193 (9th Cir. 1977) ("[I]here would of course be a 'first sale' of any film sold for salvage."); *United States v. Drebin*, 557 F.2d 1316, 1327 (9th Cir. 1977) (assuming sale of film print to a salvage company would constitute a sale for first sale purposes, but finding insufficient evidence that defendants actually purchased prints from salvage companies); *United States v. Atherton*, 561 F.2d 747, 751 (9th Cir. 1977) ("In a criminal prosecution, a sale for salvage purposes can be a first sale."); *Indep. News Co. v. Williams*, 293 F.2d 510, 516 (3d Cir. 1961) (holding that coverless comic books sold for use as waste paper could be resold because of the first sale doctrine); *Harrison v. Maynard, Merrill & Co.*, 61 F. 689, 691 (2d Cir. 1894). The *Harrison* court held:

[T]he right to restrain the sale of a particular copy of the book by virtue of the copyright statutes has gone when the owner of the copyright and of that copy has parted with all his title to it, and has conferred an absolute title to the copy upon a purchaser, although with an agreement for a restricted use. The exclusive right to vend the particular copy no longer remains in the owner of the copyright by the copyright statutes. The new purchaser cannot reprint the copy. He cannot print or publish a new edition of the book; but, the copy having been absolutely sold to him, the ordinary incidents of ownership in personal property, among which is the right of alienation, attach to it. If he has agreed that he will not sell it for certain purposes or to certain persons, and violates his agreement, and sells to an innocent purchaser, he can be punished for a violation of his

Again, if these courts were willing to ignore the breach of such a central term of these agreements and nonetheless find a first sale, present day courts that allow litigants to defeat the first sale doctrine through restrictive terms in licensing agreements appear to have the analysis completely backwards. The courts should instead seek to harmonize their decisions with these cases or at least address the cases to explain why they are distinguishable in a given instance.²²⁰

2. *Should Heed the Warning of the Supreme Court's Recent Quanta Computer Decision*

In *Quanta Computer, Inc. v. LG Electronics, Inc.*,²²¹ the Supreme Court reaffirmed the long-standing doctrine of patent exhaustion, and in so doing, included a discussion of how the Court came to reject its holding in *Henry v. A. B. Dick Co.*,²²² which had permitted, in 1912, post-sale restrictions on the use of a patented article.

A. B. Dick Co. sold its patented Rotary Mimeograph machine with a notice attached that read, "LICENSE RESTRICTION. This machine is sold by the A.B. Dick Co. with the license restriction that it may be used only with the stencil paper, ink and other supplies made by A.B. Dick Company, Chicago, U.S.A."²²³

While the Court upheld this restriction, it did so against vigorous dissent which predicted an uncontrolled multiplication of such license restrictions that would permit contract law to cause the expansion of patent law beyond its intended limits.²²⁴

The following year, 1913, the Court found an opportunity to limit the *A. B. Dick Co.* holding in *Bauer & Cie v. O'Donnell*.²²⁵ James O'Donnell was a proprietor of a drug store who refused to honor a restriction placed by Bauer Chemical on packages of its patented "Sanatogen," a water-soluble

agreement; but neither is guilty, under the copyright statutes, of an infringement.

61 F. at 691.

220. This is another failing of the *Vernor* panel opinion, which distinguished none of these cases. *See Vernor*, 621 F.3d 1102.

221. 553 U.S. 617 (2008).

222. 224 U.S. 1 (1912).

223. *Id.* at 11. The *A. B. Dick* court writes, "the patentee sold its machines at cost, or less, and depended upon the profit realized from the sale of other non-patented articles adapted to be used with the machine, and that it had put out many thousands of such machines under the same license restriction." *Id.* at 26.

224. *Id.* at 51–55 (White, C.J., dissenting).

225. 229 U.S. 1 (1913).

albumenoid.²²⁶ The restriction purported to require all such packages to retail for no less than one dollar. O'Donnell purchased the packages first from Bauer and, after they severed relations with him, from jobbers²²⁷ and sold them for less than a dollar. O'Donnell indicated that he intended to keep doing so.²²⁸

There, despite a license “for sale and use at a price not less than one dollar,” terms which purported to state that purchase was an acceptance of the conditions,²²⁹ and terms stating that any violation of the restrictions caused all rights to revert to the patentee, the Court rejected the characterization of the license as “a license to use the invention.”²³⁰ Noting that “the patentee had no interest in the proceeds of subsequent sales, no right to any royalty thereon or to participation in the profits thereof”²³¹ the Court concluded that the

packages were sold with as full and complete title as any article could have when sold in the open market, excepting only the attempt to limit the sale or use when sold for not less than one dollar. . . . There was no transfer of a limited right to use this invention, and to call the sale a license to use is a mere play upon words. . . . The right to vend conferred by the patent law has been exercised, and the added restriction is beyond the protection and purpose of the act. This being so, the case is brought within that line of cases in which this court from the beginning has held that a patentee who has parted with a patented machine by passing title to a purchaser has placed the article beyond the limits of the monopoly secured by the patent act.²³²

Not content to merely distinguish *A. B. Dick Co.*, the Court took the opportunity four years later to explicitly overrule it in *Motion Picture Patents Co. v. Universal Film Manufacturing Co.*²³³ The *Quanta Computer* Court's summary of the case will suffice here:

[A] patent holder attempted to limit purchasers' use of its film projectors to show only film made under a patent held by the same company. The Court noted the “increasing frequency” with which patent holders were using

226. *See id.* at 8.

227. “Jobbers” are wholesalers who operate on a small scale or who sell only to retailers and institutions, often “by the job.” BLACK'S LAW DICTIONARY, *supra* note 35, at 840.

228. *O'Donnell*, 229 U.S. at 9.

229. Compare to today's shrinkwrap and click-wrap agreements.

230. *O'Donnell*, 229 U.S. at 16.

231. *Id.*

232. *Id.* at 16–17.

233. 243 U.S. 502 (1917). In the interim, on October 15, 1914, Congress enacted the Clayton Antitrust Act, Pub. L. 63-212, 38 Stat. 730, reflecting a similar concern about those with monopolies over-reaching.

A.B. Dick-style licenses to limit the use of their products and thereby using the patents to secure market control of related, unpatented items Observing that “the primary purpose of our patent laws is not the creation of private fortunes for the owners of patents but is “to promote the progress of science and useful arts,” (quoting U.S. CONST. art. I, § 8, cl. 8), the Court held that “the scope of the grant which may be made to an inventor in a patent, pursuant to the [patent] statute, must be limited to the invention described in the claims of his patent.” Accordingly, it reiterated the rule that “the right to vend is exhausted by a single, unconditional sale, the article sold being thereby carried outside the monopoly of the patent law and rendered free of every restriction which the vendor may attempt to put upon it.”²³⁴

The importance of this line of patent exhaustion cases is that their vitality has been affirmed as recently as 2008 by the Supreme Court in *Quanta*.²³⁵ The Supreme Court has found all manner of contractual restrictions on the sale and use of a tangible thing embodying a copyrighted or patented invention invalid. Despite this, some software and music CD distributors believe that they have found a magical restriction that can avoid this problem. By adding a contractual restriction that purports to reserve title in the tangible object itself, while allowing every other feature of the transaction to look like a sale (or vending), they believe they have avoided over a century of Supreme Court precedent and Congressional policy. The only thing more surprising than the fact of this hubris is the success they have had in persuading courts to fall for it.

C. MUST RESPECT CONGRESSIONAL CHOICES

Software manufacturers have invented a new form of distributing software that they have named “licensing.” Because of alleged essential differences between software and other forms of copyrighted works, these distributors seek a special dispensation from the courts to treat the transactions they engage in as immune from the first sale doctrine. Even if software is uniquely situated with respect to its methods of distribution, this is not a determination for the courts to make. As the Supreme Court has written, “[s]ound policy, as well as history, supports our consistent deference to Congress when major technological innovations alter the market for copyrighted materials. Congress has the constitutional authority and the institutional ability to accommodate fully the varied permutations of competing interests that are inevitably implicated by such new

234. *Quanta Computer, Inc. v. LG Elecs., Inc.*, 553 U.S. 617, 626 (2008) (citations omitted).

235. *See id.* at 625–30.

technology.”²³⁶ This judicial deference to congressional policy choices should be reflected in at least two ways that are often neglected in discussions of copy ownership. Judicial deference requires a determination of whether: (1) contractual provisions that seek to subvert the limitations and exceptions articulated in § 107 through § 122 are federally preempted; and (2) whether an approach to determining copy ownership could be used that would preserve the national uniformity that copyright law seeks.

D. MUST TAKE INTO ACCOUNT THAT THE COPYRIGHT ACT EXPLICITLY ENUMERATES THOSE METHODS OF DISTRIBUTION WITHIN THE EXCLUSIVE RIGHTS OF COPYRIGHT OWNERS AND A DISTRIBUTION IN WHICH TITLE TO THE COPY IS RESERVED IS NOT AMONG THEM.

In § 106(3), Congress has given copyright holders the exclusive right to distribute copies or phonorecords of the copyrighted work to the public by:

- sale or
- other transfer of ownership, or
- by rental,
- lease, or
- lending.²³⁷

Upon looking carefully, the first thing to notice is that an exclusive right to distribute by *license* is nowhere to be found. Take each enumerated form of distribution in turn. First we have “sale,” which is most certainly not what copyright holders mean by *license*. Then there is “other transfer of ownership,” but the key feature of *licensing* is that ownership of the copy is not transferred. The next two, “rental” and “lease,” are synonymous in most contexts, though some distinguish the two on the basis of whether the period of possession is auto-renewing. In either case, they both fundamentally involve a limited period of possession in exchange for consideration.²³⁸ It is that feature, consideration, that distinguishes the fifth and final recognized method of distribution, “lending.” Lending is what libraries and neighbors do, and it is done without payment. But all three of these methods of distribution—rental, lease, and lending—fundamentally involve a limited period of possession. For that reason, none of those three methods of distribution can be what is referred to as a *license*, because copyright holders

236. Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 431 (1984).

237. 17 U.S.C. § 106(3) (2006).

238. That consideration is something less than the full value of permanent possession of the thing rented or leased.

who use the term *license* routinely do so in transactions where the copy is provided to the transferee for an unlimited period of possession.

The Copyright Act never granted to copyright holders the exclusive right to distribute by means of *license*.²³⁹ So, copyright holders made it up. But what are the consequences of that fact?

One might argue that copyright holders only have the exclusive right to distribute in the five ways specified in the statute. Any other form of distribution is simply not a form of distribution recognized by the Copyright Act or granted exclusively to copyright holders.²⁴⁰ Even on this view, the consequences are not immediately clear. Does this mean it is forbidden to distribute by *license* or simply that copyright holders can do so, but when doing so they will not be exercising one of their exclusive rights under copyright? If the latter, it would seem to be a consequence of that fact that copyright holders would also not be able to make use of the remedies of copyright to prevent such non-copyright based distributions. That is, if distribution by *license* is not one of the exclusive rights granted to copyright holders, then anyone can do it without infringement of any right of the copyright holder! This would be a puzzling transaction because, assuming a first distribution by *license*, the *licensee* would, perhaps while acknowledging that he holds no title to the copy, offer to *license* the copy to another, which, since he could pass no greater right than he held, would amount to a mere transfer of possession and nothing more. Arguably this might violate some contractual provision of the first exchange, but *ex hypothesi* it could not violate any right of copyright.²⁴¹

239. This is a statement that sounds strange only if one fails to keep the specific definition of *license* that is under consideration in mind.

240. This argument could rely in part on the choice made by Congress to narrow the publication right which had previously used the phrases “publish” and “vend.” The old terminology had forced courts to attempt to interpret the scope of “vend” in particular. In the 1976 Act, Congress clarified by enumerating the methods of distribution it intended to grant exclusively to copyright holders. This strict reading of the statute receives some support from the dissent in *Hotaling v. Church of Jesus Christ of Latter-Day Saints*, 118 F.3d 199 (4th Cir. 1997) (Hall, J., dissenting). Judge Hall argued:

The owner of a copyright does not possess an exclusive right to “distribute” the work in any conceivable manner; instead, it has the exclusive right “to distribute copies . . . of the copyrighted work to the public by sale or other transfer of ownership, or by rental, lease, or lending.”

Id. at 205.

241. An assumption, for the sake of this bizarre scenario, is that the *licensee* does not purport to grant any of the actual rights of copyright of the copyright holder, such as the right to reproduce or prepare derivative works.

One problem with this scenario is that nothing about the transaction would appear any different from a § 106(3) distribution. If the *licensee* offered it on a perpetual basis, it would *look* in every respect like a sale or other transfer of ownership, and if the *licensee* offered it for a limited period of time, it would look either like a rental/lease or lending, depending on whether consideration was received. This is, of course, precisely the problem with distribution by *license* in the first instance by the copyright holder. The copyright holder wishes to insist that it is not an exercise of the distribution right that gives rise to the first sale doctrine, but, in those cases where there is a right of perpetual possession, the transaction looks in every relevant respect like a sale or other transfer of ownership.

Approaching the problem from another angle, we should agree that either distribution by *license* is an exercise of the distribution right or it is not. If it is not, then we have the problems described above: anyone can do it without infringing. This is clearly unacceptable, at least to copyright holders. However, if distribution by *license* is an exercise of the distribution right, then the first sale doctrine should apply. That is the overriding purpose of the first sale doctrine: Congress gives the copyright holder one exercise of the distribution right and after that, the right is exhausted and the first sale doctrine takes over. Thus, advocates of the *licensing* paradigm face a dilemma: either it fails in its purpose because the first sale doctrine still applies or it fails in its purpose because the recipient is free to pass the copy along to another without infringing an exclusive right of the copyright holder.

If this sounds absurd, then recall how we got here. We were faced with two options: either distribution by *license* is forbidden or it is not. If it is forbidden, then all the questions of copy ownership are solved. If it is not forbidden, then given that it is not among the enumerated rights of copyright holders, what are its consequences? And we found the consequences are certainly not what copyright holders would hope and are in fact borderline absurd. In order to get the result that they desire, defenders of distribution by *license* would have to argue both that copyright holders are entitled to distribute by means not enumerated in the Copyright Act and that they are entitled to exclude others from doing the same. That is, they have an exclusive right that is not enumerated. If this is the argument defenders of distribution by *license* are forced into, then things have not gone well for them

at all. Such unenumerated exclusive rights would pose serious due process problems.²⁴²

Even without fully developing the above argument, it is apparent that the better view is that distribution by *license*, since it is not among the methods of distribution enumerated in § 106(3), is not recognized and not available. That is, there is simply no such thing as distribution of a copy by means of a *license*. Instead, every transaction falls into one of two categories: those that provide perpetual possession are sales or other transfers of ownership (e.g., gifts) and those that provide a limited period of possession are either a rental/lease or a lending, depending on whether consideration is required. In the case of copies of software and music, those transactions that fall into the sale or gift category provide first sale rights and those transactions that fall into the lease or lending category do not. As a peripheral benefit, such a view provides enormous administrative simplicity, because nearly the only fact about a transaction that matters to determining copy ownership is whether it allows for a right of perpetual possession.²⁴³

242. *Baggett v. Bullitt*, 377 U.S. 360, 367 (1964) (“[A] law forbidding or requiring conduct in terms so vague that men of common intelligence must necessarily guess at its meaning and differ as to its application violates due process of law.”).

243. One could critique this view by arguing that software distributors could get nearly all of the benefits of distribution by license by simply leasing software for thirty- or forty-year terms, something clearly less than the copyright term but clearly beyond the software’s useful life. The argument would continue that if this is so, then there is no reason to object to distribution by *license*. On the contrary, I believe courts could reject sham leases just as they have rejected “license labels.” There would also be tangible consumer benefits if software distributors were required to label their transactions accurately. Courts could reasonably imply a warranty of merchantability for the period of the lease and refuse to recognize sham lease periods that purport to run beyond the “end of life” or “support lifecycle” of the software. At least one court has already recognized this potential equivalence. *Datalex Ltd. v. PSA, Inc.*, No. 01-06482, 2003 U.S. Dist. LEXIS 27563, at *7 (C.D. Cal. Jan. 30, 2003) (“Due to the rapid changes in technology, which outdates nearly all computer software and hardware every few years, having a 5-year license may well be the same as a perpetual purchase, due to the limited lifetime of any software purchase.”). Once lease periods were constrained to realistic usage periods, consumers would benefit from the clarity provided by calling the transaction what it is and presumably from the lower prices that accompany leases when compared to sales. A purported requirement of return at some distant future date should also be interpreted as a lease or sham lease (i.e., sale) regardless of how the transaction is labeled.

1. *Must Recognize the Limiting Role Played by §§ 107 Through 122 and Examine Whether Contrary Contractual Provisions are Federally Preempted*

The rights of copyright enumerated in 17 U.S.C. § 106 are “Subject to sections 107 through 122”²⁴⁴ This language sets up a structure through which a list of the rights of copyright will be enumerated in § 106, but the precise scope of those rights can only be determined by reading § 106 in conjunction with the limitations and exceptions in §§ 107 through 122.

When one of these limitations is found to apply, it does not represent an “excused” infringement of a § 106 right, but rather indicates that the § 106 right at issue never extended that far to begin with. Congress has carefully chosen the contours of these rights, and courts should raise doubts about the ability of one party to a typically un-negotiated contract of adhesion between parties of unequal bargaining power to re-draw these lines that Congress has chosen. At a minimum, it seems erroneous to look at whether an agreement purports to limit the permissions given with respect to the rights of copyright in an effort to determine whether the licensee has rights Congress put beyond the control of the copyright holder to begin with.

The point may be best expressed by way of an example from another context. Imagine trying to determine whether individuals could make a fair use of a work under § 107 by looking to any agreements accompanying the copyrighted work to see if there were terms in the agreement that restricted the recipient’s fair use rights. It would be a bizarre methodology that would entitle every copyright holder to simply do away with the fair uses of his work by fiat. Congress could never have intended that result.²⁴⁵ Nonetheless, courts routinely accept a similar methodology when trying to determine if individuals have first sale rights under § 109 or “essential step” rights under § 117. Courts would reach better results if they thought of § 109 and § 117 as specific fair uses that Congress felt so important that it decided to articulate them separately, rather than simply relying on § 107 to cover these cases.

244. 17 U.S.C. § 106 (2006).

245. It would also raise constitutional issues under the First Amendment, but those are not the focus here. The handful of decisions that have suggested that a consumer may contract away fair use rights through a click-wrap were wrongly decided. *See, e.g.*, *Davidson & Assocs. v. Jung*, 422 F.3d 630, 639 (8th Cir. 2005); *Bowers v. Baystate Techs., Inc.*, 320 F.3d 1317, 1325–26 (Fed. Cir. 2003). The better view is found in Judge Dyk’s opinion in *Bowers* and in other decisions. *See Bowers*, 320 F.3d at 1335–38 (Dyk, J., concurring in part and dissenting in part); *Vault Corp. v. Quaid Software Ltd.*, 847 F.2d 255, 268–70 (5th Cir. 1988).

Seen through that lens, the approaches taken by many courts on this issue leave much to be desired.

A full discussion of whether such contractual arrangements are federally preempted is beyond the scope of this Article. Briefly, when the *Step-Saver Data* court provided its useful history of how the software industry came to “license” but not sell copies of its works in an effort to avoid the reach of the first sale doctrine, it noted, “Questions remained, however, as to whether the use of state contract law to avoid the first sale doctrine would be preempted either by the federal copyright statute (statutory preemption) or by the exclusive constitutional grant of authority over copyright issues to the federal government (constitutional preemption).”²⁴⁶

But while the rental of software that concerned the industry was addressed by Congress, the courts have not seriously taken up the other questions posed by the *Step-Saver Data* court, as none of the courts to address this issue have a substantive discussion of federal preemption and only a handful mention it at all.²⁴⁷

Courts have, in the past, found the scope of federal preemption to be quite broad in the copyright context. The Supreme Court wrote:

These [patent and copyright] laws, like other laws of the United States enacted pursuant to constitutional authority, are the supreme law of the land. When state law touches upon the area of these federal statutes, it is “familiar doctrine” that the federal policy “may not be set at naught, or its benefits denied” by the state law.²⁴⁸

Similarly, another court has noted, “even apart from Section 301, the general proposition pertains in copyright law, as elsewhere, that a state law is

246. See *Step-Saver Data Sys. v. Wyse Tech.*, 939 F.2d 91, 96 n.7 (3d Cir. 1991) (citations omitted).

247. *Vault Corp.*, 847 F.2d at 269–70; *SoftMan Prods. Co. v. Adobe Sys. Inc.*, 171 F. Supp. 2d 1075, 1083 n.10 (C.D. Cal. 2001) (citing *Step-Saver Data*, 939 F.2d at 96 n.7); *Foresight Res. Corp. v. Pfortmiller*, 719 F. Supp. 1006, 1010 (D. Kan. 1989) (citing *Vault*, 847 F.2d at 268–70).

248. *Sears, Roebuck & Co. v. Stiffel Co.*, 376 U.S. 225, 229 (1964) (quoting *Sola Elec. Co. v. Jefferson Elec. Co.*, 317 U.S. 172, 173, 176 (1942)); see also *Bonito Boats v. Thunder Craft Boats*, 489 U.S. 141, 168 (1989). The question courts should address is whether Congress, in enacting § 107 through § 122 has determined that these limitations are among those within the national interest that cannot be set at naught. Unfortunately, courts have done too little to develop this line of inquiry. See David Nimmer, Elliot Brown & Gary N. Frischling, *The Metamorphosis of Contract into Expand*, 87 CALIF. L. REV. 17, 41 (1999).

invalid that ‘stands as an obstacle to the accomplishment of the full purposes and objectives of Congress.’”²⁴⁹

The precise contours of this preemption can only begin to be determined if a court takes up the question of whether the license agreement that it typically would take at face value “stands as an obstacle to the accomplishment of the full purposes and objectives of Congress”²⁵⁰ or “sets at naught” a federal policy embodied in the Copyright Act.²⁵¹

2. *Should Preserve as Far as Possible the National Uniformity Copyright Law Seeks*

The Second Circuit in *Krause* provided an argument that it put forward as a reason against interpreting § 117(a) to require formal title, but which is also an argument against using an Agreement Controls approach to determining copy ownership.²⁵² The *Krause* court explained that if we must look to the agreement to determine ownership of title, then we must invariably interpret that agreement. When courts interpret agreements, they tend to look to the governing law clause which then leads them to adopt the varying interpretative canons of construction of each state. This leads to conflict with the Copyright Act’s “express objective of creating national, uniform copyright law by broadly preempting state statutory and common-law copyright regulation.”²⁵³ The Perpetual Possession approach, by focusing almost exclusively on the term of possession, provides an approach that does not pose such conflicts with congressional will.

Alternatively, courts could develop a consistent federal common law approach to the question of copy ownership by more explicitly adopting the U.C.C.’s passage of title rules. Many courts have concluded that the sale of software is a sale of goods governed by the U.C.C. or a given state’s implementation of the U.C.C.²⁵⁴ The critical question in copy ownership cases is a determination of the owner of a tangible good. This is a subject that is directly addressed by the U.C.C., which reflects an extensive experience

249. *Facenda v. N.F.L. Films, Inc.*, 542 F.3d 1007, 1028 (3d Cir. 2008) (citing 1 NIMMER ON COPYRIGHT § 1.01[B][3][a] (2008)).

250. *Id.*; *Nimmer et al.*, *supra* note 248, at 40–41. (quoting *Hines v. Davidowitz*, 312 U.S. 52, 67 (1941)).

251. *Sola Elec. Co.*, 317 U.S. at 176.

252. *Krause v. Titleserv, Inc.*, 402 F.3d 119, 123 (2d Cir. 2005).

253. *Id.* (citing *Cnty. for Creative Non-Violence v. Reid*, 490 U.S. 730, 740 (1989)); *see also* 17 U.S.C. § 301(a) (2006).

254. *See, e.g.*, Mark Lemley, *Intellectual Property and Shrinkwrap Licenses*, 68 S. CAL. L. REV. 1239, 1244 n.23 (1995) (collecting cases); Note, *Computer Programs as Goods Under the U.C.C.*, 77 MICH. L. REV. 1149 (1979).

with such commercial transactions. Thus, whether federal common law or some other rule governs the transfer of title in copies of copyrighted works, such a scheme should at least be informed by the approach taken by the U.C.C.²⁵⁵

The U.C.C.'s rules for the passing of title suggest by analogy that consumers purchasing copies of software should be deemed to receive title to the copies at the time and place at which the copyright owner (or its retailer) completes performance by delivering the copy.²⁵⁶ Furthermore, any purported reservation of title to the copy should be limited to a reservation of a security interest. Where a consumer pays full price in one lump sum, no security interest would exist and the consumer should simply be deemed the owner of the copy immediately upon full payment.

Furthermore, the sale of tangible goods such as watches, bottles of shampoo, or boxes of diapers are precisely the sorts of things governed by Article 2 of the U.C.C. However, these tangible goods can also embody or contain a copyrighted work.²⁵⁷ Since not every watch, shampoo bottle, or box of diapers contains a copyrighted work, we have a situation where ownership of some tangible goods might be determined solely by reference to the U.C.C. principles in section 2-401. Conversely, ownership of other tangible goods containing copyrighted materials might be determined by any one of the other approaches described in this Article, depending upon the court. Those courts that apply approaches that are extremely deferential to the copyright owner will reach results that yield whatever ownership status the copyright owner prefers, which may well be at odds with the result U.C.C. section 2-401 would have counseled. Such contradictory results are extremely unsatisfying, suggesting that harmonizing the federal common law regarding transfer of title to copies of copyrighted works with the U.C.C.'s title-passing principles would be far more sensible.

This Article advocates for the Perpetual Possession approach, but at least one understanding of a U.C.C. Controls approach could yield equivalent results. Section 2-401 basically does not permit a reservation of title, and thus

255. *United States v. Ravel*, 930 F.2d 721, 725 (9th Cir. 1991) (noting that the U.C.C. may be relevant to the import of a term used in a federal statute); *Sierra Pac. Indus. v. Lyng*, 866 F.2d 1099, 1110 n.16 (9th Cir. 1989) ("Federal courts frequently draw upon the Uniform Commercial Code in creating federal common law.").

256. *See supra* Section III.C.

257. *See Quality King Distributions v. L'Anza Research Int'l*, 523 U.S. 135 (1998) (shampoos, conditioners, and other hair care products with copyrighted labels affixed thereto); *Omega S.A. v. Costco Wholesale Corp.*, 541 F.3d 982 (9th Cir. 2008) (watches with copyrighted engraving on back); *Denbicare U.S.A. Inc. v. Toys "R" Us, Inc.*, 84 F.3d 1143 (9th Cir. 1996) (diapers in copyrighted packaging).

under both approaches, when a permanent transfer occurs, a right of perpetual possession exists, and a sale is deemed to have occurred.

VI. HOW TO ADDRESS *MDY INDUSTRIES LLC V. BLIZZARD ENTERTAINMENT, INC.*

Purchasers of Blizzard's World of Warcraft ("WoW") pay a single amount up front for the client software, and nothing in Blizzard's EULA or TOU reflect a requirement that WoW purchasers return the software to Blizzard after any specified period. It is theirs to keep, destroy, or give away for as long as they like. Under the Perpetual Possession approach, this leads to the conclusion that WoW purchasers are owners of their copies. As owners of copies, they are entitled, under § 117(a)(1), to make another copy of WoW, provided that such a new copy is created as an essential step in the utilization of WoW in conjunction with a machine and that it is used in no other manner.²⁵⁸ When an owner of a copy of WoW runs WoW in conjunction with Glider, a RAM copy of portions of WoW is created as an essential step in the utilization of WoW in conjunction with their computer.²⁵⁹ That RAM copy is used in no other manner. The creation of the RAM copy is an essential step in the utilization of WoW in conjunction with the user's computer because the creation of RAM copies is an automated process that necessarily occurs in the utilization of any program. WoW simply could not be utilized at all in conjunction with a computer without the creation of a RAM copy, making its creation an "essential step" in its utilization. The RAM copy is used in "no other manner" because Glider is *not* a cracking or copying tool that enables the creation of copies.²⁶⁰

Thus, if WoW users were entitled to make RAM copies of WoW under § 117, then even if such users breached their agreement with Blizzard with respect to using bots, they were not direct infringers of Blizzard's copyright. Consequently, MDY Industries is not liable for secondary infringement. This result is reasonable because it is profoundly odd that assisting someone to break the rules of a game would constitute secondary copyright infringement.

258. 17 U.S.C. § 117(a)(1) (2006).

259. It is probably never the case that the entirety of a complex software program, such as WoW, is in a user's RAM simultaneously, but instead, primarily those portions necessary for operating the software at that moment are in RAM. Whether these serially-present snippets of the program in RAM should even be considered a "copy" or "fixed" are further issues worthy of careful reflection. For the purposes of this argument, I will refer to what is created in RAM as a "copy."

260. The "no other manner" requirement seems motivated by an interest in preventing the creation of infringing copies. The interpretation of each of the elements of § 117 are adeptly addressed at length in *Krause v. Tittelser, Inc.*, 402 F.3d 119, 125–30 (2d Cir. 2005).

No one likes cheaters, but Blizzard has alternative methods of dealing with bots, and could adopt a “carrot and a stick” approach. It could offer the carrot of creating alternative Battle.net server networks where bots are allowed, and utilize the stick of continuing to terminate users who use bots on servers where they are not allowed. Blizzard also has the ability to terminate the Battle.net account of a user violating the terms of that service, forcing them to sacrifice the remainder of the month’s service already paid for, and could also use authorization codes with the client software, forcing banned users to purchase another copy of WoW in order to play again. The cost to users of new Battle.net accounts and new authentication codes would be a strong deterrent to repeat botting, especially where alternative networks existed where it was allowed.

VII. CONCLUSION

To determine whether title to a copy has been transferred, courts should look to whether the transferee has a right of perpetual possession of the copy. This is the key factor that distinguishes sales and gifts on the one hand, and leases and lending on the other. The invented notion of “*licensing*” software, where that means transferring perpetual possession of a copy but retaining title to the copy, is both incoherent and not found in the Copyright Act. Courts that look to the agreement accompanying a copy are focusing on the wrong factors for determining ownership of the copy, unless they stay focused on the right of perpetual possession, as the controlling precedents in both the Ninth and Second Circuits say that they should. Where other courts have gone astray, the *Augusto* and *Vernor* district courts found the correct path and demonstrated an approach that is logical, that respects controlling precedent and congressional choices, and that is easy to apply yielding results that square with settled consumer expectations. Thus, the *Augusto* and *Vernor* district court opinions show us not only how to resolve *MDY Industries*, but also the question of copy ownership in many future cases.

IQBAL-ING SEAGATE: PLAUSIBILITY PLEADING OF WILLFUL PATENT INFRINGEMENT

Damon C. Andrews[†]

TABLE OF CONTENTS

I.	INTRODUCTION	1956
II.	BACKGROUND: PLEADING STANDARDS AND WILLFUL INFRINGEMENT	1958
	A. <i>TWOMBLY: GETTING THE BALL ROLLING</i>	1959
	B. <i>IQBAL-ING TWOMBLY: PLAUSIBILITY FOR ALL</i>	1961
	C. THE NATURE OF PATENT INFRINGEMENT AND THE HISTORY OF INCREASED DAMAGES	1963
	1. <i>Patent Infringement as a Strict-Liability Offense</i>	1963
	2. <i>The History of Increased Damages</i>	1964
	D. OBJECTIVELY HIGHER: THE FEDERAL CIRCUIT’S NEW STANDARD OF WILLFULNESS	1965
III.	GETTING COMPLAINTS THROUGH THE (SEA)GATE: APPLYING IQBAL TO WILLFUL INFRINGEMENT	1967
	A. POSSIBILITY, PLAUSIBILITY, PROBABILITY: WHAT IS THE ACTUAL STANDARD?.....	1967
	B. WHAT ABOUT THE RULES?.....	1969
	1. <i>The Sample Complaint of Form 18</i>	1970
	2. <i>Liberal Constructions Under Rule 8</i>	1972
	3. <i>Amending Pursuant to Rule 15</i>	1973
IV.	THE IQBAL AFTERMATH: DAMAGES, DICHOTOMY, AND THE DESERTION OF WILLFULNESS	1975
	A. DAMAGES: TREBLE DAMAGES AND THE INCENTIVE TO SUE.....	1976

© 2010 Damon C. Andrews.

† J.D. Candidate, University of Iowa College of Law, 2011; B.S., Clemson University, 2008. For their mentorship, guidance, and insightful comments on an earlier draft, I thank Professors Christina Bohannon and Herbert Hovenkamp. I am also grateful to Mark Lemley for his helpful comments on an earlier draft. For their support and encouragement, I thank my parents, Richard and Tonya Andrews; my sisters, Lindsay Walker and Kelsey Andrews; and my grandfather, Howard Hillegas.

B.	DICHOTOMY: COMPARING WILLFUL PATENT INFRINGEMENT TO OTHER BAD-FAITH INFRINGEMENTS.....	1978
1.	<i>Copyright Law's Recklessness Standard</i>	1979
2.	<i>Inequitable Conduct: An Expressly Higher Threshold?</i>	1981
C.	DESERTION: THE RETIREMENT OF WILLFULNESS.....	1983
1.	<i>Close-Knit Technology Communities: A Solution?</i>	1984
2.	<i>Alternative Criteria for Increasing Damages</i>	1985
V.	CONCLUSION	1988

I. INTRODUCTION

“Few issues in civil procedure jurisprudence are more significant than pleading standards, which are the key that opens access to courts.”¹ Practically speaking, pleadings are the necessary gateway to any civil action and unlock the doors to discovery.² In general, a complaint should set forth the basic facts known to the plaintiff and serves to put defendants on notice of the allegations and resulting harm that the plaintiff seeks to remedy.³ However, ascertaining the depth of the harm and defendants’ specific injurious conduct frequently requires extensive discovery procedures.⁴ In the interest of serving judicial economy and avoiding frivolous litigation,⁵ these

1. Phillips v. Cnty. of Allegheny, 515 F.3d 224, 230 (3d Cir. 2008).

2. See, e.g., Christopher M. Fairman, *The Myth of Notice Pleading*, 45 ARIZ. L. REV. 987, 988 (2003) (“Pleading is the gateway to the federal courts.”); Lonny S. Hoffman, *Burn up the Chaff with Unquenchable Fire: What Two Doctrinal Intersections Can Teach Us About Judicial Power over Pleadings*, 88 B.U. L. REV. 1217, 1232 (2008) (considering the possible influences for “more rigorous judicial gatekeeping at the pleading stage”); Douglas G. Smith, *The Twombly Revolution?*, 36 PEPP. L. REV. 1063, 1099 (2009) (discussing “the importance of the Rule 8(a) pleading standard as the gateway to further proceedings under the generous discovery provisions of the Federal Rules”).

3. See Michael Moffit, *Pleadings in the Age of Settlement*, 80 IND. L.J. 727, 729–33 (2005) (“[M]odern pleadings provide most defendants, in most circumstances, with all of the information they genuinely need to understand the legal and factual scope of the litigation at hand.”).

4. See Rivera v. NIBCO, Inc., 384 F.3d 822, 824 (9th Cir. 2004) (“[P]re-trial discovery has been recognized as an essential means for evaluation of damages . . .”).

5. See Robert G. Bone, *Modeling Frivolous Suits*, 145 U. PA. L. REV. 519, 520 (1997) (noting the “widespread belief that frivolous litigation is out of control”); Arthur R. Miller, *The Pretrial Rush to Judgment: Are the “Litigation Explosion,” “Liability Crisis,” and Efficiency Clichés Eroding Our Day in Court and Jury Trial Commitments*, 78 N.Y.U. L. REV. 982, 984 (2003) (“[E]xcessive and frivolous litigation overwhelms the judicial system’s capacity to administer speedy and efficient justice, leads to higher costs for litigants and society at large, and even hinders America’s competitive position in the global economy.”).

discovery procedures are generally only available once a complainant has sufficiently pled facts upon which he has shown that he is entitled to relief.⁶

In 2009, the Supreme Court in *Ashcroft v. Iqbal*⁷ raised the bar for what must be pled to survive defendants' motions to dismiss and to begin the discovery process. The Court abandoned the "no set of facts" language that prevailed as the judicial interpretation of Federal Rule of Civil Procedure 8(a)(2)⁸ for half a century, and elevated the standard to one where the pleadings must "state a claim to relief that is plausible on its face."⁹ Just two years earlier, in *In re Seagate Technology, LLC*,¹⁰ the Court of Appeals for the Federal Circuit similarly abandoned its own well-established negligence standard for proving willful patent infringement. Relying on the Supreme Court's latest interpretation of "willful,"¹¹ the Federal Circuit elevated the threshold of willfulness to a level of "objective recklessness."¹²

This Note argues that when taken together, *Iqbal* and *Seagate* will raise the bar for sufficiently pleading willful patent infringement beyond any attainable level. Part II describes the Supreme Court's recent decisions that heighten the standard for Rule 8(a)(2) pleadings and the Federal Circuit's shift from a negligence standard to a recklessness standard for proving willful patent infringement. Part III provides a review of the plausibility-standard language and suggests three mechanisms rooted in the Federal Rules of Civil Procedure that patentee-plaintiffs can employ to circumvent the stringent threshold to survive defendants' motions to dismiss. Part IV evaluates the new standard for willful patent infringement by comparing it to standards for bad-faith infractions in other areas of intellectual property law. Finally, Part IV also contemplates whether the courts should desert willfulness as the hallmark indicium for increased damages in favor of other bad-faith characteristics of the defendant or conduct surrounding the infringement.

6. See Patricia M. Wald, *Summary Judgment at Sixty*, 76 TEX. L. REV. 1897, 1925 (1998) (stating that pleadings "establish[] narrow gateways through which plaintiffs must pass before they can obtain any discovery").

7. 129 S. Ct. 1937 (2009).

8. "A pleading that states a claim for relief must contain . . . a short and plain statement showing that the pleader is entitled to relief." FED. R. CIV. P. 8(a)(2).

9. *Iqbal*, 129 S. Ct. at 1949 (quoting *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 570 (2007)).

10. *In re Seagate Tech., LLC*, 497 F.3d 1360 (Fed. Cir. 2007) (en banc), cert. denied sub nom. *Convolve, Inc. v. Seagate Tech., LLC*, 552 U.S. 1230 (2008).

11. See *infra* notes 76–80 and accompanying text (discussing two Supreme Court cases that elaborated on the meaning of "willful").

12. *Seagate*, 497 F.3d at 1371.

II. BACKGROUND: PLEADING STANDARDS AND WILLFUL INFRINGEMENT

The Supreme Court's decision in *Conley v. Gibson*¹³ governed the standard of notice pleadings in civil cases for half a century. In an opinion by Justice Black, the Court announced that "a complaint should not be dismissed for failure to state a claim unless it appears beyond doubt that the plaintiff can prove no set of facts in support of his claim which would entitle him to relief."¹⁴ However, in 2007, the Court fractured the "no set of facts" bedrock with its decision in *Bell Atlantic Corp. v. Twombly*,¹⁵ raising the bar for complaints—at least in antitrust cases—to survive Rule 12(b)(6) motions.¹⁶ For nearly two years to the day,¹⁷ the legal community debated the extent of the rift between the *Twombly* and *Conley* interpretations of the Rule 8(a)(2) foundation.¹⁸ Finally, in May 2009, that crack, since widened to a crevice,¹⁹ became a canyon when the Court announced its decision in *Ashcroft v. Iqbal*,²⁰ leaving no doubt that the *Twombly* standard applied to all civil actions.²¹

While the Supreme Court was reviewing *Twombly* on certiorari,²² the Court of Appeals for the Federal Circuit was occupied with its own standard-setting opinion, *In re Seagate Technology, LLC*.²³ In *Seagate*, the Federal Circuit established a new benchmark for proving that defendants willfully infringed a plaintiff's patent: objective recklessness.²⁴ In doing so, the court expressly

13. 355 U.S. 41 (1957).

14. *Id.* at 45–46.

15. 550 U.S. 544 (2007).

16. FED. R. CIV. P. 12(b)(6) (permitting dismissal of a complaint when a plaintiff "fail[s] to state a claim upon which relief can be granted").

17. *Twombly*, 550 U.S. 544, was decided on May 21, 2007, and *Ashcroft v. Iqbal*, 129 S. Ct. 1937, was decided on May 18, 2009.

18. *See, e.g.*, A. Benjamin Spencer, *Plausibility Pleading*, 49 B.C. L. REV. 431, 458 (2008) ("[I]t is hard to understand how the *Twombly* approach would not apply in other types of cases."); Amy J. Wildermuth, *What Twombly and Mead Have In Common*, 102 NW. U. L. REV. 276, 277 (2008) (discussing ways in which the Supreme Court can "end the speculation regarding what *Twombly* means"); *The Supreme Court, 2006 Term Leading Cases—Pleading Standards*, 121 HARV. L. REV. 305, 310 (2007) (expressing uncertainty over "how big of an effect *Twombly* will ultimately have on pleading practice").

19. Even before the Supreme Court announced its *Iqbal* decision, other courts across the country were already applying the *Twombly* plausibility standard to non-antitrust cases. *See* sources cited *infra* note 40.

20. 129 S. Ct. 1937 (2009).

21. *Id.* at 1953 (stating that the *Twombly* Court's interpretation of Rule 8 "expounded the pleading standard for 'all civil actions'" (quoting *Twombly*, 550 U.S. at 554)).

22. *Bell Atl. Corp. v. Twombly*, 548 U.S. 903 (2006) (granting certiorari).

23. 497 F.3d 1360 (Fed. Cir. 2007) (en banc).

24. *Id.* at 1371.

overruled its own twenty-four-year-old negligence standard established in *Underwater Devices, Inc. v. Morrison-Knudsen Co.*²⁵

It is now only a matter of time until the *Twombly–Iqbal* pleading standard collides with the *Seagate* willfulness standard.²⁶ In considering how to resolve this possible conflict, it is necessary to understand the backdrop upon which these decisions were founded. While most legal scholarship offers an extensive review of Rule 8(a)(2) in the years following *Conley* and focuses on the groundbreaking implications of *Twombly*, this Note employs *Twombly* as its departure point. As professors Kevin Clermont and Stephen Yeazell hypothesize, just as “*Twombly* generated a mountain of commentary from academics . . . a new round of tumult over the combined cases [of *Twombly* and *Iqbal*] surely lies ahead.”²⁷ This Note adds to that tumult in an effort to analyze the effects of *Iqbal* on an important and growing area of the law.

A. TWOMBLY: GETTING THE BALL ROLLING

In *Bell Atlantic Corp. v. Twombly*, two members of a class of local telephone and Internet subscribers alleged that the defendant service providers violated section 1 of the Sherman Act.²⁸ The Southern District of New York dismissed the plaintiffs’ complaint pursuant to Rule 12(b)(6)²⁹ On appeal, the Second Circuit reversed, citing *Conley*, and held that a plaintiff need not plead specific facts in addition to parallel conduct to survive a motion to dismiss.³⁰

25. 717 F.2d 1380 (Fed. Cir. 1983).

26. *Ashcroft v. Iqbal* has been cited a combined 20,081 times by federal circuit courts since its adoption in May 2009. However, only five of those citations belong to the Court of the Appeals for the Federal Circuit, none of which relate to willful patent infringement. Focused Shepard’s Search Results for “129 S. Ct. 1937,” LEXISNEXIS.COM, <http://www.lexisnexis.com> (last visited Mar. 5, 2011).

27. Kevin M. Clermont & Stephen C. Yeazell, *Inventing Tests, Destabilizing Systems*, 95 IOWA L. REV. 821, 823 n.4 (2010).

28. 15 U.S.C. § 1 (2006); *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 550 (2007).

29. *Twombly v. Bell Atl. Corp.*, 313 F. Supp. 2d 174, 189 (S.D.N.Y. 2003) (dismissing the complaint for failure to set forth a “reason to believe that defendants’ parallel conduct was reflective of any agreement”), *vacated*, 425 F.3d 99 (2d Cir. 2005), *rev’d*, 550 U.S. 544 (2007).

30. *Twombly*, 425 F.3d at 114 (stating that “plus factors are not *required* to be pleaded to permit an antitrust claim based on parallel conduct to survive dismissal”); *see also* 15 U.S.C. § 1 (2006) (requiring a “contract, combination . . . or conspiracy, in restraint of trade or commerce” to violate section 1 of the Sherman Act). For a discussion and explanation of parallel conduct, *see* 6 PHILLIP E. AREEDA & HERBERT HOVENKAMP, *ANTITRUST LAW: AN ANALYSIS OF ANTITRUST PRINCIPLES AND THEIR APPLICATION* ¶ 1425a, at 181–82 (3d ed. 2010) (“‘Unnatural’ parallelism refers to parallel behavior that would probably not result from chance, coincidence, independent responses to common stimuli, or mere interdependence unaided by an advance understanding among the parties.”).

The Supreme Court granted certiorari “to address the proper standard for pleading an antitrust conspiracy through allegations of parallel conduct.”³¹

In a highly criticized opinion,³² the Court reversed the Second Circuit.³³ Writing for the Court, Justice Souter abandoned the “no set of facts” standard set forth in *Conley* and condemned the phrase as an “incomplete, negative gloss on an accepted pleading standard.”³⁴ The Court reasoned that the language of *Conley*, when read literally and in isolation, would allow any theory of a claim to survive a motion to dismiss so long as the pleadings were not factually impossible.³⁵ While noting that a well-pleaded complaint may still proceed even if it appears that recovery is “very remote and unlikely,”³⁶ the Court nevertheless required the plaintiffs to plead specific facts that would “nudge[] their claims across the line from conceivable to plausible.”³⁷

In his dissent, Justice Stevens anticipated the uncertainty that would occupy scholars’ and judges’ minds alike in the wake of *Twombly*: “Whether the Court’s actions will benefit only defendants in antitrust treble-damages cases, or whether its test for the sufficiency of a complaint will inure to the benefit of all civil defendants, is a question that the future will answer.”³⁸ From the language of the opinion, it appeared that the standard applied only

31. *Twombly*, 550 U.S. at 553.

32. See, e.g., Stephen B. Burbank, *Pleading and the Dilemma of “General Rules,”* 2009 WIS. L. REV. 535, 560 (noting the “enormous confusion and transaction costs as a result of [the] uncertainty” imposed by *Twombly*); Lee Goldman, *Trouble for Private Enforcement of the Sherman Act: Twombly, Pleading Standards, and the Oligopoly Problem*, 2008 BYU L. REV. 1057, 1058 (2008) (“*Twombly* is a confusing opinion replete with inconsistent statements . . . [I]t is impossible to discuss the [*Twombly*] opinion without being critical of its lack of clarity.”); A. Benjamin Spencer, *Understanding Pleading Doctrine*, 108 MICH. L. REV. 1, 6, 9 (2009) (criticizing the Court’s opinion for “inconsistent rhetoric” and a “lack of precision”). But see, e.g., Robert G. Bone, *Twombly, Pleading Rules, and the Regulation of Court Access*, 94 IOWA L. REV. 873, 890 (2009) (“[I]t is wrong to condemn *Twombly*’s plausibility standard for being inconsistent with the language of Rule 8(a)(2) or the intent of the 1938 Advisory Committee.”); Allan Ides, *Bell Atlantic and the Principle of Substantive Sufficiency Under Federal Rule of Civil Procedure 8(a)(2): Toward a Structured Approach to Federal Pleading Practice*, 243 F.R.D. 604, 605 (2006) (“Assertions that [Rule 8(a)(2)] does not require the pleading of facts . . . [are] demonstrably false.”); Smith, *supra* note 2, at 1097 (“The Court’s decision [in *Twombly*] is fully consistent with the text of the rules as well as their underlying rationale. As such, it articulates a standard that is neither ‘vague’ nor ‘confusing,’ nor a ‘startling’ deviation from existing precedent.”).

33. *Twombly*, 550 U.S. at 570.

34. *Id.* at 563 (concluding that the criticism and questioning by courts of the *Conley* “no set of facts” language earned the phrase its retirement).

35. *Id.* at 561.

36. *Id.* at 556 (citing *Scheur v. Rhodes*, 416 U.S. 232, 236 (1974)).

37. *Id.* at 570.

38. *Id.* at 596 (Stevens, J., dissenting).

to antitrust cases.³⁹ Courts across the country, however, opined (correctly, as the Supreme Court later determined in *Iqbal*) that the *Twombly* plausibility standard reached beyond the scope of antitrust complaints.⁴⁰ The question Justice Stevens posed would be answered when the Court granted certiorari to determine the sufficiency of a Pakistani prisoner's allegations that he was deprived of constitutional protections while in federal custody.⁴¹

B. *IQBAL*-ING *TWOMBLY*: PLAUSIBILITY FOR ALL

In the aftermath of September 11, 2001, the FBI launched an investigation to seek out and identify the individuals involved in the terrorist attacks.⁴² The Bureau detained a target of the investigation, petitioner and Pakistani–Muslim Javaid Iqbal, and held him in the Administrative Maximum Special Housing Unit of the Metropolitan Detention Center in Brooklyn, New York.⁴³ In his *Bivens*⁴⁴ complaint, Iqbal alleged that he was deprived of various constitutional protections during his confinement, including being subjected to harsh conditions on account of his race, religion, or national

39. *Id.* at 554–55 (majority opinion) (“This case presents the antecedent question of what a plaintiff must plead in order to state a claim under § 1 of the Sherman Act.”); *id.* at 556 (“In applying these general standards to [a 15 U.S.C.] § 1 claim, we hold that stating such a claim requires a complaint with enough factual matter (taken as true) to suggest that an agreement was made.”).

40. *See, e.g.*, *Robinson v. Am. Honda Motor Corp.*, 551 F.3d 218, 222 (4th Cir. 2009) (applying *Twombly* to a breach-of-warranty claim); *Phillips v. Cnty. of Allegheny*, 515 F.3d 224, 234 (3d Cir. 2008) (“[W]e decline at this point to read *Twombly* so narrowly as to limit its holding on plausibility to the antitrust context.”); *McZeal v. Sprint Nextel Corp.*, 501 F.3d 1354, 1362 (Fed. Cir. 2007) (Dyk, J., concurring in part and dissenting in part) (“Plainly, *Bell Atlantic* applies outside the antitrust context.”); *Iqbal v. Hasty*, 490 F.3d 143, 157 (2d Cir. 2007) (“We are reluctant to assume that all of the language of *Bell Atlantic* applies only to [15 U.S.C.] section 1 allegations based on competitors’ parallel conduct or, slightly more broadly, only to antitrust cases.”), *rev’d on other grounds sub nom.* *Ashcroft v. Iqbal*, 129 S. Ct. 1937 (2009).

41. *Ashcroft v. Iqbal*, 128 S. Ct. 2931 (2008) (granting certiorari).

42. *Iqbal*, 129 S. Ct. at 1943.

43. *Id.* According to Iqbal’s complaint, he was arrested in November 2001 on charges of fraud in relation to identification documents and conspiracy to defraud the United States. *Iqbal*, 490 F.3d at 147–48.

44. Iqbal’s complaint included a *Bivens* claim and allegations that federal officials violated his First and Fifth Amendment rights through targeting him by using constitutionally prohibited factors. *Iqbal*, 129 S. Ct. at 1943–44. For a brief explanation of the facts of *Bivens v. Six Unknown Named Agents of the Federal Bureau of Narcotics*, 403 U.S. 388 (1971), and the related cause of action, see Cornelia T.L. Pillard, *Taking Fiction Seriously: The Strange Results of Public Officials’ Individual Liability under Bivens*, 88 GEO. L.J. 65, 68–72 (1999).

origin.⁴⁵ The complaint named John Ashcroft, former U.S. Attorney General, as the “principal architect” of the policy that resulted in Iqbal’s treatment, and Robert Mueller, Director of the FBI, as being “instrumental” in its adoption and execution.⁴⁶

In its decision to grant the defendants’ motion to dismiss, the Supreme Court relied heavily on the language of *Twombly* and scarcely, if at all, on the “no set of facts” language of *Conley*.⁴⁷ The Court shed light on the outcome of *Twombly* by illuminating a fundamental principle underlying the decision: “Only a complaint that states a plausible claim for relief survives a motion to dismiss.”⁴⁸ In fleshing out the plausibility requirement, the Court placed “plausible” along the likelihood spectrum between “possible” and “probable.”⁴⁹

The thunder of the Court’s *Iqbal* decision, however, came not from its in-depth explanation of plausibility under *Twombly*, but rather from its announcement of *Twombly*’s far-reaching application beyond section 1 of the Sherman Act. Specifically, the Court rejected Iqbal’s contention that the *Twombly* standard was limited to antitrust cases and held that the decision in *Twombly* “expounded the pleading standard for *all* civil actions.”⁵⁰ The Court reasoned that pursuant to Federal Rule of Civil Procedure 1,⁵¹ the *Twombly* interpretation of the pleading standard governing Rule 8 applies to all civil cases.⁵²

45. *Iqbal*, 129 S. Ct. at 1942. In particular, Iqbal claimed he was abused, denied food and medical treatment, unnecessarily stripped and body-cavity searched, and verbally accosted as a “terrorist” and “Muslim killer.” *Id.* at 1955.

46. *Id.* at 1944.

47. In fact, the Court mentioned *Conley v. Gibson* a total of only two times in its *Iqbal* opinion—once in reference to the petitioners’ argument, and once in reference to the circuit court’s evaluation of the district court’s test for a motion to dismiss. *Id.* at 1944, 1961. Whatever doubt *Twombly* left as to the retirement of *Conley*, *Iqbal* made it clear that the “no set of facts” regime was gone for good.

48. *Id.* at 1950.

49. *Id.* at 1949 (“The plausibility standard is not akin to a ‘probability requirement,’ but it asks for more than a sheer possibility that a defendant has acted unlawfully.”).

50. *Id.* at 1953 (emphasis added) (internal quotation marks omitted).

51. In relevant part, Rule 1 states that “[the Federal Rules of Civil Procedure] govern the procedure in all civil actions and proceedings in the United States district courts.” FED. R. CIV. P. 1.

52. See *Iqbal*, 129 S. Ct. at 1953 (stating that the standard announced in *Twombly* applies to “antitrust and discrimination suits alike”). While applying the plausibility standard in *Iqbal* as mandated by Rule 1 may seem to be the obvious reason for the Court’s holding, there may have been a subsurface motivation for dismissing the petitioners’ complaint. The Court addressed the issue of costs pertaining to government officials involved in costly litigation.

The Court's *Iqbal* analysis sets the stage for a heightened pleading standard in all civil actions, including patent infringement cases.⁵³ The following Section details the long history of pleading bad-faith infringement in patent actions to recover treble damages and discusses how the *Iqbal* standard will affect such bad-faith claims.⁵⁴

C. THE NATURE OF PATENT INFRINGEMENT AND THE HISTORY OF INCREASED DAMAGES

The United States Constitution sets forth the framework for a legal monopoly providing inventors the exclusive right to their discoveries for a limited period of time.⁵⁵ Consequently, any person who, “without authority, makes, uses, offers to sell, or sells any patented invention, within the United States or imports into the United States any patented invention during the term of the patent therefor, infringes the patent.”⁵⁶ Upon such infringement, the patentee is entitled to a civil remedy.⁵⁷ If defendants are found liable, relief may be granted in the form of an injunction⁵⁸ or damages.⁵⁹ As discussed below, damages may be increased based on the defendants’ conduct surrounding the infringement.

1. Patent Infringement as a Strict-Liability Offense

Patent infringement is a strict-liability tort.⁶⁰ To determine liability, courts need only construe the language of a patent’s claims and then compare it to

Id. Like *Twombly*, the Court noted that litigation can yield heavy financial burdens and can take away from government officials’ proper execution of their duties. *Id.*

53. 35 U.S.C. § 281 (2006) (“A patentee shall have remedy by civil action for infringement of his patent.”). Professors Clermont and Yeazell predict that under the Court’s new plausibility pleading standard, “patent cases . . . [will] cause experienced judges to shudder.” Clermont & Yeazell, *supra* note 27, at 836.

54. *See Jurgens v. CBK, Ltd.*, 80 F.3d 1566, 1570 (Fed. Cir. 1996) (“Increased damages also may be awarded to a party because of the bad faith of the other side.”).

55. U.S. CONST. art. I, § 8, cl. 8 (“To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.”).

56. 35 U.S.C. § 271(a) (2006).

57. *Id.* § 281.

58. *Id.* § 283.

59. *Id.* § 284.

60. *In re Seagate Tech., LLC*, 497 F.3d 1360, 1368 (Fed. Cir. 2007) (en banc) (“[P]atent infringement is a strict liability offense.”); *Hilton Davis Chem. Co. v. Warner-Jenkinson Co.*, 62 F.3d 1512, 1519, 1527 (Fed. Cir. 1995) (en banc) (“Intent is not an element of infringement. . . . Infringement is, and should remain, a strict liability offense.”), *rev’d on other grounds*, 520 U.S. 17 (1997); *see Blair v. Westinghouse Elec. Corp.*, 291 F. Supp. 664, 670

the allegedly infringing device.⁶¹ If, after comparison, the fact-finder determines that the defendant's device does not meet each and every claim limitation, there is no infringement.⁶² Thus, because the strict-liability nature of direct patent infringement requires no inquiry into the alleged infringer's intent,⁶³ and because claim construction is a matter of law,⁶⁴ the scope of discovery sought by patentee-plaintiffs largely relates to the nature of the infringement rather than to the simple finding of liability.⁶⁵

2. *The History of Increased Damages*

Congress has expressly authorized courts to increase damages for patent infringement. Such a provision first appeared in the Act of February 21, 1793.⁶⁶ Similarly, the Act of April 17, 1800, permitted a patentee to recover in an action at law "a sum equal to three times the actual damage sustained."⁶⁷ Not until the Act of July 4, 1836, however, were courts given the discretion to determine *when* increased damages should be awarded.⁶⁸ Under this legislation, courts could render judgments of increased damages "according to the circumstances of the case."⁶⁹

(D.D.C. 1968) ("[A]n infringement may be entirely inadvertent and unintentional and without knowledge of the patent.").

61. *See* Tate Access Floors, Inc. v. Interface Architectural Res., Inc., 279 F.3d 1357, 1365 (Fed. Cir. 2002) ("Infringement analysis involves two steps: the court first construes the scope of the asserted claims and then compares the accused device to the properly construed claims to determine whether each and every limitation of a claim is present, either literally or equivalently, in the accused device.").

62. *Dynacore Holdings Corp. v. U.S. Phillips Corp.*, 363 F.3d 1263, 1273 (Fed. Cir. 2004) ("To prove infringement, the patentee must show that the accused device meets each claim limitation, either literally or under the doctrine of equivalents.").

63. This Note is concerned only with direct patent infringement under 35 U.S.C. § 271(a). *See* Intel Corp. v. U.S. Int'l Trade Comm'n, 946 F.2d 821, 832 (Fed. Cir. 1991) ("[T]here is no intent element to *direct* infringement."). *Compare* 35 U.S.C. § 271(a) (2006) (lacking a mens rea for direct infringement), *with* § 271(c) (requiring a contributory infringer to aid the direct infringement knowingly).

64. *Markman v. Westview Instruments, Inc.*, 52 F.3d 967, 979 (Fed. Cir. 1995), *aff'd*, 517 U.S. 370 (1996) ("[T]he court has the power and obligation to construe as a matter of law the meaning of language used in the patent claim.").

65. *Seagate*, 497 F.3d at 1371-72 ("[I]t is indisputable that the proper legal standard for willful infringement informs the relevance of evidence relating to that issue and, more importantly here, the proper scope of discovery.").

66. Act of Feb. 21, 1793, ch. 11, sec. 5, 1 Stat. 318, 322 (allowing a patentee to recover "a sum, that shall be at least equal to three times the price, for which the patentee has usually sold or licensed to other persons, the use of [the invention]").

67. Act of Apr. 17, 1800, ch. 25, sec. 3, 2 Stat. 37, 38.

68. Act of July 4, 1836, ch. 357, sec. 14, 5 Stat. 117, 123.

69. *Id.*

Currently, 35 U.S.C. § 284 allows courts to “increase the damages up to three times the amount found or assessed.”⁷⁰ Despite the longstanding power to award increased damages, no statute has ever provided courts with any standard or guidelines for determining when to do so.⁷¹ Rather, these criteria have been left largely for courts to decide on a case-by-case basis.⁷²

D. OBJECTIVELY HIGHER: THE FEDERAL CIRCUIT’S NEW STANDARD OF WILLFULNESS

The Federal Circuit first crafted a standard for evaluating willfulness in *Underwater Devices Inc. v. Morrison-Knudsen Co.*⁷³ The court set forth the following criteria for evaluating willful infringement: “Where . . . a potential infringer has actual notice of another’s patent rights, he has an affirmative duty to exercise due care to determine whether or not he is infringing.”⁷⁴ Expounding upon this affirmative duty, the Federal Circuit indicated that an alleged infringer is considered “willful” if he is merely negligent.⁷⁵

The Supreme Court considered the meaning of “willfulness” in its 2007 decision *Safeco Insurance Co. of America v. Burr*,⁷⁶ a case governed by the Fair Credit Reporting Act.⁷⁷ The Court concluded that “reckless disregard” qualifies as a “willful violation.”⁷⁸ This finding was not surprising in light of the Court’s decision nearly two decades earlier in *McLaughlin v. Richmond Shoe Co.*⁷⁹ In *McLaughlin*, the Court held that “[t]he word ‘willful’ . . . is generally

70. 35 U.S.C. § 284 (2006).

71. *In re Seagate Tech., LLC*, 497 F.3d 1360, 1368 (Fed. Cir. 2007) (en banc) (noting that treble damages are awarded “[a]bsent a statutory guide” and stating that § 284 is “devoid of any standard for awarding [enhanced damages]”). This judicial discretion for determining when to award enhanced damages will prove to be the saving grace of treble damages after the *Iqbal* and *Seagate* decisions. See *infra* Section IV.C (recommending alternative criteria that courts can adopt for awarding increased damages).

72. See *Jurgens v. CBK, Ltd.*, 80 F.3d 1566, 1570 (Fed. Cir. 1996) (“[T]he court . . . exercising its sound discretion, [determines] whether, and to what extent, to increase the damages award given the totality of the circumstances.”).

73. 717 F.2d 1380, 1389–90 (Fed. Cir. 1983).

74. *Id.* at 1389.

75. *Seagate*, 497 F.3d at 1371 (noting that the standard of willfulness pursuant to *Underwater Devices* was “more akin to negligence”); see also Stephanie Pall, Note, *Willful Patent Infringement: Theoretically Sound? A Proposal to Restore Willful Infringement to Its Proper Place Within Patent Law*, 2006 U. ILL. L. REV. 659, 659 (stating that the Federal Circuit’s previous duty-of-care standard in *Underwater Devices* was a “concept[] mainly associated with negligence law”).

76. 551 U.S. 47 (2007).

77. 15 U.S.C. § 1681 (2006).

78. *Safeco*, 551 U.S. at 71.

79. 486 U.S. 128 (1988).

understood to refer to conduct that is not merely negligent.”⁸⁰ Thus, the negligence standard in *Underwater Devices* set the threshold for proving willful infringement too low.⁸¹ Following the language of the Supreme Court, the Federal Circuit expressly overruled *Underwater Devices* in *Seagate*: “Accordingly, we overrule the standard set out in *Underwater Devices* and hold that proof of willful infringement permitting enhanced damages requires at least a showing of objective recklessness.”⁸² Despite articulating the new standard, the Federal Circuit did nothing to elaborate on what conduct constitutes “objective recklessness” in the context of patent infringement.⁸³

After *Seagate*, to succeed on a claim of willful infringement, patentee–plaintiffs must demonstrate that the objectively defined risk of infringement, “determined by the record developed in the infringement proceeding,” was “either known or so obvious that it should have been known to the accused infringer.”⁸⁴ Therefore, the Federal Circuit set forth a framework in which, absent substantial knowledge about the actions of the alleged infringer prior to filing suit, the plaintiff must rely exclusively on discovery procedures to produce a record that demonstrates the defendant’s recklessness.⁸⁵ Thus, the new framework for proving willful conduct places a high premium on a plaintiff’s complaint getting to the discovery phase.⁸⁶

Herein lies the paradox presented by the combination of *Iqbal* and *Seagate*: patentee–plaintiffs must plead sufficient factual allegations to warrant discovery related to defendants’ willful conduct in accordance with *Iqbal*; however, as the Federal Circuit stated in *Seagate*, the factual foundation for

80. *Id.* at 133.

81. *See In re Seagate Tech., LLC*, 497 F.3d 1360, 1371 (Fed. Cir. 2007) (en banc) (“[The negligence] standard fails to comport with the general understanding of willfulness in the civil context.”).

82. *Id.* The court further delineated the standard by stating that an accused infringer acts with objective recklessness by proceeding “despite an objectively high likelihood that its actions constituted infringement of a valid patent.” *Id.*

83. *Id.* (“We leave it to future cases to further develop the application of this standard.”).

84. *Id.*

85. *See* Justin P. Huddleson, Note, *Objectively Reckless: A Semi-Empirical Evaluation of In re Seagate*, 15 B.U. J. SCI. & TECH. L. 102, 117–18 (2009) (“*Seagate* clearly raises the bar for proving willfulness . . . [and] limits a patentee’s ability to prove willful infringement through discovery.”).

86. *See* James E. Hopenfeld, *A Proposal for a “Good-Faith Offer” Standard for Evaluating Allegations of Willful Infringement—With Thanks to Major League Baseball*, 20 FED. CIR. B.J. 5, 22 (2010) (discussing how the current version of the Patent Reform Act incorporates the *Seagate* standard of willfulness, and stating that under the current version of the Act, one of the most important limitations is that patentees “must plead willful infringement with particularity”).

demonstrating recklessness will often be established by the discovery proceedings themselves.⁸⁷ In other words, plaintiffs are expected to sufficiently plead allegations of willfulness to warrant discovery related to defendants' conduct while they are simultaneously allowed—and indeed, *expected*—to rely on the discovery proceedings themselves to expose the alleged willful conduct.

III. GETTING COMPLAINTS THROUGH THE (SEA)GATE: APPLYING *IQBAL* TO WILLFUL INFRINGEMENT

The language of the Supreme Court's decision in *Iqbal*—much of which is taken from *Twombly*—lays the groundwork for a new plausibility regime in civil pleadings under Federal Rule of Civil Procedure 8. The Court's standard will serve as the gatekeeper for all claims to pass through on their way to discovery, dispositive motions, and ultimately, trial. Since *Iqbal* adopts the *Twombly* standard, it is important to understand the exact criteria that Justice Souter established in *Twombly*. Section III.A will revisit the language of the *Twombly* opinion (and consequently, the *Iqbal* standard), and attempt to reconcile the Court's inconsistent language, which seems hesitant about retiring *Conley*.⁸⁸ Section III.B will then consider three ways in which patentee-plaintiffs can use the Federal Rules of Civil Procedure to overcome the high *Iqbal* pleading standard and satisfy the Court's plausibility threshold.

A. POSSIBILITY, PLAUSIBILITY, PROBABILITY: WHAT IS THE ACTUAL STANDARD?

The wave of confusion and uncertainty that resulted from the *Twombly* decision was not entirely unexpected in light of Justice Souter's inconsistent

87. The Federal Circuit indicated in *Seagate* that it could proceed on the issue of infringement alone without the same scope of discovery that willfulness allegations required:

The ultimate dispute in this case is the proper scope of discovery. While it is true that the issue of willful infringement, or even infringement for that matter, has not been decided by the trial court, it is indisputable that the proper legal standard for willful infringement informs the relevance of evidence relating to that issue and, more importantly here, the proper scope of discovery.

Seagate, 497 F.3d at 1371–72.

88. In his *Twombly* dissent, Justice Stevens appears to accuse the majority of putting the *Conley* “no set of facts” language to rest with a sense of melancholy and admiration. *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 577 (2007) (Stevens, J., dissenting) (“If *Conley*’s ‘no set of facts’ language is to be interred, let it not be without a eulogy.”).

opinion.⁸⁹ Moreover, while *Iqbal* stated that the *Twombly* standard applies to all civil actions, it did very little to explain “plausibility” in a way that could be understood without relying—almost entirely—on the language of *Twombly* itself.⁹⁰ Thus, absent a clarification of the Court’s intent, plaintiffs and judges are left without any explanation regarding an apparently complete overhaul⁹¹ to one of the cornerstones of the civil justice system.⁹²

As a point of departure for analyzing the language of *Iqbal*, it is significant to note that the floor for pleading requirements under the *Conley* standard is drastically—and objectively—different from that set forth in *Twombly*. A literal reading of *Conley* establishes that a complaint should not be dismissed “whenever the pleadings [leave] open the *possibility* that a plaintiff might later establish some set of [undisclosed] facts to support recovery.”⁹³ In other words, claims should not be dismissed unless they are shown to be factually impossible.⁹⁴ In contrast, *Twombly* made it clear that pleadings must set forth more than just mere possibility of recovery: allegations that “stop[] short of the line between possibility and plausibility” should be dismissed.⁹⁵ While this language may suggest that the *Twombly–Iqbal* plausibility standard is significantly more stringent than the *Conley* threshold (as this Note contends), the Court’s explicit language calls that conclusion into question. For example, in dismissing the plaintiff’s complaint, the *Twombly* majority

89. See *Iqbal v. Hasty*, 490 F.3d 143, 155 (2d Cir. 2008) (“[T]he [*Twombly*] Court’s explanation contains several, not entirely consistent, signals . . .”), *rev’d sub nom.* *Ashcroft v. Iqbal*, 129 S. Ct. 1937 (2009); see also Benjamin W. Cheesbro, Note, *A Pirate’s Treasure?: Heightened Pleadings Standards for Copyright Infringement Complaints After Bell Atlantic Corp. v. Twombly*, 16 J. INTELL. PROP. L. 241, 248 (2009) (“Some language indicates that a new and ‘heightened’ pleading standard was intended, while other language explicitly or implicitly rejects that notion.”).

90. In his opinion in *Ashcroft v. Iqbal*, Justice Kennedy cited *Twombly* over two dozen times.

91. Arthur Miller, a civil procedure scholar and professor at New York University School of Law, stated the following: “I have spent my entire life with the Federal Rules of Civil Procedure, and I firmly believe that [*Twombly* and *Iqbal*] represent a philosophical sea of change in American civil litigation.” *Access to Justice Denied: Ashcroft v. Iqbal: Hearing Before the Subcomm. on the Constitution, Civil Rights, and Civil Liberties of the H. Comm. on the Judiciary*, 111th Cong. 6 (2009) (testimony of Arthur R. Miller), available at http://judiciary.house.gov/hearings/printers/111th/111-36_53090.PDF.

92. See Spencer, *supra* note 18, at 447 (framing the *Conley* “no set of facts” standard as the “fifty-year-old statement providing the bedrock understanding of the general pleading standard in our system”).

93. *Twombly*, 550 U.S. at 561 (emphasis added) (internal quotation marks omitted).

94. See *id.* (“This ‘no set of facts’ language can be read in isolation as saying that any statement revealing the theory of the claim will suffice unless its factual impossibility may be shown from the face of the pleadings . . .”).

95. *Id.* at 557.

averred that the Court was “not requir[ing] heightened fact pleading of specifics.”⁹⁶

Furthermore, while one can argue that what was really rejected in *Twombly* was the “no set of facts” language, *Twombly* nevertheless expressly held that stating a plausible claim “requires a complaint [to have] *enough* factual matter (taken as true) to suggest an agreement was made” that would violate section 1 of the Sherman Act.⁹⁷ Hence, there appears to be a distinct discrepancy between “short and plain statement” and “enough factual matter” such that the two standards require different levels of specificity. Consequently, the broad application of *Iqbal* affirmatively raises the bar for pleading.⁹⁸

B. WHAT ABOUT THE RULES?

The *Twombly* opinion was widely criticized for failing to clarify the extent to which the Court’s new plausibility standard would apply to other areas of the law beyond antitrust litigation.⁹⁹ However, now that the Court has answered the question posed by Justice Stevens in his *Twombly* dissent, the legal community has shifted focus to asking whether the *Iqbal* standard of plausibility comports with the purpose and meaning of the Federal Rules of Civil Procedure.¹⁰⁰ This Section will examine three approaches to circumventing *Iqbal*’s plausibility standard in an effort to maintain—at least as closely as possible—patentees’ short and plain statements as sufficient to open the gates of discovery related to willful infringement.

96. *Id.* at 570.

97. *Id.* at 556 (emphasis added).

98. See Spencer, *supra* note 32, at 4 (“Although . . . the Court itself would perhaps deny it, *Twombly* appeared to be a departure from the simple ‘notice’ pleading standard announced in *Conley v. Gibson* . . .”).

99. See, e.g., sources cited *supra* note 32.

100. See Kenneth S. Klein, *Ashcroft v. Iqbal Crashes Rule 8 Pleadings Standards on to Unconstitutional Shores*, 88 NEB. L. REV. 261, 262 (2009) (“[T]he [*Iqbal*] decision inescapably interprets *Federal Rule of Civil Procedure* 8 in a manner that is unconstitutional.”); Michelle Spiegel, Comment, *Ashcroft v. Iqbal: The Question of a Heightened Standard of Pleading in Qualified Immunity Cases*, 4 DUKE J. CONST. L. & PUB. POL’Y SIDEBAR 375, 383 (2009) (finding the plausibility standard to be a distinct departure from the “short, plain statement” interpretation of the Federal Rules). Even before *Iqbal* was granted certiorari, members of the Court already shared the sentiment that *Twombly* may “rewrite the Nation’s civil procedure textbooks and call into doubt the pleading rules of most of its States.” *Twombly*, 550 U.S. at 579 (Stevens, J., dissenting).

1. *The Sample Complaint of Form 18*

The Federal Rules of Civil Procedure provide model complaints for various causes of action.¹⁰¹ Of particular importance is Form 18—the Complaint for Patent Infringement.¹⁰² Form 18 sets forth the basic elements of a cause of action for patent infringement: (1) the plaintiff owned a patent throughout the defendant’s infringing period and still owns the patent; (2) the defendant infringed the plaintiff’s patent and is still infringing the patent by making, selling, or using the patented invention; and (3) the plaintiff complied with the statutory requirement of placing notice of the patent on the patented invention and gave the defendant written notice of the infringement.¹⁰³ Form 18 even provides a formulaic recitation of the harm the plaintiff seeks to remedy and sets forth relief in the form of a preliminary and final injunction against the continuing infringement, and accounts for damages, interests, and costs.¹⁰⁴ However, Form 18 fails to go so far as to set forth the elements of willful infringement that would allow for an increase in damages. Because treble damages are codified in the Patent Act¹⁰⁵ and allegations of willfulness have become a routine element of infringement complaints,¹⁰⁶ Form 18, in its current form, inadequately reflects the current state of pleading in patent litigation.

The Federal Circuit provided very little guidance to potential patentee-plaintiffs as to what the elements of willfulness are in the post-*Seagate* era. While *Seagate* itself may have been significant for altering the standard from one of negligence to one of recklessness, the opinion evaded the task of laying down a clear set of instructions for what actually constitutes reckless

101. FED. R. CIV. P. app. B; *see also* *Conley v. Gibson*, 355 U.S. 41, 47 (1957) (“The illustrative forms appended to the Rules plainly demonstrate [the pleading requirements].”).

102. FED. R. CIV. P. form 18.

103. *Id.*

104. *Id.*

105. 35 U.S.C. § 284 (2006) (allowing the court to increase damages up to three times the amount assessed).

106. *See The Fourteenth Annual Judicial Conference for the United States Court of Appeals for the Federal Circuit*, 170 F.R.D. 534, 615 (1996) [hereinafter *Judicial Conference*] (citing commentary from one patent attorney who stated: “I think if you ask most District Court Judges, they think that willful infringement is one word. I have never heard of an infringer called anything but willful.”); Danny Prati, Note, *In re Seagate Technology, LLC: A Clean Slate for Willfulness*, 23 BERKELEY TECH. L.J. 47, 52–53 n.42 (2008) (citing Kimberly A. Moore, *Empirical Statistics on Willful Patent Infringement*, 14 FED. CIR. B.J. 227, 232 (2004) (finding that between 1999 and 2000, plaintiffs pled willfulness in 92.3% of patent infringement cases)); Jon E. Wright, *Willful Patent Infringement and Enhanced Damages—Evolution and Analysis*, 10 GEO. MASON L. REV. 97, 97 (2001) (“A charge of willful infringement has become a routine adjunct to almost every pleading asserting patent infringement.”).

conduct.¹⁰⁷ Thus, patentees are left without a guidepost for determining what elements need to be pled to constitute willful conduct, let alone the specificity with which they need to be pled under *Iqbal* to open the doors of discovery. One solution to this problem may be for the Federal Circuit to explicitly address what the elements of objective recklessness are in the patent infringement context beyond any fundamental commentary the courts previously relied upon when defining “reckless.”¹⁰⁸ However, incorporating any necessary elements into Form 18 is a daunting task since the Federal Circuit itself “fully recognize[s] that the term [‘reckless’] is not self-defining.”¹⁰⁹

In a patentee’s utopia, Form 18 would include specific language regarding what must be pled in a willfulness allegation. According to Rule 84, “[t]he forms in the Appendix suffice under these rules and illustrate the simplicity and brevity that these rules contemplate.”¹¹⁰ Thus, any guidelines set forth in the form should technically be sufficient to open the gates of discovery. However, *Twombly* and *Iqbal* may say otherwise. On their face, the allegations in Form 18 appear as only facts that are “merely consistent with a defendant’s liability” and are a far cry from “contain[ing] sufficient factual matter” in accordance with *Twombly*.¹¹¹ Under the plausibility regime, such statements do not “nudge[] . . . claims . . . across the line from conceivable to plausible.”¹¹²

The Federal Circuit recently reinforced this proposition in *Colida v. Nokia, Inc.*,¹¹³ where the plaintiff alleged that Nokia infringed four design patents for cellular telephones.¹¹⁴ In a per curiam opinion affirming the Southern District of New York’s dismissal of the plaintiff’s complaint, the

107. *In re Seagate Tech., LLC*, 497 F.3d 1360, 1371 (Fed. Cir. 2007) (en banc) (“We leave it to future cases to further develop the application of [the] [objective recklessness] standard.”).

108. At common law, “recklessness” was defined as engaging in an “action entailing an unjustifiably high risk of harm that is either known or so obvious that it should be known.” *Safeco Ins. Co. of Am. v. Burr*, 551 U.S. 47, 68 (2007) (quoting *Farmer v. Brennan*, 511 U.S. 825, 836 (1994) (internal quotation marks omitted)).

109. *Seagate*, 497 F.3d at 1371 (internal quotation marks omitted) (citing *Farmer*, 511 U.S. at 836).

110. FED. R. CIV. P. 84.

111. *Ashcroft v. Iqbal*, 129 S. Ct. 1937, 1949 (2009).

112. *Id.* at 1951 (citing *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 570 (2007)).

113. 347 F. App’x 568 (Fed. Cir. 2009).

114. *Id.* at 569.

Federal Circuit cited both *Iqbal*¹¹⁵ and *Twombly*¹¹⁶ and stated that the plaintiff's infringement claims were "facially implausible."¹¹⁷ In a footnote, the court acknowledged Form 18 but ultimately rejected its sufficiency.¹¹⁸ In doing so, the court stated: "Form 18 is a sample pleading for patent infringement, but . . . was last updated before the Supreme Court's *Iqbal* decision."¹¹⁹ Therefore, absent new amendments to Form 18 that would instruct patentee-plaintiffs on how to provide more specificity in their complaints, *Iqbal* has stripped Form 18 and Rule 84 of any practical significance.

2. Liberal Constructions Under Rule 8

To the extent that Federal Rule of Civil Procedure 8 maintains a pulse after its most famous phrase was "retired" by *Twombly*¹²⁰ and then exiled in its entirety by *Iqbal*, Rule 8(e) may provide a safe harbor for plaintiffs whose allegations are somewhere on the likelihood spectrum between "possible" and "probable" (wherever "plausible" may actually be).¹²¹ Federal Rule of Civil Procedure 8(e) states that "[p]leadings must be construed so as to do justice."¹²² Therefore, a court that is uncertain how to apply the plausibility standard in light of the Supreme Court's language¹²³ may be lenient with plaintiffs' complaints that the court believes come close to satisfying the *Iqbal* threshold.¹²⁴

115. *Id.* at 570 (stating that a "complaint must have sufficient 'facial plausibility' to 'allow[] the court to draw the reasonable inference that the defendant is liable'" (citing *Iqbal*, 129 S. Ct. at 1949)).

116. *Id.* (stating that "[t]he plaintiff's factual allegations must 'raise a right to relief above the speculative level' and cross 'the line from conceivable to plausible'" (citing *Twombly*, 550 U.S. at 555, 570)).

117. *Twombly*, 550 U.S. at 570.

118. *Id.* at 575 n.2. *But see* Automated Transactions, LLC v. First Niagara Fin. Grp., Inc., No. 10-cv-00407 (A)(M), slip op. at 13 (W.D.N.Y. Aug. 31, 2010) (holding that willfulness allegations suffice under Rule 84 notwithstanding *Twombly* and *Iqbal*).

119. *Colida*, 347 F. App'x at 571 n.2.

120. *Twombly*, 550 U.S. at 562–63 (2007) (stating that the "no set of facts" language in *Conley v. Gibson*, 355 U.S. 41, 45 (1957), earned its "retirement" due to being "questioned, criticized, and explained away long enough" over the last fifty years).

121. *Ashcroft v. Iqbal*, 129 S. Ct. 1937, 1949 (2009). For discussion of the placement of "possibility," "plausibility," and "probability" along the Court's likelihood spectrum, see *supra* Section III.A.

122. FED. R. CIV. P. 8(e).

123. See sources cited *supra* note 32 (discussing the confusion and uncertainty resulting from the language of the Court's *Twombly* opinion).

124. See *Maldonado v. Fontanes*, 568 F.3d 263, 268 (1st Cir. 2009) (advocating that courts "draw on [their] judicial experience[s] and common sense" when reviewing complaints (quoting *Twombly*, 550 U.S. at 556) (quotation marks omitted)); cf. Arthur R. Miller, *From Conley to Twombly to Iqbal: A Double Play on the Federal Rules of Civil Procedure*, 60

For example, even under the more stringent plausibility standard in pleading, the Federal Circuit has already been relatively lenient with plaintiffs in the post-*Twombly* era. In *McZeal v. Sprint Nextel Corp.*,¹²⁵ the Federal Circuit allowed pro se plaintiff Alfred McZeal, Jr.'s claim to survive the defendants' motion to dismiss.¹²⁶ McZeal had alleged infringement of his utility patent related to cellular telephone technology.¹²⁷ After the U.S. District Court for the Southern District of Texas dismissed his claim for failure to plead sufficient allegations, McZeal appealed to the Federal Circuit for a review of his ninety-five page complaint—not including exhibits.¹²⁸ Writing for the court, Judge Archer determined that the pro se plaintiff should be “grant[ed] . . . leeway on procedural matters, such as pleading requirements.”¹²⁹

While the court's decision in *McZeal* was influenced by the fact that the plaintiff was appearing pro se, the court nevertheless went out of its way to state that McZeal was “no stranger to legal proceedings, having filed numerous complaints in the past.”¹³⁰ Therefore, a holding that reversed the district court's Rule 12(b)(6) dismissal indicates that the Federal Circuit may be lenient in construing pleadings in accordance with Rule 8. Thus, despite a higher standard for pleadings under *Iqbal*, construing pleadings liberally under Rule 8(e) may inject just enough support to keep the *Conley* “no set of facts” standard alive.

3. *Amending Pursuant to Rule 15*

Federal Rule of Civil Procedure 15 allows plaintiffs to amend their pleadings after the initial complaint has already been filed with the court and served upon the defendant.¹³¹ To do so, Rule 15 requires either one of two actions to occur: (1) the opposing party grants the plaintiff written permission to amend, or (2) the court grants the plaintiff leave to amend.¹³²

DUKE L.J. 1, 29 n.107 (2010) (citing cases where “courts [have been] more lenient than others in allowing a claim to go forward even if there is an alternative explanation”).

125. 501 F.3d 1354 (Fed. Cir. 2007).

126. *Id.* at 1358 (vacating the trial court's dismissal and remanding for further proceedings).

127. *Id.* at 1357.

128. *Id.* at 1355.

129. *Id.* at 1356 (referring to a lesser procedural standard for pro se litigants in *Hughes v. Rowe*, 449 U.S. 5, 9 (1980)).

130. *Id.* at 1358.

131. FED. R. CIV. P. 15.

132. FED. R. CIV. P. 15(a)(2). In limiting a plaintiff to these two actions for amending, I am of course assuming that the twenty-one-day period after the complaint has been served

While it is unlikely that a defendant staring down the explosive barrel of potentially expensive patent litigation¹³³—in addition to tens of millions of dollars in punitive damages if the court finds willfulness¹³⁴—will grant a plaintiff permission to amend the complaint to be more specific about the allegedly willful conduct, Rule 15 provides that “court[s] should freely give leave [to amend] when justice so requires.”¹³⁵ Thus, the Rules may provide a plaintiff a chance to amend the complaint to set forth and describe a course of conduct that rises to the level of objective recklessness without being defeated by *Twombly*–*Iqbal* plausibility.¹³⁶

However, the leave to amend a complaint, absent gathering any additional information between the time of filing and the time of amending, may be futile in light of *Iqbal*. Under the Court’s plausibility standard, plaintiffs will likely set forth all facts known at the time of filing the complaint to ensure that their claims rise beyond a level of being merely “conceivable.”¹³⁷ The implication of this is that a plaintiff will likely show all his cards to the defendant in the initial complaint rather than choosing to only set forth the minimal detail sufficient to get to the discovery phase.¹³⁸

in which the plaintiff can amend the complaint as a matter of course has lapsed. See FED. R. CIV. P. 15(a)(1).

133. See Christopher A. Harkins, *A Budding Theory of Willful Patent Infringement: Orange Books, Colored Pills, and Greener Verdicts*, 2007 DUKE L. & TECH. REV. 6, *7 (discussing a “2007 survey published by the American Intellectual Property Law Association [that] affirms what patent owners and attorneys knew all along: patent litigation is expensive”); Joseph Scott Miller, *Building a Better Bounty: Litigation-Stage Rewards for Defeating Patents*, 19 BERKELEY TECH. L.J. 667, 703 (2004) (“Patent litigation is, after all, quite expensive.”).

134. See *infra* Section IV.A (discussing the high compensatory damages imposed on defendants when a court finds liability for infringement); see also, e.g., *i4i Ltd. P’ship v. Microsoft Corp.*, 598 F.3d 831, 858–59 (Fed. Cir. 2010) (affirming an award of \$40 million in enhanced damages due to willful infringement).

135. FED. R. CIV. P. 15.

136. See Patricia W. Hatamyar, *The Tao of Pleading: Do Twombly and Iqbal Matter Empirically?*, 59 AM. U. L. REV. 553, 598–600 & tbl.1 (2010) (presenting evidence from a study that revealed that the number of Rule 12(b)(6) dismissals granted with leave to amend increased from six percent under *Conley* to nineteen percent under *Iqbal*, and speculating that “the newness of the *Iqbal* standard [has] caused [district courts] to err on the side of allowing the plaintiff one more chance to plead”).

137. *Ashcroft v. Iqbal*, 129 S. Ct. 1937, 1951 (2009) (citing *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 570 (2007)); Kendall W. Hannon, Note, *Much Ado About Twombly? A Study on the Impact of Bell Atlantic Corp. v. Twombly on 12(b)(6) Motions*, 83 NOTRE DAME L. REV. 1811, 1840 (2008) (noting the “dramatic and readily apparent impact” that the heightened standard will have with respect to the length of complaints for some types of cases).

138. See Hon. Colleen McMahon, *The Law of Unintended Consequences: Shockwaves in the Lower Courts after Bell Atlantic Corp. v. Twombly*, 41 SUFFOLK U. L. REV. 851, 868 (2008)

This tendency to file much more detailed complaints harkens back to archaic code pleading¹³⁹ and departs from notice pleading as intended by the Federal Rules and enforced in *Conley*.¹⁴⁰ For instance, recall that even McZeal's ninety-five-page complaint was initially dismissed by the Southern District of Texas for failure to plead sufficient allegations.¹⁴¹

Notwithstanding, while Rule 15 may allow courts to grant leave in the interest of justice, Rule 15 does not provide a patentee–plaintiff with a sneak peek into a defendant's potentially discoverable documents that may uncover indicia of willful infringement before filing an amended complaint. In fact, the Federal Circuit was clear in *Seagate* that willfulness is a function of the discovery record: “[i]f this threshold objective standard is satisfied, the patentee must also demonstrate that this objectively-defined risk (*determined by the record developed in the infringement proceeding*) was either known or so obvious that it should have been known to the accused infringer.”¹⁴² Therefore, unless patentees withhold material information related to willful conduct from their initial complaints (a strategy highly unlikely after *Iqbal*), the leave to amend will impart very little benefit to patentees if they have no new information to plead. Thus, the safe-harbor provision that allows plaintiffs to amend pleadings may provide little comfort to a plaintiff nakedly “armed with nothing more than conclusions.”¹⁴³

IV. THE *IQBAL* AFTERMATH: DAMAGES, DICHOTOMY, AND THE DESERTION OF WILLFULNESS

Iqbal's effects on treble damages in patent cases are yet to be determined. However, because plaintiffs allege willfulness in approximately ninety percent

(noting how district courts will have to spend more time deciding Rule 12(b)(6) motions due to the increase in the complexity of complaints).

139. For a general discussion of code pleading prior to the adoption of notice pleadings, see Koan Mercer, Comment, “*Even in These Days of Notice Pleadings*”: *Factual Pleading Requirements in the Fourth Circuit*, 82 N.C. L. REV. 1167, 1168–70 (2004) (noting how the code pleadings required specific “facts constituting the cause of action”).

140. See Scott Dodson, *Comparative Convergences in Pleading Standards*, 158 U. PA. L. REV. 441, 447–49 (2010) (discussing how the Federal Rules of Civil Procedure were a response to the dissatisfaction associated with code pleading).

141. *McZeal v. Nextel Sprint Corp.*, 501 F.3d 1354, 1355 (Fed. Cir. 2007).

142. *In re Seagate Tech., LLC*, 497 F.3d 1360, 1371 (Fed. Cir. 2007) (en banc) (emphasis added).

143. *Ashcroft v. Iqbal*, 129 S. Ct. 1937, 1950 (2009).

of their complaints,¹⁴⁴ it seems inevitable that the standard for obtaining discovery orders pertaining to willfulness allegations will in some way be affected by a heightened pleading threshold. Moreover, proceeding past the dismissal phase of a lawsuit without the fear of enhanced damages may make defendants more willing to fight the allegations since the risk of tripled liability¹⁴⁵ has been taken out of the equation. Section IV.A will examine the current state of damages in patent cases and the role that willfulness plays in pursuing such claims. Section IV.B will then examine bad-faith conduct in two other areas of intellectual property law. First, Section IV.B.1 overviews the willfulness standard in copyright law. Second, Section IV.B.2 compares the standard of liability for inequitable conduct, as defined by the U.S. Patent and Trademark Office (PTO), to that of willful patent infringement and the different Federal Rules applicable to each. Section IV.C then suggests how patentees may overcome *Twombly–Iqbal* plausibility and contemplates what effects eliminating willfulness will have on patent litigation and whether treble damages will soon be interred alongside the *Conley* “no set of facts” regime.

A. DAMAGES: TREBLE DAMAGES AND THE INCENTIVE TO SUE

Enhanced damages in patent cases can be extremely costly for defendants and handsomely rewarding for plaintiffs.¹⁴⁶ In fact, even without the assessment of multiplied damages, litigation alone can be rather expensive for many companies seeking to either enforce their patent rights or defend against an infringement allegation.¹⁴⁷ Thus, it is important to evaluate

144. See *Judicial Conference*, *supra* note 106, at 615 (citing Judge Sue Robinson of the District of Delaware, who stated that willfulness is pled in eighty-seven percent of her pending patent infringement cases); Prati, *supra* note 106, at 52 n.42 (referring to Kimberly Moore’s study that found that willfulness is alleged in 92.3% of patent infringement cases).

145. See 35 U.S.C. § 284 (2006) (allowing a court to increase damages up to three times the amount assessed).

146. See, e.g., Harkins, *supra* note 133, at *7 (“Allegations of willful patent infringement frequently take center stage in patent litigation, offering treble damages and attorneys’ fees to patent owners eager to turn actual damages into a windfall.”); Andrew M. Newton, *Encouraging Willful Infringement? Knorr-Bremse Leaves Due Care in Patent Litigation in a State of Flux*, 15 TEX. INTEL. PROP. L.J. 91, 111–12 (2006) (noting that treble damages can rise “well into the tens of millions of dollars, or even much higher”); Shashank Upadhye, *Understanding Willfulness in Patent Infringement: An Analysis of the “Advice of Counsel” Defense*, 8 TEX. INTEL. PROP. L.J. 39, 40 (1999) (“A large damages award can often bankrupt a company.”).

147. See, e.g., John M. Golden, Comment, “Patent Trolls” and Patent Remedies, 85 TEX. L. REV. 2111, 2128 (2007) (“[A] party’s expected litigation costs are substantial—generally at least on the order of \$1 million.”); David O. Taylor, *Wasting Resources: Reinventing the Scope of Waiver Resulting from the Advice-of-Counsel Defense to a Charge of Willful Patent Infringement*, 12 TEX.

the role that treble damages play in plaintiffs' incentives to sue and how those incentives may be influenced by a higher standard for pleading willfulness.¹⁴⁸

Although actual compensatory damages in patent infringement cases can escalate well into the hundreds of millions of dollars,¹⁴⁹ patentees are often faced with the economic dilemma of whether suing to enforce their rights will accrue any substantial benefit when balanced against the costs of litigation and the likelihood of success.¹⁵⁰ One empirical study shows that patentees prevail in just over half of patent infringement cases that go to trial.¹⁵¹ In light of such data, compensatory damages alone may not provide enough incentive to undergo potentially costly litigation. Additionally, the patentee risks paying for the defendant's attorney's fees¹⁵² if the defendant is not found liable, and if the court determines that the patentee purposefully prolonged the suit in bad faith to fish for a settlement agreement.¹⁵³ Furthermore, patentees must consider that the court may find inequitable

INTELL. PROP. L.J. 319, 322 (2004) ([T]he cost of bringing a patent infringement case, let alone defending one, is . . . substantial."); Wright, *supra* note 106, at 97 ("[P]atent infringement lawsuits are enormously expensive and often swamp balance sheets." (internal citation and quotation marks omitted)).

148. See Jonathan T. Molot, *How U.S. Procedure Skeins Tort Law Incentives*, 73 IND. L.J. 59, 70 (1997) ("For a plaintiff, the decision whether to file a suit requires a weighing of the expected benefits and anticipated expenses of litigation.").

149. Taylor, *supra* note 147, at 321–22 ("Compensatory damages in patent cases often amount to more than ten million dollars and sometimes rise into the hundreds of millions of dollars.").

150. See Nathaniel C. Love, Comment, *Nominal Reasonable Royalties for Patent Infringement*, 75 U. CHI. L. REV. 1749, 1772 (2008) (noting that where "no significant damage award can be expected or an injunction has low value against a non-competitor that can adopt an alternative," a patentee may have "insufficient incentives to file suit").

151. Kimberly A. Moore, *Judges, Juries, and Patent Cases: An Empirical Peek Inside the Black Box*, 99 MICH. L. REV. 365, 385 & fig.2 (2000) (finding that patentees prevailed in fifty-eight percent of the lawsuits that proceeded to the trial stage between 1999 and 2000).

152. 35 U.S.C. § 285 (2006) (allowing the court in exceptional cases to award reasonable attorneys fees to the prevailing party in patent cases); see also *Phonometrics, Inc. v. Westin Hotel Co.*, 350 F.3d 1242, 1246 (Fed. Cir. 2003) ("[A]ny attorney . . . who so multiplies the proceedings in any case unreasonably and vexatiously may be required by the court to satisfy personally the excess costs, expenses, and attorneys' fees reasonably incurred because of such conduct." (citing 28 U.S.C. § 1927 (2003))); Thomas F. Cotter, *An Economic Analysis of Enhanced Damages and Attorney's Fees for Willful Patent Infringement*, 14 FED. CIR. B.J. 291, 292 n.3 (2004) ("Defendants are sometimes entitled to an award of their attorney's fees, if the court determines that the plaintiff filed or litigated the suit in bad faith.").

153. See sources cited *supra* note 5 (discussing abusive litigation tactics used by plaintiffs to force a defendant into a settlement agreement).

conduct during the prosecution process.¹⁵⁴ Stated best: “Patent holders and litigators must decide whether asserting patent rights against an infringer outweighs the financial cost of defending against an inequitable conduct challenge and the risk of losing the entire patent should the defendant uncover some evidence a court finds to be inequitable conduct.”¹⁵⁵ This is a significant and costly deterrent to enforcing one’s monopoly rights and one that potential patentee–plaintiffs must contemplate.¹⁵⁶

Moreover, a patentee’s hesitance to proceed with a suit may be magnified if the outcome is determined by a judge rather than jury. As then-professor Kimberly Moore (now a Federal Circuit judge) noted, “[t]he identity of the adjudicator . . . is a statistically significant predictor of who wins the claims in the lawsuit.”¹⁵⁷ Specifically, Judge Moore conducted a study that found that patentee–plaintiffs have a seventeen-percent-lower win rate in cases tried before judges compared to juries.¹⁵⁸ Therefore, absent the anticipation of enhanced damages due to willfulness, patentees may be less likely to file suit due to the risk that the court will find the defendant not liable.¹⁵⁹

B. DICHOTOMY: COMPARING WILLFUL PATENT INFRINGEMENT TO OTHER BAD-FAITH INFRACTIONS

Increasing liability based on willful conduct is not new to intellectual property law. For instance, federal copyright law allows plaintiffs to collect enhanced damages if “the court finds[] that infringement was committed willfully.”¹⁶⁰ And analogous to *Seagate*, copyright law already requires a finding of “reckless disregard” as the threshold for obtaining enhanced

154. See *infra* Section IV.B.2 (detailing and explaining inequitable conduct during patent prosecution and the ramifications for plaintiffs).

155. Kate McElhone, *Inequitable Conduct: Shifting Standards for Patent Applicants, Prosecutors, and Litigators*, 17 TEX. INTELL. PROP. L.J. 385, 388 (2009).

156. See *Kingsdown Med. Consultants, Ltd. v. Hollister, Inc.*, 863 F.2d 867, 876 n.15 (Fed. Cir. 1988) (“[Th]e habit of charging inequitable conduct in almost every major patent case has become an absolute plague.” (quoting *Burlington Indus., Inc. v. Dayco Corp.*, 849 F.2d 1418, 1422 (Fed. Cir. 1988))).

157. Moore, *supra* note 151, at 386–87.

158. *Id.* at 386 tbl.2 (finding only a fifty-one percent win rate for patentees when the case is decided by a judge, and a sixty-eight percent win rate for patentees when the case is decided by a jury).

159. See Harkins, *supra* note 133, at *1–2 (“From a patent owner’s perspective, the possibility of recovering treble damages and attorneys’ fees may tip the scales in favor of enforcing its patent in an infringement lawsuit.”).

160. 17 U.S.C. § 504(c)(2) (2006).

damages.¹⁶¹ Additionally, many states' trade secret laws award plaintiffs increased damages for willful misappropriation. For example, section 550 of the Iowa Code awards plaintiffs "exemplary damages" where the court finds that the defendant "commits a willful and malicious misappropriation."¹⁶² Similarly, transferring liability for bad-faith conduct is also present in federal intellectual property law. For example, defendants in patent infringement cases may seek to recover attorney's fees from the plaintiff under a theory that the plaintiff engaged in inequitable conduct during the patent prosecution process, thus leading to frivolous litigation over an invalid patent.¹⁶³

1. *Copyright Law's Recklessness Standard*

For over a decade before the Federal Circuit decided *Seagate*, courts were already employing recklessness as the standard for proving willfulness in copyright infringement actions.¹⁶⁴ Thus, it is not unreasonable to believe that, even absent the Supreme Court's clarification of the meaning of "willful" in *Safeco Insurance Co. of America v. Burr*,¹⁶⁵ it was only a matter of time until the Federal Circuit established a more stringent test for patentees seeking treble damages.¹⁶⁶ For example, in *Island Software and Computer Service, Inc. v. Microsoft Corp.*,¹⁶⁷ the Second Circuit stated that one way for a plaintiff to prove willful copyright infringement is to demonstrate that "the defendant's actions were the result of 'reckless disregard' for, or 'willful blindness' to, the copyright

161. *Yurman Design, Inc. v. PAJ, Inc.*, 262 F.3d 101, 112 (2d Cir. 2001) ("Willfulness . . . means that the defendant recklessly disregarded the possibility that its conduct represented infringement." (internal citation and quotation marks omitted)).

162. IOWA CODE § 550.4 (2009).

163. *See* *Epcon Gas Sys., Inc. v. Bauer Compressors, Inc.*, 279 F.3d 1022, 1034 (Fed. Cir. 2002) (quoting 35 U.S.C. § 285 and including "inequitable conduct before the PTO" and "vexatious, unjustified, and otherwise bad faith litigation" as among the reasons that a prevailing party may receive reasonable attorney's fees).

164. *See, e.g., Yurman*, 262 F.3d at 112; *Hamil Am., Inc. v. GFI*, 193 F.3d 92, 97 (2d Cir. 1999); *Knitwaves, Inc. v. Lollytogs Ltd.*, 71 F.3d 996, 1010 (2d Cir. 1995).

165. *See supra* Section II.D (discussing the Federal Circuit's reliance in *Seagate* on the Supreme Court's interpretation of the word "willfulness").

166. *See In re Seagate Tech., LLC*, 497 F.3d 1360, 1370 (Fed. Cir. 2007) (en banc) (discussing copyright law's definition of "willfulness" as including reckless behavior); *see also* Rachel L. Emsley, Note, *Copying Copyright's Willful Infringement Standard: A Comparison of Enhanced Damages in Patent Law and Copyright Law*, 42 SUFFOLK U. L. REV. 157, 159 (2008) (stating that in adopting the recklessness standard, the Federal Circuit "import[ed] a standard from copyright law into patent law"); Justin McCarthy, Note, *In re Seagate: One Step Closer to a Rational Doctrine*, 10 MINN. J.L. SCI. & TECH. 355, 374 (2009) ("[T]he *Seagate* majority 'harmonized' willfulness between patent law and copyright law.").

167. 413 F.3d 257 (2d Cir. 2005).

holder's rights."¹⁶⁸ Several other circuit courts have adopted similar language when formulating their willful-infringement standards.¹⁶⁹ In *Zomba Enterprises, Inc. v. Panorama Records, Inc.*,¹⁷⁰ the Sixth Circuit applied this widely accepted standard when it found that the defendant acted recklessly when it continued to distribute copies of the plaintiff's copyrighted works after the district court entered a suspension order prohibiting the defendant from doing so.¹⁷¹

Since copyright law dealt with a recklessness standard for over a decade before *Iqbal* was decided, copyright law did not face the one-two punch that patent law encountered due to the relatively short period of time between the *Seagate* and *Iqbal* decisions. In other words, copyright law has had time to develop its standard for willful conduct. Patent law, on the other hand, was left high and dry in the wake of *Seagate* without much time to develop the standard for willfulness before also being faced with the task of determining what pleadings would satisfy the Court's new plausibility threshold.

However, copyright law may help to shed some light on patent law's post-*Seagate* development. For example, similar to the Second Circuit's holding in *Knitwaves, Inc. v. Lollytags, Ltd.*,¹⁷² a copyright case, the Federal Circuit adopted the view that "[t]he state of mind of the accused infringer is not relevant to th[e] objective inquiry [of willfulness]."¹⁷³ This rule is a significant departure from the Federal Circuit's previous negligence doctrine in the *Underwater Devices* regime, where the court placed substantial weight on an infringer's knowledge in determining whether the infringement was willful, and consequently, whether to award treble damages.¹⁷⁴ Therefore,

168. *Id.* at 263.

169. *E.g.*, *Superior Form Builders, Inc. v. Dan Chase Taxidermy Supply Co.*, 74 F.3d 488, 496 (4th Cir. 1996) (permitting a willfulness finding based on "any factor which the jury believes evidences the defendants . . . recklessly disregarded the fact that its conduct constituted copyright infringement"); *Wildlife Express Corp. v. Carol Wright Sales, Inc.*, 18 F.3d 502, 511–12 (7th Cir. 1994) (finding willful infringement where "the infringer has acted in reckless disregard of the copyright owner's right" (internal citation and quotation marks omitted)); *see RCA/Ariola Int'l v. Thomas & Grayston Co.*, 845 F.2d 773, 779 (8th Cir. 1988) ("[R]eckless disregard of the copyright holder's rights (rather than actual knowledge of infringement) suffices to warrant award of the enhanced damages.").

170. 491 F.3d 574 (6th Cir. 2007).

171. *Id.* at 585.

172. 71 F.3d 996, 1010 (2d Cir. 1995) ("Knowledge of infringement may be constructive rather than actual; that is, 'it need not be proven directly but may be inferred from the defendant's conduct.'" (quoting *N.A.S. Import Corp. v. Chenson Enters., Inc.*, 968 F.2d 250, 252 (2d Cir. 1992))).

173. *In re Seagate Tech., LLC*, 497 F.3d 1360, 1371 (Fed. Cir. 2007) (en banc).

174. *See Underwater Devices Inc. v. Morrison-Knudsen Co.*, 717 F.2d 1380, 1390 (Fed. Cir. 1983) (discussing the defendant's failure to obtain competent legal advice before infringing as a sign of a lack of good faith and a "fact to be weighed"). The Federal Circuit

under the objective standard in *Seagate*, the defendant's conduct *alone* provides the sufficient basis for collecting treble damages without any inquiry into the defendant's actual intent. Copyright law maintains a similar view.¹⁷⁵ Furthermore, the Federal Circuit made it clear in *Seagate* that while *Seagate* itself may have initiated the climb from negligence to recklessness, what actually constitutes reckless behavior and what must be alleged in a pleading would be for later cases to determine.¹⁷⁶ Given this lack of guidance, copyright law may provide valuable insight into the willfulness threshold and what must be pled to satisfy the *Twombly–Iqbal* plausibility standard.

2. *Inequitable Conduct: An Expressly Higher Threshold?*

Another type of bad-faith behavior that must be specifically pled in the interest of expanding the scope of discovery—like willful infringement after *Iqbal* and *Seagate*—is a plaintiff's inequitable conduct before the PTO when prosecuting a patent.¹⁷⁷ A patent is presumed valid upon issuance;¹⁷⁸ however, defendants in patent litigation frequently seek to render the patent unenforceable by alleging that the patentee failed to disclose material information during the prosecution process that, had such information been

further expounded upon its definition of “willfulness” in another pre-*Seagate* case, *Bott v. Four Star Corp.*, 807 F.2d 1567 (Fed. Cir. 1986). In *Bott*, the court set forth three criteria for assessing willfulness:

- (1) whether the infringer deliberately copied the ideas or design of another;
- (2) whether the infringer, when he knew of another's patent protection, investigated the scope of the patent and formed a good-faith belief that it was invalid or that it was not infringed; and
- (3) the infringer's behavior as a party to the litigation.

Bott, 807 F.2d at 1572.

175. See *Knitwaves*, 71 F.3d at 1010 (“[R]eckless disregard of the copyright holder's rights . . . suffices to warrant award of the enhanced damages.” (quoting *RCA/Ariola Int'l, Inc. v. Thomas & Grayston Co.*, 845 F.2d 773, 779 (8th Cir. 1988)) (internal quotation marks omitted)).

176. Although the *Seagate* court set forth what the new standard for willful infringement is, the Federal Circuit gave very little guidance on what actions satisfy that standard. Rather, the court stated that it will “leave it to future cases to further develop the application of this standard.” *Seagate*, 497 F.3d at 1371.

177. See 37 C.F.R. § 1.56 (2009). This regulation states:

Each individual associated with the filing and prosecution of a patent application has a duty of candor and good faith in dealing with the [Patent and Trademark] Office, which includes a duty to disclose to the Office all information known to that individual to be material to patentability as defined in this section.

Id.

178. 35 U.S.C. § 282 (2006) (“A patent shall be presumed valid.”).

known by the patent examiner, may have prevented such a broad claim scope from issuing.¹⁷⁹ This allegation falls within the framework of the doctrine of “unclean hands” on behalf of the then-applicant.¹⁸⁰ Conduct that can lead to a patent being rendered unenforceable includes the failure to disclose a known, material reference, or a false statement in an affidavit to the PTO.¹⁸¹

The Federal Circuit has not been shy about stating that inequitable-conduct defenses must be pled with particularity by defendants seeking to render a patent unenforceable.¹⁸² In a recent reinforcement of this standard, the Federal Circuit stated that in “pleading inequitable conduct in patent cases, Rule 9(b) requires identification of the specific who, what, when, where, and how of the material misrepresentation or omission committed before the PTO.”¹⁸³ While it did not expressly state what must be pled with

179. See Christian E. Mammen, *Controlling the “Plague”: Reforming the Doctrine of Inequitable Conduct*, 24 BERKELEY TECH. L.J. 1329, 1329, 1345 (2009) (noting how “claims directly affected by a [patentee’s] inequitable conduct . . . should be deemed unenforceable” and that the effect of the inequitable-conduct defense is to “render[] the entire patent, and possibly any descendant patents, unenforceable”); see also Kevin Mack, Note, *Reforming Inequitable Conduct to Improve Patent Quality: Cleansing Unclean Hands*, 21 BERKELEY TECH. L.J. 147, 161–62 (2006) (discussing the standard of materiality for inequitable conduct and how courts must determine if “one or more claims of the patent should not have issued or should have issued with different scope” (internal citation and quotation marks omitted)).

180. The doctrine of unclean hands is best explained by the Supreme Court’s opinion in *Keystone Driller Co. v. General Excavator Co.*, 290 U.S. 240 (1933). The Court elaborated:

[B]efore a complainant can have a standing in court he must first show that not only has he a good and meritorious cause of action, but he must come into court with clean hands. . . . [W]henever a party who, as *actor*, seeks to set the judicial machinery in motion and obtain some remedy, has violated conscience, or good faith, or other equitable principle, in his prior conduct, then the doors of the court will be shut against him *in limine*

Id. at 244–45.

181. Courts draw a fine distinction between failure to disclose a reference and making false statements. False affidavits technically fall under the category of “fraud on the Patent Office” and not inequitable conduct. See *Timely Prods. Corp. v. Arron*, 523 F.2d 288, 297 (2d Cir. 1975). However, “the distinction is without practical significance insofar as the consequences are concerned. In either case, the only value the patent has the right to exclude others from the use of the invention during the patent term is irreparably destroyed.” *Id.* at 297–98.

182. *E.g.*, *Exergen Corp. v. Wal-Mart Stores, Inc.*, 575 F.3d 1312, 1326–27 (Fed. Cir. 2009) (“Inequitable conduct, while a broader concept than fraud, must be pled with particularity under Rule 9(b).” (internal quotation marks omitted) (citing *Ferguson Beauregard/Logic Controls, Div. of Dover Res., Inc. v. Mega Sys., LLC*, 350 F.3d 1327, 1344 (Fed. Cir. 2003))); *Hebert v. Lisle Corp.*, 99 F.3d 1109, 1116 (Fed. Cir. 1996).

183. *Exergen*, 575 F.3d at 1327.

particularity, the Supreme Court in *Twombly* and *Iqbal* nevertheless set forth a standard that, at the very least, looks very similar to Rule 9(b) pleadings.¹⁸⁴

The language of the *Seagate* opinion thus puts patentees seeking to recover damages from alleged willful infringement in a difficult position: because willfulness does not require an inquiry into the defendant's state of mind,¹⁸⁵ the fact that the accused willful infringer may have had the intent to infringe the patent is of no consequence and thus, a general allegation will not suffice; however, "plead[ing] facts that are 'merely consistent with' a defendant's liability . . . 'stops short of the line between possibility and plausibility . . .'"¹⁸⁶ Therefore, it seems that Rule 8 pleadings under *Iqbal* actually set forth a higher threshold than do Rule 9 pleadings for inequitable conduct. While the *Iqbal* Court does not specifically mention the "who, what when, where, and how" facts that must be alleged, the Court does make it clear that even when well-pleaded factual allegations are set forth, and *even* when their veracity is assumed, the complainant may *still* come up short of what is necessary under Rule 8(a).¹⁸⁷ Reconciling how the Rule 8 pleading standard can—at least as far as the *Iqbal* opinion is most reasonably interpreted—rival the specificity required for Rule 9 pleadings and become more stringent illustrates a grave discrepancy among the courts as to what truly is a "short and plain statement."¹⁸⁸

C. DESERTION: THE RETIREMENT OF WILLFULNESS

Ashcroft v. Iqbal may signify the beginning of the end for treble damages in patent cases due to willfulness. The Supreme Court set what appears to be a virtually unattainable level of specificity for pleading to open the doors of discovery necessary for proving willful conduct. However, patentees may take refuge from the heightened pleading standard in two ways. First, at least in the case of high-technology patents, close-knit technology communities

184. Compare *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 558 (2007) ("[A] district court must retain the power to insist upon some specificity in pleading . . ." (quoting *Associated Gen. Contractors of Cal., Inc. v. Cal. State Council of Carpenters*, 459 U.S. 519, 528 n.17 (1983))), and *Ashcroft v. Iqbal*, 129 S. Ct. 1937, 1949 (2009) (requiring pleadings to contain "further factual enhancement" to overcome the plausibility threshold (quoting *Twombly*, 550 U.S. at 557)), with FED. R. CIV. P. 9(b) ("In alleging fraud or mistake, a party must state with particularity the circumstances constituting fraud or mistake. Malice, intent, knowledge, and other conditions of a person's mind may be alleged generally.")

185. *In re Seagate Tech., LLC*, 497 F.3d 1360, 1371 (Fed. Cir. 2007) (en banc).

186. *Iqbal*, 129 S. Ct. at 1949 (quoting *Twombly*, 550 U.S. at 557).

187. *Id.* at 1950 ("When there are well-pleaded factual allegations, a court should assume their veracity and *then* determine whether they plausibly give rise to an entitlement to relief." (emphasis added)).

188. FED. R. CIV. P. 8(a)(2).

may provide patentees the requisite information necessary to overcome the plausibility standard when pleading recklessness. Second, courts could adopt alternative criteria upon which to base treble-damages awards.

1. *Close-Knit Technology Communities: A Solution?*

Under the *Iqbal* regime, pleading sufficient facts about a defendant's conduct per *Seagate* may be just as difficult as pleading a defendant's specific state of mind in accordance with *Underwater Devices*. This is because the patent holder is only able to compare his own claims with the defendant's allegedly infringing device and may have no prior knowledge about the defendant.¹⁸⁹ However, at least in the case of high-technology patents, specialized-technology firms—like close-knit communities of artists in the copyright context¹⁹⁰—may provide a solution for plaintiffs who otherwise lack sufficient knowledge to prove willfulness.¹⁹¹ Under this theory, patentee-plaintiffs may be able to overcome the new pleading standard for objective recklessness by claiming that an alleged infringer had knowledge of the plaintiff's patent due to the close-knit communities that certain technology fields entail.¹⁹²

189. See Matthew A. Josephson, Note, *Some Things Are Better Left Unsaid: Pleading Practice After Bell Atlantic Corp. v. Twombly*, 42 GA. L. REV. 867, 894 (2008) (“Plaintiffs often face the problem of not being able to access concrete evidence at the pleading stage because the information concerning a particular controversy is accessible to only one party, usually the defendant.”); see also *supra* Section II.C.1 (discussing the strict-liability nature of patent infringement and resulting limited scope of discovery).

190. See Steven A. Hetcher, *Using Social Norms to Regulate Fan Fiction and Remix Culture*, 157 U. PA. L. REV. 1869, 1885 (2009) (“Members [of close-knit copyright communities] are inclined to take an interest in each other's actions precisely because each person, in terms of the actual utilities and disutilities, has an interest in any step that other members take to commercialize their acts of remix.”); see also Mark F. Schultz, *Fear and Norms and Rock & Roll: What Jambands Can Teach Us About Persuading People to Obey Copyright Law*, 21 BERKELEY TECH. L.J. 651, 681–82 (2006) (discussing, inter alia, four ways that potential copyright infringers communicate with one another about protected works, including “monitoring and reporting illegal activities”).

191. See Jordan Barry, *When Second Comes First: Correcting Patent's Poor Secondary Incentives Through an Optional Patent Purchase System*, 2007 WIS. L. REV. 585, 588 (“As research has advanced, technologies have become more closely knit, and the complementary and competitive relationships between them have become increasingly important.”); Lior Jacob Strahilevitz, *Social Norms from Close-Knit Groups to Loose-Knit Groups*, 70 U. CHI. L. REV. 359, 359 (2003) (defining close-knit groups as “network[s] in which power is broadly distributed and information pertinent to informal control circulates easily among network members”).

192. See, e.g., Henry E. Smith, *The Language of Property: Form, Context, and Audience*, 55 STAN. L. REV. 1105, 1175 (2003) (“Highly detailed and patent-specific information is not only indispensable, but the limited audience of potential violators *can be expected to process it.*” (emphasis added)); Strahilevitz, *supra* note 191, at 359 (“Typically, close-knit groups are made

While the *Seagate* standard imputes no value to the infringer's mental state, that an infringer was aware of another's patented technology may nevertheless bolster a plaintiff's claim that the objectively defined risk "was either known or so obvious that it should have been known to the accused infringer."¹⁹³ Berkeley Law Professor Robert Merges has expanded on this idea by claiming that there are some firms that knowingly infringe another group member's patents with the intention to either "settle up" in the future (by paying money damages) or "trade off" infringement liabilities.¹⁹⁴

2. *Alternative Criteria for Increasing Damages*

The average patentee probably has only limited knowledge of a defendant's allegedly infringing device, and almost no knowledge of the defendant's conduct leading up to his decision to manufacture that device.¹⁹⁵ Thus, the *Twombly-Iqbal* plausibility pleading standard, in combination with *Seagate's* recklessness standard, will almost certainly be beyond the reach of most plaintiffs in infringement actions. However, this is not to say that willfulness and treble damages are inexorably linked. In fact, while willfulness has been the longstanding criterion for awarding enhanced damages, a finding of willfulness does not require a court to award a plaintiff treble damages.¹⁹⁶ And, as the converse is also true, treble damages are not awarded

up of repeat players who can identify one another."); Richard Li-dar Wang, *Biomedical Upstream Patenting and Scientific Research: The Case for Compulsory Licensing Bearing Reach-Through Royalties*, 10 YALE J.L. & TECH. 251, 259–60 (2008). In his article, Professor Wang states:

Under the coordination of an "invisible hand" operating in the scientific community, however, scientists' pursuit of personal achievement weaves a closely knit web, as they tackle their common subject of research from every potential angle. Dissemination of scientific findings . . . works as a guiding compass, leading scientists [to] work[] on the same topic

Id.

193. *In re Seagate Tech., LLC*, 497 F.3d 1360, 1371 (Fed. Cir. 2007) (en banc).

194. Robert P. Merges, *Contracting into Liabilities Rules: Intellectual Property Rights and Collective Rights Organizations*, 84 CALIF. L. REV. 1293, 1354–55 (1996).

195. Cf. Kevin G. McBride, Comment, *Attributing Knowledge to a Corporation as an Element of Laches in a Suit for Patent Infringement*, 80 NW. U. L. REV. 698, 698–702 (1985) (explaining how a plaintiff's lack of knowledge about a defendant and its corresponding infringing activities can bar a defendant's laches defense); Robert P. Merges & Jeffrey M. Kuhn, *An Estoppel Doctrine for Patented Standards*, 97 CALIF. L. REV. 1, 44 (2009) ("Courts refuse to permit a laches defense in the absence of actual or constructive knowledge.").

196. *Johns Hopkins Univ. v. CellPro, Inc.*, 152 F.3d 1342, 1365 (Fed. Cir. 1998) ("[A] finding of willful infringement does not mandate that the district court enhance damages; it merely authorizes the court to do so at its discretion."); *Read Corp. v. Portec, Inc.*, 970 F.2d 816, 826 (Fed. Cir. 1992) ("[A] finding of willful infringement does not mandate that damages be enhanced, much less mandate treble damages."), *abrogated on other grounds by*

only upon a court finding that the defendant infringed willfully (although willfulness has emerged to be the hallmark indicia).¹⁹⁷ Indeed, the Federal Circuit has long recognized this proposition.¹⁹⁸

As the ease of proving willfulness becomes largely a luxury of past pleading and negligence regimes under *Conley* and *Underwater Devices*, patent law may need to develop alternative criteria for awarding treble damages. While one could argue that reasonable-royalty damages will simply become higher to compensate patentees for their inability to prove willfulness, the Federal Circuit has made it clear that compensatory damages must remain at a level “adequate to compensate for the infringement”¹⁹⁹ and must not stray into being awarded punitively where there is no finding of bad faith.²⁰⁰ Thus, *Iqbal* has presented a challenge that the Federal Circuit must address when awarding treble damages that may be founded on bad-faith conduct other than willfulness. The task is not an impossible one, however. Although the Patent Act’s damages provision provides very little to no guidance for when treble damages should be awarded,²⁰¹ and courts have previously found this

Markman v. Westview Instruments, Inc., 52 F.3d 967, 975 (Fed. Cir. 1995); Modine Mfg. Co. v. Allen Grp., Inc., 917 F.2d 538, 543 (Fed. Cir. 1990) (“The decision to increase damages is committed to the discretion of the trial judge and a district court’s refusal to award increased damages will not be overturned absent a clear showing of abuse of discretion.”).

197. See *Seagate*, 497 F.3d at 1381 (referring to numerous Supreme Court cases that recognize the “uncontroversial proposition that a finding of willfulness is *sufficient* to support an award of enhanced damages,” but none of which “hold that a finding of willfulness is *necessary* to support an award of enhanced damages”).

198. See *Portec*, 970 F.2d at 826 (stating that “[a]n award of enhanced damages for infringement, as well as the extent of the enhancement, is committed to the discretion of the trial court” and acknowledging that “no statutory standard dictates the circumstances under which the district court may exercise its discretion”).

199. 35 U.S.C. § 284 (2006) (“Upon finding for the claimant the court shall award the claimant damages adequate to compensate for the infringement, but in no event less than a reasonable royalty for the use made of the invention by the infringer . . .”).

200. *Beatrice Foods Co. v. New Eng. Printing & Lithographing Co.*, 923 F.2d 1576, 1579 (Fed. Cir. 1991) (“[E]nhanced damages may be awarded *only* as a penalty for an infringer’s increased culpability, namely willful infringement or bad faith. Damages cannot be enhanced to award the patentee additional compensation to rectify what the district court views as an inadequacy in the actual damages awarded.” (emphasis added)). The court in *Beatrice Foods* rejected the district court’s reasoning for increasing the damages by \$16 million to “compensate” the plaintiff. *Id.*

201. See 35 U.S.C. § 284 (2006) (“When the damages are not found by a jury, the court shall assess them. In either event the court may increase the damages up to three times the amount found or assessed.”); see also James G. McEwen, *Is the Cure Worse than the Disease? An Overview of the Patent Reform Act of 2005*, 5 J. MARSHALL REV. INTELL. PROP. L. 55, 67 (2005) ([T]here has been little statutory guidance as to when increased damages can be assessed.”); Wright, *supra* note 106, at 101 (“The Patent Act of 1952 merely recited past statutory

ambiguous provision troublesome in determining the appropriate circumstances for awarding treble damages, the want of constraints may inure to patentees' benefit. The Act's lack of guidance may allow courts to develop conduct standards that can be pled with specificity more easily and that can nevertheless be indicative of bad faith.²⁰²

In fact, many district courts and circuit courts other than the Federal Circuit have already established alternative criteria for determining when to increase—as well as when to limit—a patentee–plaintiff's treble damages. For example, in *American Safety Table Co. v. Schreiber*,²⁰³ the Second Circuit declined to award a plaintiff enhanced damages despite acknowledging that the defendant acted “deliberate[ly] and with knowledge of plaintiff's rights.”²⁰⁴ The Second Circuit considered the defendant's motivation for infringement and based its reasoning on the fact that the defendant's infringing conduct was prompted by economic pressures due to customer dissatisfaction.²⁰⁵ In *St. Regis Paper Co. v. Winchester Carton Corp.*,²⁰⁶ the District Court of Massachusetts cabined an enhanced-damages award based on the defendant's size.²⁰⁷ Sitting by designation, Senior Judge Aldrich of the First Circuit qualified the damages award by stating that “[i]f [the] defendant were the giant and plaintiff the small independent [business], I would make it treble.”²⁰⁸ Similarly, other courts have also considered the defendant's financial condition in deciding when and to what extent to increase damages.²⁰⁹ The First Circuit in *Russell Box Co. v. Grant Paper Box Co.*²¹⁰ even

language and offered no additional guidance to the courts. Thus, it was left to the courts to exercise discretion in awarding enhanced damages under the statute.”).

202. *See, e.g., Aro Mfg. Co. v. Convertible Top Replacement Co.*, 377 U.S. 476, 508 (1964) (noting that the patentee could recover punitive damages for “willful or bad-faith infringement” (emphasis added)); *see also Seagate*, 497 F.3d at 1381 (reviewing a line of Supreme Court cases that the Federal Circuit determined “cannot be interpreted to mean that enhanced damages are limited to a finding of willfulness”).

203. 415 F.2d 373 (2d Cir. 1969).

204. *Id.* at 379.

205. *Id.*

206. 410 F. Supp. 1304 (D. Mass. 1976).

207. *Id.* at 1309.

208. *Id.*

209. *E.g., Lightwave Tech., Inc. v. Corning Glass Works*, No. 86 Civ. 759 (KC), 1991 WL 4737, at *13 (S.D.N.Y. Jan. 18, 1991) (concluding that the defendant “can withstand some increase in damages, but not treble damages”); *Bott v. Four Star Corp.*, Civ. No. 79-71438, 1985 WL 6071, at *19 (E.D. Mich. Dec. 16, 1985) (recognizing that “a threefold increase in damages would severely affect [defendant's] financial condition”), *vacated and remanded for clarification of damage amount*, 807 F.2d 1567 (Fed. Cir. 1986).

210. 203 F.2d 177 (1st Cir. 1953).

went so far as to increase damages based on the defendant's conduct after the suit was initiated and during the trial.²¹¹ Other factors courts have used to increase or limit treble damages include remedial actions by the defendant,²¹² "closeness" of the case,²¹³ and duration and timing of the defendant's infringement.²¹⁴

These cases illustrate that even if enhanced damages due to willfulness are to drown under the weight of *Iqbal's* pleading standard and *Seagate's* objective-recklessness framework, enhanced damages as a whole do not necessarily have to meet a similar fate. Rather, the Federal Circuit can adopt other indicia of bad faith that may be more easily pled to overcome *Iqbal's* plausibility standard or that may be self-revealing during litigation. Furthermore, as illustrated in some of the above cases, the Federal Circuit can choose to broaden the criteria for awarding treble damages beyond bad faith. Because 35 U.S.C. § 284 provides no guidance to courts on when to increase damages,²¹⁵ a defendant's characteristics alone can establish grounds for awarding a plaintiff enhanced damages (e.g., a large corporate defendant being sued by a small independent inventor).

V. CONCLUSION

Under the Supreme Court's instruction in *Ashcroft v. Iqbal*, the path to treble damages due to willfulness in patent infringement cases will undoubtedly be steeper—and more expensive—for patentees. This is because access to defendants' documents and indicia of objective

211. *Id.* at 183 (enhancing damages based on the defendant's attempt to conceal the wrongdoing, and noting that "the defendant had failed to preserve its records and had failed to cooperate as it should at the trial of the issue of damages").

212. *Intra Corp. v. Hamar Laser Instruments, Inc.*, 662 F. Supp. 1420, 1439 (E.D. Mich. 1987) (limiting the enhanced damages based on the defendant's remedial conduct, and acknowledging that the defendant "voluntarily ceased manufacture and sale of infringing systems during the pendency of this litigation"), *aff'd*, 862 F.2d 320 (Fed. Cir. 1988).

213. *Modine Mfg. Co. v. Allen Grp., Inc.*, 917 F.2d 538, 543 (Fed. Cir. 1990) (weighing the "closeness of the case" in its decision to allocate litigation burdens among the parties) (citing *S.C. Johnson & Son, Inc. v. Carter-Wallace, Inc.*, 781 F.2d 198, 201 (Fed. Cir. 1986)); *Crucible, Inc. v. Stora Kopparbergs Bergslags AB*, 701 F. Supp. 1157, 1164 (W.D. Pa. 1988) ("[B]ecause the court still considers the question [of infringement] to be a close one; . . . the court concludes that double, and not treble damages are appropriate.").

214. *Bott*, 1985 WL 6071, at *19 (increasing damages by twenty percent for sales made while liability was under appeal and doubling damages for sales made after the appellate court affirmed the finding of liability).

215. 35 U.S.C. § 284 (2006) ("[T]he court may increase the damages up to three times the amount found or assessed."); *see supra* notes 71, 201 and accompanying text (noting how the damages provision is devoid of any instruction on when to award enhanced damages).

recklessness will not be attainable without sufficiently pleading facts that demonstrate a plausible showing of willful conduct. Because patent infringement is a strict-liability offense, the judge or jury will be able to determine whether the defendant is liable for compensatory damages by merely comparing the plaintiff's patent claims with the defendant's allegedly infringing article.²¹⁶ The scope of discovery, then, will be limited to matters related only to direct infringement. Thus, under the Court's new plausibility regime for Rule 8 pleadings, patentees must do more than merely plead the bare elements of willful infringement to "nudge" willfulness allegations "across the line from conceivable to plausible"²¹⁷ and open further the doors of discovery.

When assessing the validity of willfulness allegations, district courts and the Federal Circuit should construe the pleadings in accordance with Federal Rule 8(e) so as to promote justice for the plaintiff while steadfastly requiring the complaint to "contain sufficient factual matter" beyond a mere "[t]hreadbare recital[]" of the elements of [the] cause of action."²¹⁸ Moreover, the Federal Circuit should consider expounding the criteria for awarding treble damages, which has traditionally been narrowly limited to bad faith, to other indicia that may justify larger damage awards. Adhering to these principles, the Federal Circuit can ensure that discovery will be limited to only those indicia of recklessness which pass muster under the Supreme Court's heightened plausibility standard while still adequately compensating plaintiffs in patent infringement cases.

216. *See supra* Section II.C.1 (explaining the strict-liability nature of patent infringement).

217. *Ashcroft v. Iqbal*, 129 S. Ct. 1937, 1951 (2009) (internal quotation marks omitted) (quoting *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 570 (2007)).

218. *Id.* at 1949.

