

# BERKELEY TECHNOLOGY LAW JOURNAL

VOLUME 25

NUMBER 2

SPRING 2010

## TABLE OF CONTENTS

### ARTICLES

PATENT CLAIM CONSTRUCTION: A MODERN SYNTHESIS AND STRUCTURED FRAMEWORK .....	711
<i>Peter S. Menell, Matthew D. Powers, &amp; Steven C. Carlson</i>	
COMMUNICATIONS PRIVACY IN THE MILITARY .....	831
<i>Justin Holbrook</i>	
THE IMPERFECT IS THE ENEMY OF THE GOOD: ANTICIRCUMVENTION VERSUS OPEN USER INNOVATION .....	909
<i>Wendy Seltzer</i>	
LAW, TECHNOLOGY, AND SHIFTING POWER RELATIONS .....	973
<i>Bert-Jaap Koops</i>	
WHAT PAYMENT INTERMEDIARIES ARE DOING ABOUT ONLINE LIABILITY AND WHY IT MATTERS .....	1037
<i>Mark MacCarthy</i>	
LICENSING COMPLEMENTARY PATENTS: "PATENT TROLLS," MARKET STRUCTURE, AND "EXCESSIVE" ROYALTIES .....	1121
<i>Anne Layne-Farrar &amp; Klaus M. Schmidt</i>	

## SUBSCRIBER INFORMATION

The *Berkeley Technology Law Journal* (ISSN 1086-3818), a continuation of the *High Technology Law Journal* effective Volume 11, is edited by the students of the University of California School of Law, Berkeley (Boalt Hall), and published four times each year (March, June, September, January) by the Regents of the University of California, Berkeley, Journal Publications, School of Law, 2850 Telegraph Avenue, Suite 561 #7220 Berkeley, CA 94705-7220. Periodicals Postage Rate Paid at Berkeley, CA 94704-9998, and at additional mailing offices. POSTMASTER: Send address changes to Journal Publications, 311 U.C. Berkeley School of Law, University of California, Berkeley, CA 94720-7200.

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

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

**Form.** The text and citations in the *Journal* conform generally to the THE CHICAGO MANUAL OF STYLE (15th ed. 2003) and to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (Columbia Law Review Ass'n et al. eds., 18th ed. 2005). Please cite this issue of the *Berkeley Technology Law Journal* as 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.

# PATENT CLAIM CONSTRUCTION: A MODERN SYNTHESIS AND STRUCTURED FRAMEWORK

*Peter S. Menell,<sup>†</sup> Matthew D. Powers,<sup>††</sup> & Steven C. Carlson<sup>†††</sup>*

## TABLE OF CONTENTS

I.	INTRODUCTION.....	714
II.	A STRUCTURED FRAMEWORK FOR CLAIM CONSTRUCTION.....	717
A.	DERIVING MEANING FROM CLAIMS .....	718
1.	<i>Claim Drafting: The Genesis and Evolution of Claim Terms</i> .....	719
2.	<i>Sources for Deriving Claim Meaning</i> .....	720
a)	Principal Source: Intrinsic Evidence .....	722
i)	Specification .....	722
ii)	Prosecution History .....	723
iii)	Related and Foreign Applications.....	723
b)	Extrinsic Evidence Permissible, But It May Not Contradict or Override Intrinsic Evidence .....	725
i)	Illustrations of Reliance (and Non-Reliance) upon Extrinsic Evidence.....	727
ii)	Conclusory Expert Opinions Should Be Disregarded .....	729
B.	A STRUCTURED APPROACH TO CLAIM CONSTRUCTION: TWO STAGES OF ANALYSIS.....	730
1.	<i>Step 1: Is Construction of a Claim Term Required?</i> .....	730

---

© 2010 Peter S. Menell, Matthew D. Powers, and Steven C. Carlson.

† Professor of Law and Director, Berkeley Center for Law & Technology, University of California at Berkeley School of Law. This Article grew out of the Patent Case Management Judicial Guide, a treatise developed for federal judges published in 2009 by the Federal Judicial Center. We worked with leading patent jurists, practitioners, and academics in developing this guide. We want to thank the Berkeley Center for Law & Technology and the Federal Judicial Center for their generous support of these projects. We owe special thanks for Judge Ronald Whyte, Judge Kathleen O'Malley, Lynn Pasahow, James Pooley, Mark Lemley, George Pappas, Nick Brown, Carolyn Chang, Tom Fletcher, Jeff Homrig, Marc David Peters, and Sue Vastano Vaughan for their contributions to this project. We also thank Ashley Doty and Jason Romrell for their research assistance in the preparation of this Article and Laura Rocheloios and By Design Legal Graphics for their assistance with illustrations.

†† Partner, Weil, Gotshal & Manges LLP.

††† Principal, Fish & Richardson PC.

a)	Is There a Genuine Dispute About the Claim Term?.....	731
b)	Would Claim Construction of the Term Help the Jury? .....	731
c)	Is Claim Construction of the Term a Priority?.....	732
d)	Has the Term Been Construed Before?.....	732
e)	Is the Term Amenable to Construction?.....	733
i)	Lay Terms .....	734
ii)	Terms of Degree.....	736
iii)	Technical Terms .....	737
2.	<i>Step 2: Interpretation of Claim Language</i> .....	737
a)	General Framework .....	737
b)	Claim Construction Methodology .....	739
c)	Misuse of “Ordinary Meaning” .....	743
d)	Interpreting Claim Language in Light of the Specification .....	745
i)	The Role of Preferred Embodiments in Claim Construction.....	745
(1)	<i>Claim Scope Generally Includes Preferred Embodiments</i> .....	746
(2)	<i>Is the Patent Limited to the Preferred Embodiments?</i> .....	746
(3)	<i>Does the Number or Range of Embodiments Affect the Scope of the Claims?</i> .....	748
(4)	<i>Does Ambiguity in a Claim Term Limit Its Scope to Preferred Embodiment(s)?</i> .....	749
ii)	Characterizations of “The Invention” or “The Present Invention” .....	750
iii)	Distinctions Over the Prior Art .....	751
iv)	Consistent Usage of Claim Terms .....	752
e)	Prosecution Disclaimers.....	752
f)	Looking to Other Claims: The Doctrine of Claim Differentiation.....	753
g)	Significance of the “Preamble” in Claim Construction.....	755
3.	<i>Claim Terms Having Conventional, Presumed, or Established Meanings</i> .....	757
4.	<i>Interpreting Terms to Preserve Validity</i> .....	764
C.	SPECIAL CASE: MEANS-PLUS-FUNCTION CLAIMS.....	765
1.	<i>Step 1: Is the Term in Question “Means-Plus-Function”?</i> .....	766
2.	<i>Step 2: Interpretation of Means-Plus-Function Claim Terms</i> .....	768
a)	Step 2A: Identify Claim Term Function .....	768
b)	Step 2B: Identify “Structure, Material, or Acts” .....	768

	c)	Step 2C: “Equivalents Thereof” .....	769
	d)	Specific Rule for Means-Plus-Function Claims in the Computer Software Context.....	770
D.		DYSFUNCTIONAL CLAIMS: MISTAKES AND INDEFINITENESS .....	770
	1.	<i>Mistakes</i> .....	770
	2.	<i>Indefiniteness</i> .....	772
E.		DEFERENCE TO PRIOR CLAIM CONSTRUCTION RULINGS .....	774
	1.	<i>Distinguishing Among Preclusion and Estoppel Doctrines</i> .....	775
	2.	<i>Issue Preclusion and Collateral Estoppel</i> .....	777
	a)	Identity of Issues .....	777
	b)	Actual Litigation .....	778
	c)	Full and Fair Opportunity to Litigate.....	778
	d)	Determination Was Essential to the Final Judgment.....	779
	i)	Finality.....	780
		(1) <i>Summary Judgment</i> .....	780
		(2) <i>Preliminary Injunction</i> .....	780
		(3) <i>Settlement</i> .....	781
	ii)	Essential to the Final Judgment .....	783
	e)	Reasoned Deference as a Prudent Approach to Issue Preclusion .....	783
	3.	<i>Judicial Estoppel</i> .....	784
	4.	<i>Stare Decisis</i> .....	786
III.		<b>CLAIM CONSTRUCTION PROCEDURE</b> .....	787
A.		PATENT LOCAL RULES.....	788
B.		TIMING OF <i>MARKMAN</i> HEARINGS.....	792
C.		STREAMLINING THE PRE- <i>MARKMAN</i> PROCESS .....	793
	1.	<i>Mandatory Disclosure of Positions</i> .....	793
	a)	Early Disclosure of Infringement and Invalidity Contentions .....	794
	b)	Disclosure of Claims to Construe and Proposed Constructions .....	794
	c)	Mechanisms for Limiting the Number of Claim Terms to Construe .....	795
	d)	Severance versus Postponement.....	797
	e)	Recommended Approach: Mandatory Disclosure of Impact of Proposed Constructions .....	798
	2.	<i>Use of Tutorials, Experts, and Advisors in Claim Construction</i> .....	800
	a)	Technology Tutorials .....	801
	b)	Court-Appointed Experts .....	803
	i)	Technical Advisor.....	803
	ii)	Special Master .....	805
	iii)	Expert Witness.....	806

D.	SUMMARY JUDGMENT AND CLAIM CONSTRUCTION .....	806
1.	<i>Summary Judgment and Claim Construction</i> .....	807
2.	<i>Recommended Dual-Track Approach to Summary Judgment</i> .....	807
a)	“First-Track” Summary Judgment Motions.....	808
b)	“Second-Track” Summary Judgment Motions.....	810
c)	Implementing a Dual-Track Approach to Summary Judgment.....	810
d)	Recognizing First-Track Summary Judgment Motions .....	811
3.	<i>Summary Judgment Independent from Claim Construction (Off- Track)</i> .....	813
E.	CONDUCT OF THE MARKMAN HEARING.....	814
1.	“Evidentiary” Nature of Markman Hearings.....	814
2.	Safeguards on Extrinsic Evidence .....	815
3.	Evidence of the Accused Device.....	816
4.	Evidence of the Prior Art.....	816
5.	The Need to Focus Markman Proceedings on Claim Construction.....	817
6.	Sequence of Argument.....	818
F.	THE MARKMAN RULING .....	818
1.	Interrelationship to Jury Instructions.....	818
2.	Basis for Appellate Review.....	818
3.	The Court May Adopt Its Own Construction .....	819
4.	Tentative Rulings Prior to the Markman Hearing.....	819
5.	Integrating the Markman Ruling into Trial .....	820
a)	Amendments to Infringement and Invalidity Contentions .....	820
b)	Integrating the <i>Markman</i> Ruling into Jury Instructions.....	820
c)	Interlocutory Appeal of <i>Markman</i> Rulings .....	821
IV.	CONCLUSIONS.....	823
	<b>APPENDIX: NARROWING OR BROADENING "ORDINARY MEANING"</b> .....	824

## I. INTRODUCTION

The construction of patent claims plays a critical role in nearly every patent case. It is central to the evaluation of infringement and validity, and can affect or determine the outcome of other significant issues such as unenforceability, enablement, and remedies. Over the past two decades, the substantive standards and process for delineating patent claim terms have

undergone significant evolution. The Supreme Court's decision in *Markman v. Westview Instruments, Inc.*<sup>1</sup> marked the beginning of the new era. However, the Federal Circuit's search for workable standards as well as the experimentation of district courts with case management process—most notably the development and spread of Patent Local Rules—also played major roles in the reformation of patent litigation. The result is a bewildering array of cases and rules that can overwhelm litigants, counsel, law clerks, and jurists.<sup>2</sup> Scholars have found a relatively high reversal rate for claim construction rulings<sup>3</sup> and shown that even experienced patent jurists fare little better than new judges.<sup>4</sup> Consequently, scholars roundly criticize the jurisprudence of claim construction for lacking theoretical or practical coherence.<sup>5</sup>

---

1. 517 U.S. 370 (1996).

2. Even experienced district court judges have expressed deep frustration with the reversal rates for claim construction. *E.g.*, Anandashankar Mazumdar, *Federal District Courts Need Experts That Are Good 'Teachers,' Judges Tell Bar*, 70 PAT. TRADEMARK & COPYRIGHT J. (BNA) 536, 537 (2005) (quoting a district court judge suggesting that given the high reversal rate on claim construction “you might as well throw darts”); Kathleen M. O'Malley et al., *A Panel Discussion: Claim Construction from the Perspective of the District Judge*, 54 CASE W. RES. L. REV. 671, 682 (2004) (noting that some district court judges are “demoralize[d]” by the high reversal rate). The Federal Circuit has noted the concern. *See Merck & Co. v. Teva Pharm. USA, Inc.*, 395 F.3d 1364, 1381 (Fed. Cir. 2005) (Rader, J., dissenting) (noting that the Federal Circuit “often hears criticism from district court judges that its reversal rate on claim construction issues far exceeds that of other circuit courts”); *Ultratech, Inc. v. Tamarak Scientific Co.*, No. C. 03-03235 CRB, 2005 WL 2562623, at \*7 (N.D. Cal. Oct. 12, 2005) (“Nor can the Court say that Ultratech’s claim construction position is so frivolous as to warrant sanctions; to be candid, this Court is reluctant to hold that any claim construction is frivolous, given the well-known reversal rate in the Federal Circuit.”).

3. Christian A. Chu, *Empirical Analysis of the Federal Circuit’s Claim Construction Trends*, 16 BERKELEY TECH. L.J. 1075, 1143 (2001); Kimberly A. Moore, *Markman Eight Years Later: Is Claim Construction More Predictable?*, 9 LEWIS & CLARK L. REV. 231, 232–34 (2005); Michael Saunders, *A Survey of Post-Phillips Claim Construction Cases*, 22 BERKELEY TECH. L.J. 215, 233 (2007); Andrew T. Zidel, *Patent Claim Construction in the Trial Courts: A Study Showing the Need for Clear Guidance from the Federal Circuit*, 33 SETON HALL L. REV. 711, 743 (2003). Although a thirty percent reversal rate appears troublingly high, it is not significantly above reversal rates in other areas of complex litigation. *See Jeffrey A. Lefstin, Claim Construction, Appeal, and the Predictability of Interpretive Regimes*, 61 U. MIAMI L. REV. 1033, 1038–39 (2007).

4. *See David L. Schwartz, Courting Specialization: An Empirical Study of Claim Construction Comparing Patent Litigation Before Federal District Courts and the International Trade Commission*, 50 WM. & MARY L. REV. 1699, 1720 (2009); David L. Schwartz, *Practice Makes Perfect? An Empirical Study of Claim Construction Reversal Rates in Patent Cases*, 107 MICH. L. REV. 223, 258–59 (2008).

5. *See, e.g.*, Gretchen A. Bender, *Uncertainty and Unpredictability in Patent Litigation: The Time Is Ripe for a Consistent Claim Construction Methodology*, 8 J. INTELL. PROP. L. 175, 209 (2001); Dan L. Burk & Mark A. Lemley, *Fence Posts or Sign Posts? Rethinking Patent Claim Construction?*, 157 U. PA. L. REV. 1743, 1744 (2009); Russell B. Hill & Frank P. Cote, *Ending the Federal Circuit Crapshoot: Emphasizing Plain Meaning in Patent Claim Interpretation*, 42 IDEA 1,

If nothing else, the past two decades revealed the inherent difficulties of using language to define the boundaries of abstract and intangible rights. These challenges grew with the rise of information technologies. The boundaries of patent claims to software and business methods have proven particularly ambiguous.<sup>6</sup>

This Article provides a cohesive framework and roadmap for navigating this rapidly evolving landscape as well as guidance on the best practices for managing claim construction. It reflects the culmination of more than a decade of working with the Federal Judicial Center, leading jurists in districts with the largest patent dockets, experienced litigators, and academics to understand the specialized field of patent litigation. From a conceptual standpoint, the Article takes a pragmatic and experiential approach. Part II begins with a step-by-step approach to the task of construing patent claim terms. This Article integrates the many principles, canons, and doctrines within a structured framework. With that architecture in place, we organize and explore the various doctrines, with emphasis on their practical significance. Part III turns to the role of procedure in claim construction. The Article discusses the pioneering work of jurists and litigators in the Northern District of California—a prominent technology center and patent district—in developing a pragmatic set of case management and disclosure rules for managing the claim construction process. Many of the patent-intensive districts throughout the nation have adopted some version of these rules. The Article then examines additional best practices for structuring the claim construction determinations, including determining how many claim terms to construe (and when to make those determinations), the use of tutorials in conjunction with claim construction, and integrating claim construction and dispositive motions.

---

2 (2002); Timothy R. Holbrook, *Substantive Versus Process-Based Formalism in Claim Construction*, 9 LEWIS & CLARK L. REV. 123, 151 (2005); Joseph S. Miller, *Enhancing Patent Disclosure for Faithful Claim Construction*, 9 LEWIS & CLARK L. REV. 177, 177 (2005); Kelly C. Mullally, *Patent Hermeneutics: Form and Substance in Claim Construction*, 59 FLA. L. REV. 333, 336 (2007); Craig A. Nard, *A Theory of Claim Interpretation*, 14 HARV. J.L. & TECH. 1, 82 (2000); Kristen Osenga, *Linguistics and Patent Claim Construction*, 38 RUTGERS L.J. 61, 62–63 (2006); R. Polk Wagner & Lee Petherbridge, *Is the Federal Circuit Succeeding? An Empirical Assessment of Judicial Performance*, 152 U. PA. L. REV. 1105, 1171 (2004).

6. See JAMES BESSEN & MICHAEL J. MEURER, PATENT FAILURE: HOW JUDGES, BUREAUCRATS, AND LAWYERS PUT INNOVATORS AT RISK 201–03 (2008).

## II. A STRUCTURED FRAMEWORK FOR CLAIM CONSTRUCTION

It is useful to have some historical and jurisprudential context for claim construction in place before delving into the details. With a growing trend of using juries in patent cases since 1980, the issue emerged of whether the judge or the jury should construe the terms of patent claims. Until 1996, it was common in jury trials for courts to include claim construction as part of the jury's charge. Resolving the scope of patent claims in this manner, however, significantly increased the complexity and uncertainty of trials. The question of who should have responsibility to determine the meaning of patent claims came before the Supreme Court in the seminal case of *Markman v. Westview Instruments, Inc.*,<sup>7</sup> from which the term “*Markman* hearing” is derived.

In *Markman*, Markman sued Westview Instruments for infringement of its patent on a system for tracking articles of clothing in a dry-cleaning operation. After a jury found infringement, Westview Instruments moved for judgment as a matter of law on the ground that the patent and its prosecution history made clear that the patent claims at issue did not extend to the defendant's accused device. The trial court granted the motion based on its examination of the patent and other evidence presented. On appeal, the patentee asserted that the trial court's judgment violated its Seventh Amendment right to a jury trial on claim construction. Markman called attention to the fact that it had introduced expert testimony on the issue. Based on the historical allocation of responsibilities between judge and jury as well as functional considerations (the training and experience of judges in interpreting written instruments and the technical nature of patent claims), the Supreme Court held that claim construction is a matter for the court and hence beyond the province of the jury. The Court emphasized that judges are better equipped than juries to construe the meaning of patent claim terms given their training and experience interpreting written instruments (such as contracts and statutes). And even though cases may arise in which the credibility of competing experts affects the determination of claim meaning, the Court anticipated that claim construction determinations will be “subsumed within the necessarily sophisticated analysis of the whole document, required by the standard construction rule that a term can be defined only in a way that comports with the instrument as a whole.”<sup>8</sup> The Court also emphasized that judges are better able to promote uniformity and

---

7. 517 U.S. 370 (1996).

8. *Id.* at 389.

certainty in claim construction.<sup>9</sup> The Court specifically noted that treating claim construction as a “purely legal” issue would serve stare decisis principles as courts are better situated to give due weight to decisions of other courts that have previously ruled on the same issues.<sup>10</sup>

Although resolving an important issue for patent litigation, *Markman* spawned a complex set of substantive and procedural questions regarding when and how courts should construe patent claims. This Article begins with the framework and substantive rules governing claim interpretation and then presents the procedural matters relating to claim construction.

#### A. DERIVING MEANING FROM CLAIMS

Although providing some guidance on the approach for construing patent claims, the *Markman* decision spawned many issues relating to the proper framework for determining claim meaning. The Federal Circuit has issued over 1,000 opinions since *Markman* addressed this subject. Over the years, the Federal Circuit shifted its approach and, therefore, it is critical for courts to focus on the most current and authoritative decisions. The Federal Circuit’s en banc decision in *Phillips v. AWH Corp.*<sup>11</sup> stands as the most authoritative synthesis of claim construction doctrine. While *Phillips* put to rest various controversies, many core tensions in claim construction persist. Moreover, the decision itself does not provide a step-by-step approach to construing claims. This Section provides a systematic process for approaching the *Markman* determination.

This Section begins by explaining the process of claim drafting so as to understand the genesis and evolution of claim terms. It then previews the sources for determining claim meaning and the general hierarchy set forth in *Phillips*. With this background in place, this Section offers a structured analysis of claim construction. At the highest level of abstraction, claim construction entails analysis of several threshold questions regarding whether and when to interpret a claim term and then working through the construal process. The court begins the process with an initial interpretation of the claim term in question based on its own reading. To the extent that the parties identify additional sources of guidance from the intrinsic evidence or extrinsic sources, the court must then systematically work through the various sources to reach a proper construction. There are several special cases as well: commonly interpreted terms, means-plus-function claim terms,

---

9. *Id.* at 390–91.

10. *Id.* at 391.

11. 415 F.3d 1303 (Fed. Cir. 2005) (en banc).

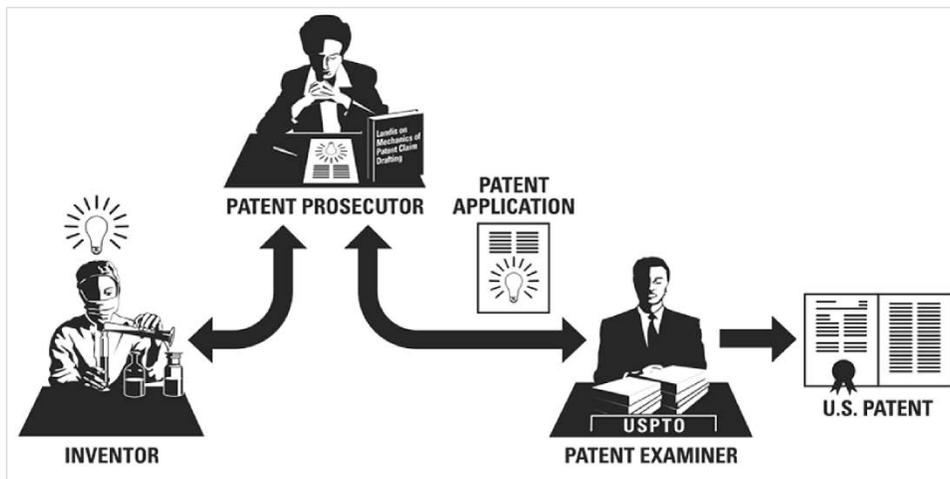
and mistaken or indefinite claim terms. We also explore the appropriate deference accorded to prior claim construction rulings.

1. *Claim Drafting: The Genesis and Evolution of Claim Terms*

Patent claim terms emerge through a process typically involving multiple contributors employing at least three distinct and distinctive vocabularies: plain English, conventions of claim drafting, and scientific or technical terminology. The court is comfortable with the former but may need assistance interpreting terms that derive from the fields of science and claim drafting. Understanding the process of claim term drafting will assist in surmounting that semantic challenge.

Chart 1 illustrates the drafters and lines of communication and collaboration leading to the ultimate words used in patent claims. The claim drafting process begins with the invention and inventor(s). Whether independent or employed in a corporate or university research and development unit, the inventor(s) will, in most cases, communicate their ideas to a trained patent attorney or agent. That person will typically have some familiarity with the field of invention (although not necessarily to the level of the inventor) as well as substantial training in the drafting of patent applications. Her job is to describe and claim the invention in terms that will satisfy the requirements of the Patent Act. She will seek to write the claims with sufficient specificity to clear the validity hurdles while providing the patentee with significant breadth to cover the foreseeable uses of the invention. As indicated in Chart 1 by the two-headed arrow between the inventor and the patent prosecutor, there is often substantial back and forth between the inventor and the drafter before filing of the initial application.

Chart 1: Crafting of Patent Claim Terms



The process of claim drafting does not end with the submission of the patent application. The patent examiner will often play a role in the ultimate claim language of patents. Like the patent prosecutor, examiners have some knowledge of the technical field as well as experience in the process of claim drafting and evaluation. As with the application drafting process, communication between the prosecutor and the examiner travels in both directions. Prosecutors frequently amend patent claims during the prosecution process based on the examiner's actions. The examiner's interest is in ensuring that the claims are valid: (1) not anticipated, obvious, or indefinite; and (2) adequately described.

After that initial filing, prosecution of the application and continuations may continue for years. There is often minimal or no interaction between the patent attorney and the inventors during this period, which causes a further drift in nomenclature, which in turn complicates claim construction. (This can lead to the anomalous and surprisingly common situation, many years later, in which a court must construe a claim term that appears nowhere in the specification.) Whereas the inventor may be steeped in the language of his or her field, the patent drafter will use terms from science as well as claim drafting to achieve a delicate balance of clarity, breadth, and flexibility.

Thus, patent claim language can be an amalgam of multiple vocabularies and perspectives. Patent case law instructs courts to interpret patent claims from the perspective of a person having ordinary skill in the art (i.e., the scientist, technologist, or artisan in the relevant field of invention). This characterization, however, glosses over the role of the patent draftsman and the examiner in actual claim drafting practice. Whereas some claim terms—such as “hydroxypropyl, methylcellulose”—undoubtedly derive their meaning from the pertinent technical art, other terms—such as the transitional phrase “comprising”—are better understood from the perspective of the person having ordinary skill in *claim drafting*. Still other terms—which frequently are the focus of the greatest disputes—are simply being used in their plain English sense. Courts need to be sensitive to these distinctions in determining which terms require construction and how individuals skilled in the art interpret those terms.

## 2. *Sources for Deriving Claim Meaning*

Claim construction draws upon two general categories of evidence: intrinsic and extrinsic. Chart 2 summarizes the main components of these sources.

Chart 2: Sources of Evidence for Claim Construction

<p><b>Intrinsic Evidence:</b></p> <ul style="list-style-type: none"><li>• Patent</li><li>• Prosecution History</li><li>• Foreign and Related Patents (and their Prosecution Histories)</li><li>• Prior Art that is cited or incorporated by reference in the Patent-in-Suit and Prosecution History</li></ul> <p><b>Extrinsic Evidence:</b></p> <ul style="list-style-type: none"><li>• Inventor Testimony</li><li>• Expert Testimony</li><li>• Other Documentary Evidence<ul style="list-style-type: none"><li>○ Dictionaries</li><li>○ Treatises</li></ul></li></ul>
--

Prior to the en banc *Phillips* decision, the Federal Circuit doctrine on whether courts should consider extrinsic evidence and what role it should play shifted significantly. From 1996 until 2002, the Federal Circuit heavily disfavored consideration of extrinsic evidence beyond educating the court about the technology.<sup>12</sup> But nearly contemporaneous decisions cautioned against such a strong reading.<sup>13</sup> In 2002, the Federal Circuit appeared to elevate dictionaries, a special category of extrinsic evidence, to a central role in claim construction.<sup>14</sup> Within a short time, however, the limitations of this approach became apparent:

The main problem with elevating the dictionary to such prominence is that it focuses the inquiry on the abstract meaning of words rather than on the meaning of claim terms within the context of the patent. . . . [H]eavy reliance on the dictionary divorced from the intrinsic evidence risks transforming the meaning of the claim term to the artisan into the meaning of the term in the abstract, out of its particular context, which is the specification.<sup>15</sup>

*Phillips* shifted attention back toward the intrinsic record while recognizing that courts could consider extrinsic evidence, although with

---

12. See *Vitronics Corp. v. Conceptronic, Inc.*, 90 F.3d 1576, 1583 (Fed. Cir. 1996) (finding it was “improper to rely on extrinsic evidence”) (emphasis omitted).

13. See, e.g., *Key Pharms. v. Hercon Labs. Corp.*, 161 F.3d 709, 716 (Fed. Cir. 1998) (noting that *Vitronics* “might be misread by some members of the bar as restricting a trial court’s ability to hear [extrinsic] evidence. We intend no such thing.”).

14. See *Tex. Digital Sys., Inc. v. Telegenix, Inc.*, 308 F.3d 1193 (Fed. Cir. 2002).

15. *Phillips*, 415 F.3d at 1321.

healthy skepticism. The court may consider extrinsic evidence if it deems it helpful to “educate [itself] regarding the field of invention . . . [and to] determine what a person of ordinary skill in the art would understand claim terms to mean.”<sup>16</sup> The Federal Circuit emphasized, however, that extrinsic evidence must be considered “in the context of the intrinsic evidence,” but is “less reliable than the patent and its prosecution history in determining how to read claim terms.”<sup>17</sup> Since *Phillips*, the law is clear that intrinsic evidence serves as the principal source for claim construction and that it trumps any extrinsic evidence contradicting it.

a) Principal Source: Intrinsic Evidence

“Intrinsic” evidence refers to the patent and its file history, including any reexaminations and reissues. Intrinsic evidence also includes related patents and their prosecution histories. In addition, the Federal Circuit generally treats the prior art that is cited or incorporated by reference in the patent-in-suit and prosecution history as intrinsic evidence.

i) Specification

The patent specification provides “a written description of the invention, and of the manner and process of making and using it.”<sup>18</sup> It includes the field and background of the invention, the drawings, detailed description of the invention, preferred embodiments, best mode of practicing the invention (although it need not be labeled as such), and the patent claims. Noting “the close kinship between the written description and the claims”<sup>19</sup> as required by the Patent Act, the Federal Circuit in *Phillips* emphasized that claims “must be read in view of the specification, of which they are a part”<sup>20</sup> and that the specification “is always highly relevant to the claim construction analysis. Usually, it is dispositive; it is the single best guide to the meaning of a disputed term.”<sup>21</sup> Where the specification reveals a special meaning to a claim term or an intentional disclaimer, such definition or limitation governs claim construction.<sup>22</sup> It is common and “entirely appropriate for a court, when

---

16. *Id.* at 1319.

17. *Id.* at 1318–19.

18. 35 U.S.C. § 112 (2006).

19. *Phillips*, 415 F.3d at 1316.

20. *Id.* at 1315 (quoting *Markman v. Westview Instruments, Inc.*, 52 F.3d 967, 979 (Fed. Cir. 1995) (en banc), *aff'd*, 517 U.S. 370 (1996)) (internal quotations omitted).

21. *Id.* (quoting *Vitronics Corp. v. Conceptor, Inc.*, 90 F.3d 1576, 1582 (Fed. Cir. 1996)) (internal quotations omitted).

22. *See id.*

conducting claim construction, to rely heavily on the written description for guidance as to the claim's meaning."<sup>23</sup>

ii) Prosecution History

Beyond the specification and other claims, an important source of evidence in claim construction is a patent's prosecution history. A prosecution history "consists of the complete record of the proceedings before the PTO and includes the prior art cited during the examination of the patent."<sup>24</sup> During those exchanges, the Patent Office will commonly reject the pending patent claims as unpatentable in light of prior art technologies. In response, the patent applicants will typically explain why their claimed inventions are patentable over what had come before. The Federal Circuit cautions that "because the prosecution history represents an ongoing negotiation between the PTO and the applicant, rather than the final product of that negotiation, it often lacks the clarity of the specification and thus is less useful for claim construction purposes."<sup>25</sup>

More specifically, the patentee may expressly limit the scope of its patent through disclaimers in order to avoid prior art. Because the inherent tension between validity and infringement issues often plays out in claim construction, it can be particularly illuminating in determining what the claims *do* cover to analyze what the applicant said the claims do *not* cover in order to get the patent issued. However, courts must carefully evaluate such disclaimers, which can be ambiguous, during claim construction.

The communications between the applicant and the Patent Office may reveal the "ordinary meaning" of a claim term—i.e., the communications may show the meaning of a claim term in the context of the patent.<sup>26</sup> For example, in *Nystrom v. TREX Co.*, the prosecution history of the patent confirmed that the claim term "board" referred to wooden boards, and not plastic boards.<sup>27</sup>

iii) Related and Foreign Applications

Some patents issue from a single application, with a single prosecution history. Other patents are members of large families of related patents, with a web of underlying patent applications, along with counterparts filed in foreign countries. In such instances, when one patent is in suit, parties may

---

23. *Id.* at 1317.

24. *Id.*

25. *Id.*

26. *Id.* ("Like the specification, the prosecution history provides evidence of how the PTO and the inventor understood the patent.")

27. 424 F.3d 1136, 1145 (Fed. Cir. 2005).

find statements in its related patents and patent applications, and in its foreign counterparts, that bear on claim construction. To what extent these statements in related filings affect the construction of the patent-in-suit is a common dispute in patent litigation.

Where there are a series of patent applications, with the patent-in-suit issuing from a later-filed application, disputes frequently arise over the implications of statements made during prosecution of an earlier-filed application (i.e., in a “parent” application). The statements in the parent application are most relevant where the earlier statements address common claim terms with the patent being construed.<sup>28</sup> Moreover, where an amendment in a “parent application distinguishes prior art and thereby specifically disclaims a later (though differently worded) limitation in the continuation application,” prosecution disclaimer may apply.<sup>29</sup> The earlier disclaimer may continue to apply throughout a patent family, particularly if the applicants do not later inform the Patent Office that they want to rescind the earlier disclaimer.<sup>30</sup> However, the general rule is that when different claim terms are present in the parent and descendant applications, the earlier statements have no bearing on claim construction.<sup>31</sup>

Statements to foreign patent offices in counterpart filings may be relevant to construing a U.S. patent where the statements made to the foreign office demonstrate the ordinary meaning of a claim term.<sup>32</sup>

---

28. *See* *Advanced Cardiovascular Sys., Inc. v. Medtronic, Inc.*, 265 F.3d 1294, 1305 (Fed. Cir. 2001).

29. *Invitrogen Corp. v. Clontech Labs., Inc.*, 429 F.3d 1052, 1078 (Fed. Cir. 2005).

30. *See* *Hakim v. Cannon Avent Group, PLC*, 479 F.3d 1313, 1318 (Fed. Cir. 2007) (“Although a disclaimer made during prosecution can be rescinded, permitting recapture of the disclaimed scope, the prosecution history must be sufficiently clear to inform the examiner that the previous disclaimer, and the prior art that it was made to avoid, may need to be re-visited.”).

31. *See* *Ventana Med. Sys., Inc. v. Biogenex Labs., Inc.*, 473 F.3d 1173, 1182 (Fed. Cir. 2006) (“[T]he doctrine of prosecution disclaimer generally does not apply when the claim term in the descendant patent uses different language.”); *ResQNet.com, Inc. v. Lansa, Inc.*, 346 F.3d 1374, 1383 (Fed. Cir. 2003) (“Although a parent patent’s prosecution history may inform the claim construction of its descendant, the [parent] patent’s prosecution history is irrelevant to the meaning of this limitation because the two patents do not share the same claim language.”).

32. *See* *Glaxo Group Ltd. v. Ranbaxy Pharms., Inc.*, 262 F.3d 1333, 1337 (Fed. Cir. 2001) (noting that a statement in a related U.K. prosecution history “bolsters this reading” of the claimed “essentially free from crystalline material” limitation in the asserted U.S. patent); *see also* *Tanabe Seiyaku Co., Ltd. v. U.S. Int’l Trade Comm’n*, 109 F.3d 726, 733 (Fed. Cir. 1997) (“In the present case, the representations made to foreign patent offices are relevant to determine whether a person skilled in the art would consider butanone or other ketones to be interchangeable with acetone in Tanabe’s claimed N-alkylation reaction.”). However, because legal requirements for obtaining a patent in other countries may be unique to those

b) Extrinsic Evidence Permissible, But It May Not Contradict or Override Intrinsic Evidence

“Extrinsic evidence” refers to all other types of evidence, including inventor testimony, expert testimony, dictionaries, and documentary evidence of how the patentee and alleged infringer have used the claim terms. Dictionaries are considered to be “extrinsic” evidence.<sup>33</sup> *Phillips* reaffirmed that the intrinsic evidence is of paramount importance in construing patent claims.<sup>34</sup> Nonetheless, extrinsic evidence can be useful, and *Phillips* confirms that district courts are free to consider extrinsic evidence, including expert testimony, dictionaries, treatises, and other such sources.<sup>35</sup> Litigants continue to argue that it is improper to consider extrinsic evidence in *Markman* rulings, citing *Vitronics Corp. v. Conceptorionics, Inc.*<sup>36</sup> However, the Federal Circuit disavowed any such interpretation of *Vitronics*,<sup>37</sup> and *Phillips* puts to rest any suggestion it is wrong to consider extrinsic evidence.<sup>38</sup>

A key to relying on extrinsic evidence is recognizing its limitations. *Phillips* spells out five reasons why extrinsic evidence is inherently less reliable than intrinsic evidence:

First, extrinsic evidence by definition is not part of the patent and does not have the specification’s virtue of being created at the time of patent prosecution for the purpose of explaining the patent’s scope and meaning. Second, while claims are construed as they would be understood by a hypothetical person of skill in the art,

---

countries, statements made to comply with those requirements are generally disregarded in interpreting a U.S. patent. *See* *Pfizer, Inc. v. Ranbaxy Labs., Ltd.*, 457 F.3d 1284, 1290 (Fed. Cir. 2006) (“[T]he statements made during prosecution of foreign counterparts to the . . . [patent-in-suit] are irrelevant to claim construction because they were made in response to patentability requirements unique to Danish and European law.”).

33. *Phillips v. AWH Corp.*, 415 F.3d 1303, 1318 (Fed. Cir. 2005) (en banc).

34. *Id.* at 1324. The *Phillips* court stated:

[T]here is no magic formula or catechism for conducting claim construction. Nor is the court barred from considering any particular sources or required to analyze sources in any specific sequence, as long as those sources are not used to contradict claim meaning that is unambiguous in light of the intrinsic evidence.

*Id.*

35. *See id.* at 1318.

36. 90 F.3d 1576 (Fed. Cir. 1996).

37. *See, e.g., MBO Labs., Inc. v. Becton, Dickinson & Co.*, 474 F.3d 1323, 1329 (Fed. Cir. 2007) (“Extrinsic evidence—testimony, dictionaries, learned treatises, or other material not part of the public record associated with the patent—may be helpful but is less significant than the intrinsic record in determining the legally operative meaning of claim language.”).

38. *See Phillips*, 415 F.3d at 1318.

extrinsic publications may not be written by or for skilled artisans and therefore may not reflect the understanding of a skilled artisan in the field of the patent. Third, extrinsic evidence consisting of expert reports and testimony is generated at the time of and for the purpose of litigation and thus can suffer from bias that is not present in intrinsic evidence. . . . Fourth, there is a virtually unbounded universe of potential extrinsic evidence of some marginal relevance that could be brought to bear on any claim construction question. . . . Finally, undue reliance on extrinsic evidence poses the risk that it will be used to change the meaning of claims in derogation of the “indisputable public records consisting of the claims, the specification and the prosecution history,” thereby undermining the public notice function of patents.<sup>39</sup>

Thus, courts must always probe expert testimony for bias, and should ensure that any expert’s offered opinion be subject to cross-examination. The chief risk of relying on dictionaries, treatises, and other outside documents is pertinence—there is often a gap between how such outside sources characterize a technology and the way it is presented and claimed in a patent.

Nonetheless, extrinsic evidence is an increasingly important source for claim construction. Extrinsic evidence is inherently factual in nature, undermining the doctrine—traceable to *Cybor Corp. v. FAS Technologies, Inc.*<sup>40</sup>—that claim construction is purely a question of law. The Federal Circuit appears to be on the verge of recognizing, en banc, that claim construction may involve underlying questions of fact, particularly in regard to the assessment of extrinsic evidence.<sup>41</sup>

---

39. *Id.* at 1318–19 (Fed. Cir. 2005) (quoting *Southwall Techs., Inc. v. Cardinal IG Co.*, 54 F.3d 1570, 1578 (Fed. Cir. 1995)).

40. 138 F.3d 1448, 1455 (Fed. Cir. 1998) (en banc).

41. *See Amgen Inc. v. Hoechst Marion Roussel, Inc.*, 469 F.3d 1039, 1041 (Fed. Cir. 2006) (Michel, C.J., dissenting from denial of petition for rehearing en banc) (“I believe the time has come for us to re-examine *Cybor*’s no deference rule. I hope that we will do so at our next opportunity, and I expect we will.”); *id.* at 1043 (Newman, J., dissenting from denial of petition for rehearing en banc) (“And if the meaning is recognized as a case-specific finding of fact, appellate review warrants deference to the trier of fact, a deference here lacking.”); *id.* at 1044 (Rader, J., dissenting from denial of petition for rehearing en banc) (“I urge this court to accord deference to the factual components of the lower court’s claim construction.”); *id.* at 1045 (Gajarsa, Linn, and Dyk, JJ., concurring in denial of petition for rehearing en banc) (stating that reconsideration of *Cybor* may be appropriate in a case “in which the language of the claims, the written description, and the prosecution history on their face did not resolve the question of claim interpretation, and the district court found it necessary to resolve conflicting expert evidence to interpret particular claim terms in the field of the art”); *id.* at 1046 (Moore, J., dissenting from denial of petition for rehearing en banc) (“I dissent because I believe this court should have taken this case en banc to

Thus, the Federal Circuit is likely to formally rule that there is a role for district court fact-finding in the claim construction process, especially with regard to assessing the credibility of competing expert witnesses. In the meantime, it appears that the Federal Circuit may be informally according such deference.<sup>42</sup> Thus, reliance on extrinsic evidence can be an important way for trial courts to bolster the “factual” nature of their findings and promote deferential review on appeal.<sup>43</sup> What follows are some lessons from post-*Phillips* case law as to the appropriate, and inappropriate, roles for extrinsic evidence.

i) Illustrations of Reliance (and Non-Reliance) upon  
Extrinsic Evidence

Where the specification supports two interpretations of a disputed claim, the court can use extrinsic evidence to confirm which interpretation is more consistent with what a person having ordinary skill in the art would have understood at the time of invention. For example, in *Conoco Inc. v. Energy & Environmental International*,<sup>44</sup> the question was whether a “stable” suspension of polymer required sufficient stability to remain suspended when stored for a long period of time, or just stability at the time the suspension was introduced into a pipeline.<sup>45</sup> The court determined from the intrinsic evidence that the appropriate frame of reference was stability at the time the suspension was introduced into the pipeline.<sup>46</sup> The court confirmed its interpretation against the extrinsic evidence, which indicated that all suspensions eventually separate, and thus that the appropriate time frame for assessing stability is at the time the suspension is introduced into the pipeline.<sup>47</sup>

*Tap Pharmaceutical Products, Inc. v. Owl Pharmaceuticals, LLC*<sup>48</sup> is another example of a court using extrinsic evidence to decide between two plausible

---

reconsider its position on deference to district court claim construction articulated in *Cybor . . .*”).

42. See *Ortho-McNeil Pharm., Inc. v. Caraco Pharm. Labs., Ltd.*, 476 F.3d 1321, 1328 (Fed. Cir. 2007) (affirming construction based in part on approval of expert testimony that claim term “about 1:5” means “approximately 1:5, encompassing a range of ratios no greater than 1:3.6 to 1:7.1”).

43. See *Phillips*, 415 F.3d at 1332 (Mayer, C.J., dissenting) (“[W]e are obligated by Rule 52(a) to review the factual findings of the district court that underlie the determination of claim construction for clear error.”).

44. 460 F.3d 1349, 1361–62 (Fed. Cir. 2006).

45. *Id.* at 1361.

46. *Id.* at 1362.

47. *Id.*

48. 419 F.3d 1346 (Fed. Cir. 2005).

interpretations from the specification. *Tap Pharmaceutical* concerned claims to a composition “comprising a copolymer . . . of lactic acid and . . . of glycolic acid.”<sup>49</sup> The question was whether the claims were limited to compositions resulting from a polymerization of lactic acid and glycolic acid, or whether the claims also covered the polymer resulting from cyclic precursors that transformed into lactic acid and glycolic acid during polymerization.<sup>50</sup> The district court properly relied on treatises that recognize that copolymers of lactic acid and glycolic acid can be made either by direct polymerization or by ring opening, and on expert testimony that a person of ordinary skill in the art would use the terms “lactic acid” and “glycolic acid” interchangeably with their cyclic analogs.<sup>51</sup>

Attempts to use extrinsic evidence as the *source* for claim construction are more problematic. Basing the meaning of claim terms on sources external to the patent raises concerns about the notice function of patents. Courts must take special care to ensure that the extrinsic evidence is consistent with the patentee’s own description of the invention. For example, an appropriate use of extrinsic evidence concerned claims to a “scanner,” where the specification contained only one illustrative embodiment having a moving scanner head but lacked a definition of the term scanner.<sup>52</sup> Faced with the question of whether a digital camera qualified as a “scanner,” the court turned to dictionaries and concluded that a scanner required “movement between a scanning element and an object being scanned.”<sup>53</sup> This definition was appropriate because it tracked what the patentee had disclosed in the specification describing a scanner.<sup>54</sup>

In a more tenuous example, the Federal Circuit approved the use of expert testimony to set numeric limits on a claim. The claim concerned a pharmaceutical composition with a ratio of “about 1:5” for two chemical components.<sup>55</sup> The court reviewed the intrinsic evidence, including claims directed to other ratios, and experimentation disclosed in the specification directed to a range of ratios, and credited the testimony of an expert who opined that “about 1:5” meant “a ratio up to and including 1:7.1 and a ratio down to and including 1:3.6.”<sup>56</sup> The Federal Circuit credited the expert

---

49. *Id.* at 1349.

50. *Id.*

51. *Id.* at 1349–50.

52. *Mass. Inst. of Tech. v. Abacus Software*, 462 F.3d 1344, 1351 (Fed. Cir. 2006).

53. *Id.*

54. *Id.* at 1351–52.

55. *Ortho–McNeil Pharm., Inc. v. Caraco Pharm. Labs., Ltd.*, 476 F.3d 1321, 1326–28 (Fed. Cir. 2007).

56. *Id.* at 1328.

testimony, which justified this range as appropriate because it was not statistically different from the claimed ratio of 1:5.<sup>57</sup>

An example of expert testimony that strayed too far afield from the patent disclosures is in *Biagro Western Sales, Inc. v. Grow More, Inc.*,<sup>58</sup> wherein the proffering party sought to use expert testimony to reconceptualize the claims. *Biagro* concerned claims to a fertilizer “wherein said phosphorous-containing acid or salt thereof is present in an amount of about 30 to about 40 weight percent.”<sup>59</sup> The amount of phosphorus-containing acid actually present in the accused fertilizer product did not meet the levels stated in the claim, but the patentee tried to use expert testimony to argue that the amount of phosphorous-containing acid in the claim limitation should be read to refer to a “chemical equivalent amount,” rather than the amount actually present.<sup>60</sup> In support, the patentee cited fertilizer labeling guidelines and standards and expert declarations, asserting that phosphorus levels in fertilizer are measured by chemically equivalent amounts.<sup>61</sup> This evidence was unpersuasive for the trial court or the Federal Circuit, because *Biagro* could not tie its measurement approach to the patent’s own description of the invention.<sup>62</sup>

#### ii) Conclusory Expert Opinions Should Be Disregarded

Parties should ground expert opinions both in the intrinsic evidence and have support in other independent, reliable sources. Where these criteria are lacking, courts should not rely upon these expert opinions. For example, in *Network Commerce, Inc. v. Microsoft Corp.*, a patentee sought a construction based upon its expert declaration that a claimed “download component” need not contain a boot program.<sup>63</sup> The expert declaration failed to explain why quoted passages from the specification supported his opinion, and failed to support the expert’s conclusion with any reference to industry publications or other independent sources. Accordingly, the court properly disregarded the declaration.<sup>64</sup>

---

57. *Id.*

58. 423 F.3d 1296 (Fed. Cir. 2005).

59. *Id.* at 1302.

60. *Id.* at 1304.

61. *Id.* at 1303.

62. *Id.*

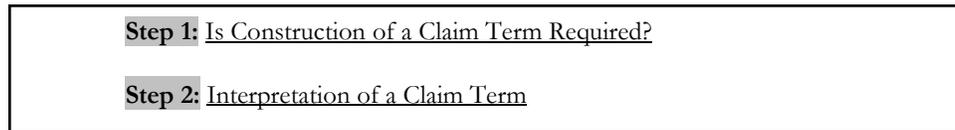
63. 422 F.3d 1353, 1361 (Fed. Cir. 2005).

64. *Id.*

B. A STRUCTURED APPROACH TO CLAIM CONSTRUCTION: TWO STAGES OF ANALYSIS

With that background in place, we are ready to map out the overarching structure of claim construction. Chart 3 presents the two distinct steps. Litigants sometimes skip over the first inquiry—whether (and when) claim construction is necessary—and jump right into the complexities of claim construction. Many courts—through Patent Local Rules<sup>65</sup> or case management—focus attention on the threshold issues. Before the court confronts the challenge of construing a claim term, it must consider a series of threshold doctrines and principles that determine whether construction is required (as well as the proper timing).

Chart 3: Claim Construction Flowchart



1. *Step 1: Is Construction of a Claim Term Required?*

Chart 4 presents the series of threshold issues that the court should consider in determining whether and when interpretation of a claim term is appropriate.

---

65. *See, e.g.*, N.D. CAL. PATENT LOC. R.

**Chart 4: Step 1: Is Construction of a Claim Term Required?**

<p>A. <u>Disputed Meaning that Can Be Derived from the Patent/PHOSTA:</u></p> <ol style="list-style-type: none"> <li>1. <u>Disputed Meaning:</u> Is the meaning of the claim term the subject of legitimate disagreement?</li> <li>2. <u>Meaning Derivable from Patent/PHOSTA:</u> For non-technical terms, is there a special meaning that can be ascertained from the patent?</li> </ol> <p>B. <u>Priority/Discretion/Timing: Courts Have Broad Discretion to Limit and Phase Claim Construction</u></p> <ol style="list-style-type: none"> <li>1. Some courts limit first and usually final <i>Markman</i> proceeding to ten terms.</li> <li>2. Court can revisit claim construction; it must eventually construe all legitimately disputed and construable terms before trial.</li> <li>3. Means + Function claims (in dispute) must be interpreted to identify corresponding structure, material, or acts.</li> </ol> <p>C. <u>Issue Preclusion: Deference to Prior <i>Markman</i> Ruling</u></p> <ol style="list-style-type: none"> <li>1. Issue preclusion cannot be applied offensively against a party not represented in prior proceeding; but it can be applied defensively if four-part test is satisfied. <ol style="list-style-type: none"> <li>i. Judicial estoppel can be applied where patentee changes positions.</li> <li>ii. Reasoned deference under stare decisis principles.</li> </ol> </li> </ol> <p>D. <u>Is the Term Amenable to Construction?</u></p> <ol style="list-style-type: none"> <li>1. See Table A</li> </ol>
--

a) Is There a Genuine Dispute About the Claim Term?

It is all too common for the parties to propose differing construction but be unable to articulate why the differences matter. Courts generally order a structured meet-and-confer process to address this problem and thereby narrow the number of claim terms requiring the court's resolution.<sup>66</sup> Holding a brief telephone conference prior to claim construction briefing at which the parties must articulate the basis for the dispute often narrows the number of terms further.

b) Would Claim Construction of the Term Help the Jury?

The point of claim construction is to instruct the jury on what the claim means from the perspective of a person having ordinary skill in the art. For many claim terms, attempting to "construe" the claim language adds little in the way of clarity. Where the perspective of a person having ordinary skill in

---

<sup>66</sup> See PETER S. MENELL, LYNN H. PASAHOW, JAMES POOLEY & MATTHEW D. POWERS, PATENT CASE MANAGEMENT JUDICIAL GUIDE § 2.1.1 (2009) [hereinafter PCMJG]; N.D. CAL. PATENT LOC. R.; *infra* Section III.A.

the art would add nothing to the analysis, there may be no need to construe the terms. Thus, non-technical terms (e.g., “on” or “above” or “surround”) and terms of degree (e.g., “approximately” or “about” or “substantially”) may not require construction by the court. Where “construing” a claim term would involve simply substituting a synonym for the claim term, it may be appropriate to allow the claim language to speak for itself.

Construction of a term is clearly appropriate in the case of technical terms, i.e., where a typical juror would not understand the term without assistance. Of course, in all cases, where the intrinsic and applicable extrinsic evidence provide further meaning to a term (such as disclaimers, descriptions of “the present invention,” and claim differentiation), the court should account for such added evidence in the claim construction. But where the intrinsic evidence and extrinsic evidence do not meaningfully add to the definition of a term, it is appropriate (and often preferred) to allow straightforward claim language to stand as-is.

c) Is Claim Construction of the Term a Priority?

Courts need not construe all of the terms in the initial *Markman* hearing. Indeed, courts increasingly focus the initial *Markman* hearing on about ten “priority” terms, with the expectation that resolving the key terms may dispose of the case. Courts are free to revisit any remaining disputes later in the case, but are required to construe all disputed claim terms before the case is submitted to the jury. How courts wish to balance the priorities of early decision-making, versus overall completeness, will depend on the circumstances of the case.

d) Has the Term Been Construed Before?

There may be prior proceedings involving the same patents-in-suit or closely related patents. Where a proceeding previously construed the term, the court needs to learn the context of the prior proceedings to determine the impact of doctrines of issue preclusion, claim preclusion, judicial estoppel, and stare decisis. Although the prior proceedings may not be binding in the present litigation, the court should hear from parties to determine the factors that determine any preclusive effect or basis for according deference to the prior claim construction.<sup>67</sup>

Similarly, in the increasingly common scenario where the patent-in-suit becomes the subject of patent reexamination proceedings, the district court

---

67. See *infra* Section II.E.

may wish to stay claim construction until the patent examiner resolves the collateral proceedings.

e) Is the Term Amenable to Construction?

As illustrated in Table A, claim terms can usefully be categorized among three potentially overlapping general types: (1) lay terms; (2) terms of degree; and (3) technical terms (including seemingly lay terms which have a different meaning in a technical context). As discussed previously,<sup>68</sup> not all terms in a claim require construction by the court. It can be improper to construe terms that do not have special meaning that can be derived from the patent. A fourth category—means-plus-function claim terms—must be construed by the court if the parties dispute their meaning so as to determine corresponding structure, materials, or acts from the specification.<sup>69</sup>

**Table A: Typology of Claim Terms**

Type	Lay Terms	Terms of Degree	Technical Terms
Examples	a, above, below, in, surround, to	approximately, essentially, substantial, dose	hydroxypropyl, methylcellulose, cyclic redundancy, oligonucleotide
Amenability to Claim Construction	Such terms are often understood by fact-finder; to construe arguably trenches upon jury's domain. But such terms may have conventional or established meaning in the technical field.	Such terms are often understood by jury; to construe arguably trenches upon jury's domain. Such terms are inherently contextual. Must be careful not to inappropriately import limitations from specification. But must base interpretation on standard set forth in the spec.: if no basis set forth in spec., then no basis for construction.	Must be interpreted if meaning is disputed; PHOSITA perspective is essential.

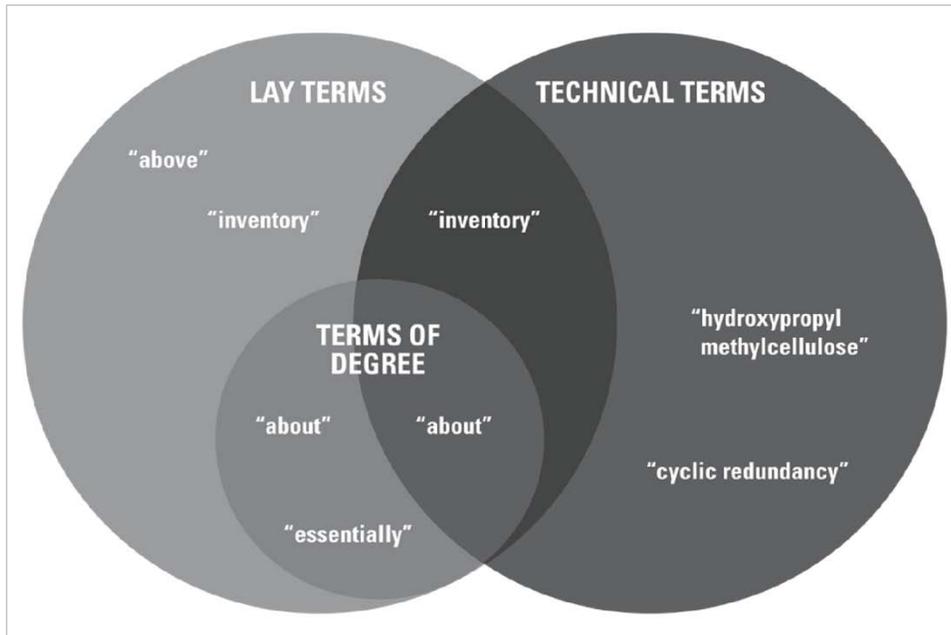
As reflected in Chart 5, the three types of claim terms are not mutually exclusive and the question of which category is most appropriate will not always be evident based solely on a reading of the claim. The court will need to examine the intrinsic record in making this assessment. Some plain

68. *See supra* Section II.B.1.

69. *See infra* Section II.C.

English terms can have technical meanings in particular fields. For example, the word “inventory” can, depending upon on the context, be considered a lay term (“an itemized list of merchandise or supplies” or a “detailed list of all items in stock”) as well as more specialized meaning in the fields of dry cleaning process inventions.<sup>70</sup>

Chart 5: Typology of Claim Terms



Some technical terms, such as “hydroxypropyl methylcellulose,” may well be self-evident. Terms of degree, however, can be ambiguous. For example, the word “about” can obviously have a non-technical meaning. But when used in describing the scope of a particular invention, it may well take on meaning that is delimited by intrinsic, and possibly even extrinsic, evidence.<sup>71</sup>

i) Lay Terms

Patent law has long struggled with how precisely claims should be construed. Many claim terms are inherently imprecise. These include terms of degree, such as “substantially,” “about,” and “approximately,” which we deal with separately below because they have been the focus of substantial

70. See *Markman v. Westview Instruments, Inc.*, 52 F.3d 967, 973 (Fed. Cir. 1995) (en banc) (interpreting “inventory” as used in patent claim to mean “articles of clothing” rather than cash or inventory receipts), *aff’d*, 517 U.S. 370 (1996).

71. See *Ortho-McNeil Pharm., Inc. v. Caraco Pharm. Labs., Ltd.*, 476 F.3d 1321, 1326–28 (Fed. Cir. 2007).

jurisprudence. District courts are commonly asked to give lay terms additional clarity in claim construction. When imprecise language should be left to the jury remains a subtle, confounding, and thorny aspect of patent adjudication.

Efforts to construe lay terms with precision are in tension with *Markman*'s division of authority between judges and juries.<sup>72</sup> It is the court's role to construe the claims, while it is the jury's role to apply that construction to an accused device or piece of prior art. That is, "Step 1" of the infringement analysis is to construe the claims, and "Step 2" is to apply the construed claims to a specific set of facts. Construing terms of degree with more precise language may be error, not only because it "imports limitations" from the specification into the claims, but also because it can impinge on the role of the jury in resolving the question of infringement or validity. The Federal Circuit has recently observed that "line-drawing" questions over what meets the scope of the claims is appropriately left to the jury in some contexts.<sup>73</sup>

On the other hand, the Federal Circuit's decision in *O2 Micro International Ltd. v. Beyond Innovation Technology Co.* states that although "district courts are not (and should not be) required to construe *every* limitation present in a patent's asserted claims,"<sup>74</sup> the court must interpret the scope of any claim term for which the parties have presented a "fundamental dispute."<sup>75</sup> In that case, the district court had declined to construe the term "only if" on the ground that it has a well-understood meaning that is capable of application by the jury without judicial interpretation. The parties in the case agreed that "only if" had a common meaning, but the parties disputed the scope of the claim based on this phrase and argued that dispute to the jury. The Federal Circuit vacated the jury verdict and permanent injunction and remanded the case for reconsideration. If this decision remains valid, the prudent course for district courts will be to construe any claim term—including lay words or phrases—for which there is a legitimate dispute. Nonetheless, courts should be skeptical of construing lay terms for which neither party can produce intrinsic evidence indicating a specialized meaning.

---

72. See *Markman v. Westview Instruments, Inc.*, 517 U.S. 370, 384 (1996).

73. *Acumed LLC v. Stryker Corp.*, 483 F.3d 800, 806 (Fed. Cir. 2007) ("[A] sound claim construction need not always purge every shred of ambiguity. The resolution of some line-drawing problems—especially easy ones like this one—is properly left to the trier of fact.")

74. 521 F.3d 1351, 1362 (Fed. Cir. 2008) (emphasis in original).

75. *Id.*

## ii) Terms of Degree

Determining how far courts should go in construing lay terms arises with particular frequency in the context of terms of degree, such as “about,” “approximately,” and “essentially.” The issues are whether such words are used in a technical sense or otherwise derive meaning from the specification.

When construing a term of degree, a key question is whether the intrinsic evidence provides some standard for measuring that degree.<sup>76</sup> Often there may be no such standard, and the Federal Circuit has frequently ruled that it would be error to impose a more exact construction on terms of degree.<sup>77</sup>

A standard for measuring a term of degree may come from the patent specification and the working examples. As noted above, a recent case concerns construction of the term “about 1:5,” referring to a pharmaceutical composition having a particular ratio of two components.<sup>78</sup> The Federal Circuit approved its construction as “a ratio up to and including 1:7.1 and a ratio down to and including 1:3.6.”<sup>79</sup> This construction was derived from the specification, which contained other examples of ratios that were tested and claimed, and from expert testimony, declaring that a range of 1:7.1 and a ratio down to and including 1:3.6 was not statistically different from the stated ratio of 1:5.<sup>80</sup> This case may represent the high water mark in terms of extrapolating examples from the specification and imposing numerical limits on claim scope, and may suggest willingness to credit district court fact-finding based on extrinsic evidence. By contrast, other cases have refused to assign numerical bounds to the scope of the claim term “about.”<sup>81</sup>

---

76. *Exxon Research & Eng'g Co. v. United States*, 265 F.3d 1371, 1381 (Fed. Cir. 2001) (“When a word of degree is used the district court must determine whether the patent’s specification provides some standard for measuring that degree.”).

77. *See, e.g., Playtex Prods., Inc. v. Procter & Gamble Co.*, 400 F.3d 901, 907 (Fed. Cir. 2005) (“[T]he definition of ‘substantially flattened surfaces’ adopted by the district court introduces a numerical tolerance to the flatness of the gripping area surfaces of the claimed applicator. That reading contradicts the recent precedent of this court, interpreting such terms of degree.”) (citing *Cordis Corp. v. Medtronic AVE, Inc.*, 339 F.3d 1352, 1361 (Fed. Cir. 2003) and *Anchor Wall Sys., Inc. v. Rockwood Retaining Walls, Inc.*, 340 F.3d 1298, 1311 (Fed. Cir. 2003)).

78. *Ortho–McNeil Pharm., Inc. v. Caraco Pharm. Labs., Ltd.*, 476 F.3d 1321, 1326 (Fed. Cir. 2007).

79. *Id.* at 1328.

80. *Id.*

81. *See Modine Mfg. Co. v. U.S. Int’l Trade Comm’n*, 75 F.3d 1545, 1551, 1554 (Fed. Cir. 1996) (stating that “[i]t is usually incorrect to read numerical precision into a claim from which it is absent” because “it is a question of technologic fact whether the accused device meets a reasonable meaning of ‘about’ in the particular circumstances”), *abrogated on other grounds by Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 234 F.3d 558 (Fed. Cir. 2000) (en banc).

A standard for measuring a term of degree may come from the applicant's statements distinguishing the prior art. For example, in *Glaxo Group Ltd. v. Ranbaxy Pharmaceuticals, Inc.*,<sup>82</sup> the Federal Circuit found that the claim phrase "essentially free of crystalline material" could be properly construed as requiring a crystalline content of less than ten percent, based in part on the applicant's statements describing the prior art. Similarly, in *Biotec Biologische Naturverpackungen GmbH & Co. KG v. Biocorp, Inc.*,<sup>83</sup> the Federal Circuit affirmed the district court's construction of the term "substantially water free" as having a water content below five percent, finding that the court below may properly rely on the applicant's statements distinguishing prior art having a water content from five to thirty percent water content during prosecution history.

Terms of degree frequently do not warrant a more precise construction, and it is often appropriate to pass imprecise terms to the jury in its role as fact-finder. However, intrinsic evidence may suggest an appropriate standard for providing a more concrete measure of claim scope. The right approach is one that recognizes the tension between the goals of clarifying claim scope and of avoiding imposing extra limitations on claim language, and then carefully assesses the objective measures that can be used to give standards for the claim terms.

### iii) Technical Terms

The easiest call relates to technical terms. When these are disputed, there is no doubt that construction by the court is required. As reflected in Chart 5, however, some lay terms, such as "about," might have a technical meaning in the context of the patent and hence will require interpretation by the court.<sup>84</sup>

## 2. Step 2: Interpretation of Claim Language

### a) General Framework

Once it is determined that claim language must be construed and is ripe for construction, the court must then apply the various substantive rules to the claim language to arrive at the proper construction. Before discussing the disputes that commonly arise in claim construction, it will be useful to state the principles that are generally *not* in dispute. The *Phillips* en banc decision is

---

82. 262 F.3d 1333, 1337 (Fed. Cir. 2001).

83. 249 F.3d 1341, 1346 (Fed. Cir. 2001).

84. See *O2 Micro Int'l Ltd. v. Beyond Innovation Tech. Co.*, 521 F.3d 1351, 1362 (Fed. Cir. 2008) (holding that failure to construe the term "only if" was error where parties engaged in technical dispute over its scope).

the most recent and authoritative attempt by the Federal Circuit to distill these principles.

“A ‘bedrock principle’ of patent law [is] that ‘the claims of a patent define the invention to which the patentee is entitled the right to exclude.’”<sup>85</sup> Courts must interpret claims from the perspective of “how a person of ordinary skill in the art understands a claim term . . . in the context of the entire patent.”<sup>86</sup> This frame of reference “is based on the well-settled understanding that inventors are typically persons skilled in the field of the invention and that patents are addressed to and intended to be read by others of skill in the pertinent art.”<sup>87</sup> Often, other evidence will provide context for characterizing the person having ordinary skill in the art. Indeed, courts look to what the meaning of the term would have to a person of ordinary skill in the art “at the time of the invention, i.e., as of the effective filing date of the patent application.”<sup>88</sup> The “effective filing date” is the earlier of the actual filing date or the filing date of an application from which priority is accorded. This is quite significant (and can generate evidentiary challenges) because the meaning of scientific and technical terms can change significantly during the life span of a patent. In the field of digital technology, for example, change can occur unbelievably rapidly given the exponential rate of advance in computer technology. Litigation over patent claims can occur multiple technological generations after the patent claim term was drafted.

Claim interpretation is highly context-dependent. The person of ordinary skill in the art “is deemed to read the words used in the patent documents with an understanding of their meaning in the field, and to have knowledge of any special meaning and usage in the field.”<sup>89</sup> The meaning that this person would give to claim language, after having considered the intrinsic and extrinsic evidence, is the “ordinary meaning” of the claim terms. This ordinary meaning is considered to be the “objective baseline” for claim construction. Thus in interpreting patent claims, a court must consider “the same resources as would [a person in the same field of technology] viz., the patent specification and the prosecution history.”<sup>90</sup> The patent and its prosecution history “usually provide[] the technological and temporal context to enable the court to ascertain the meaning of the claim to one of ordinary

---

85. *Phillips v. AWH Corp.*, 415 F.3d 1303, 1312 (Fed. Cir. 2005) (en banc).

86. *Id.* at 1313.

87. *Id.*

88. *Id.*

89. *Id.* (quoting *Multiform Desiccants, Inc. v. Medzam Ltd.*, 133 F.3d 1473, 1477 (Fed. Cir. 1998)).

90. *Id.* (quoting *Multiform Desiccants*, 133 F.3d at 1477).

skill in the art at the time of the invention.”<sup>91</sup> Thus, courts should interpret patent claims in light of this “intrinsic” evidence (i.e., the patent specification and its prosecution history) as well as pertinent “extrinsic” evidence (i.e., evidence showing the usage of the terms in the field of art).

b) Claim Construction Methodology

As noted above, *Phillips* holds that the “ordinary meaning” of a claim term is the “objective baseline” for construing patent claims. The court must adopt this perspective when interpreting claim language. The phrase “ordinary meaning” is deeply engrained in the case law, but it is a slippery concept. The “ordinary meaning” of a term is what a court arrives at after doing the work of reviewing the specification, the other claims, the file history, the cited prior art, and the pertinent extrinsic evidence. Thus, the “ordinary meaning” is not the first step in the analysis. Nor is it the endpoint, as *Phillips* and its progeny have confirmed—the proper construction is frequently *not* a term’s ordinary meaning. Thus *Phillips*’ identification of ordinary meaning as the “objective baseline” puts tremendous emphasis on this term, which can create unfortunate confusion and error.

Focusing on “ordinary meaning” has other shortcomings. The term “ordinary meaning” tends to drive the claim construction analysis to the meaning of a single word, or at most to a short phrase. But atomizing the dispute down to a word, or a short phrase, often does not make sense. Most patent disputes go to the overall approach of a patent claim, and focusing on a single word tends to lose the forest for the trees. When the overall approach of a patented invention is the central issue in a patent case, there may be no “ordinary meaning” that attaches. Trying to boil down the overall approach of an invention to a few selected words often misses the point of the dispute. There is a real danger that resolving a dispute over the meaning of a particular claim term will be mistaken for a resolution on the merits of a more fundamental infringement or validity dispute.

A more simple and useful description of the claim construction process starts with the “initial understanding” of claim language. This is the understanding that comes from the first reading of the claims, and from getting a sense as to what the patentee is trying to claim. This “initial understanding” may be focused on a particular claim term of interest, or may take into consideration larger blocks of claim text. The endpoint of the analysis is the “proper construction.” Between this starting point and this

---

91. *Id.* (quoting *V-Formation, Inc. v. Benneton Group SpA*, 401 F.3d 1307, 1310 (Fed. Cir. 2005)).

ending point, is an analytical framework represented by the black box shown in Chart 6.

**Chart 6: Claim Construction Process: Starting Point and Destination**

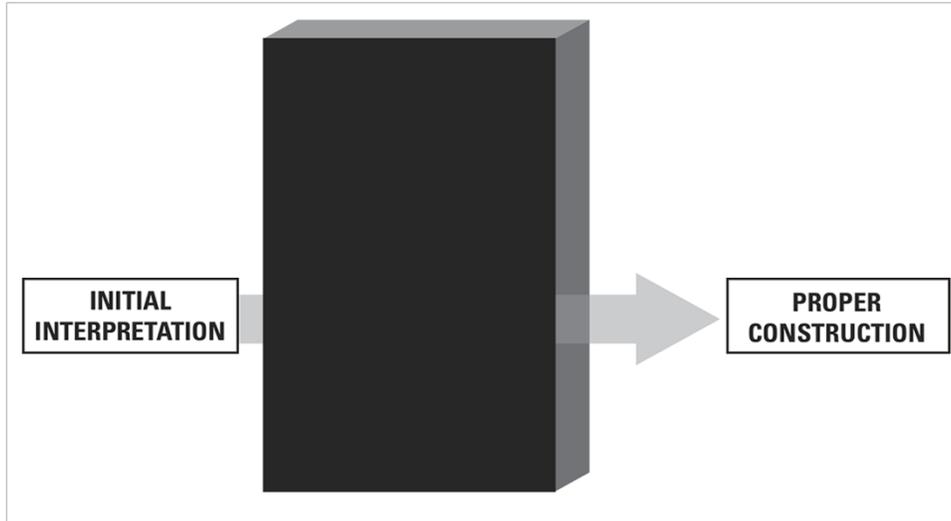
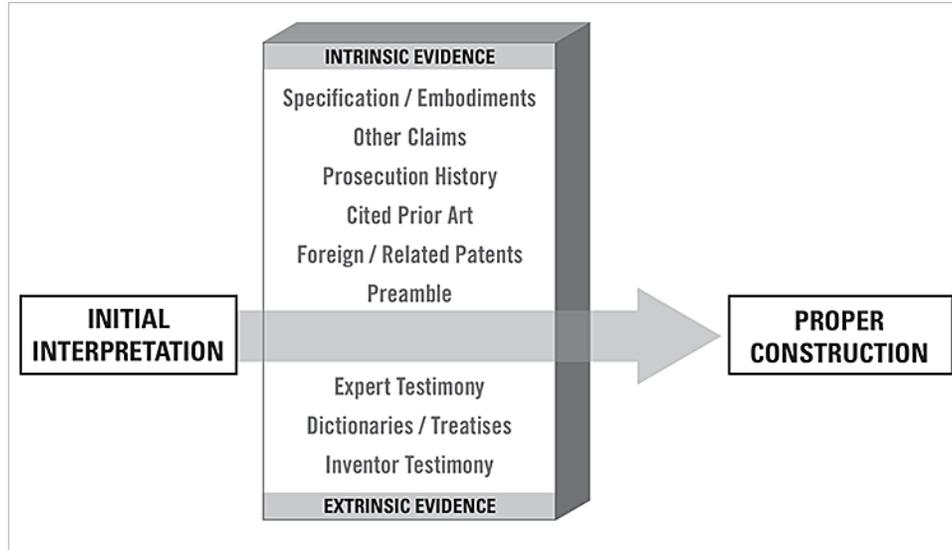


Chart 6 illustrates the starting and ending points for claim construction. The first step is to consider the claim itself, and to account for the initial understanding the court ascribes to it. If the claim language employs common, non-technical language, then its scope will immediately begin to take on meaning. If the claim language term is technical, the court may ascribe little if any meaning to the term without further review of the patent and surrounding evidence.

The ultimate destination for this process is the “proper construction.” Arriving at the proper construction requires filtering the claim language at issue through a number of rules of claim construction, taking into consideration the pertinent statements in the intrinsic and extrinsic evidence. This process requires that the court view the evidence from the appropriate perspective of a person of ordinary skill in the art from the relevant time period. The court should take into consideration the doctrine of claim differentiation, the rules for reviewing the specification for meanings of claim terms, prosecution history estoppel, and a review of related patents. The various rules that the court must take into analysis are sometimes contradictory, and typically involve a balancing of considerations. Chart 7 illustrates the principal points of analysis.

Chart 7: Claim Construction Process: Inside the Black Box



In Chart 7, the various factors that govern claim construction are vertically aligned in roughly the order of persuasiveness, with intrinsic evidence at the top, and extrinsic evidence below. The Federal Circuit has often emphasized, and the *Phillips* decision affirms, that the specification is the “primary basis for construing the claims”<sup>92</sup> and is in most cases “the best source for understanding a technical term.”<sup>93</sup> However, no fixed hierarchy of claim construction rules exists:

[T]here is no magic formula or catechism for conducting claim construction. Nor is the court barred from considering any particular sources or required to analyze sources in any specific sequence, as long as those sources are not used to contradict claim meaning that is unambiguous in light of the intrinsic evidence.<sup>94</sup>

The parties’ briefing will inform the court which sources of evidence are most relevant to interpreting the claim and what specific evidence bears on the proposed interpretation. If no evidence is adduced or if the evidence cited is not illuminating, then the court’s initial interpretation will probably be the proper construction. More commonly, the parties will call attention to various sources of meaning from the specification, file wrapper, or extrinsic sources.

92. *Id.* at 1315 (quoting *Standard Oil Co. v. Am. Cyanamid Co.*, 774 F.2d 448, 452 (Fed. Cir. 1985)).

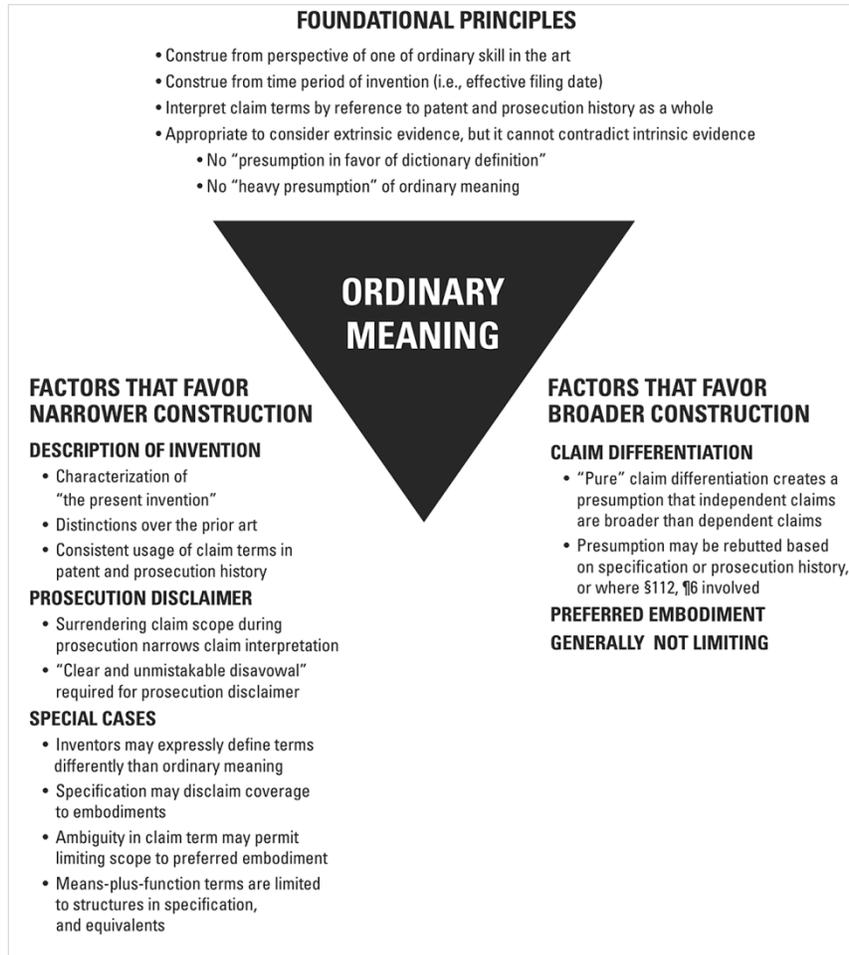
93. *Id.* (quoting *Multiform Desiccants*, 133 F.3d at 1477).

94. *Id.* at 1324.

Note that the term “ordinary meaning” is not reflected in Chart 7. This viewpoint is not the first step in the analysis, and it is not the endpoint. It is a helpful reference point, and probably occurs somewhere along the path. The “ordinary meaning” might be determined after doing the work of reviewing the pertinent intrinsic and extrinsic evidence, but before the final construction is rendered.

This ordinary meaning then might be found to be the proper construction, or the proper construction may be broader, or narrower, than the ordinary meaning based on the application of the various claim construction doctrines. Some of these doctrines tend to narrow claim scope, while others broaden it. These doctrines push and pull on the concept of “ordinary meaning,” and drive the final construction. Chart 8 reflects this dynamic. The principles set forth at the top of the chart are foundational principles of claim construction which ground the inquiry. The factors on the left tend to narrow the construction (but may in some cases broaden it), and the factors on the right tend to broaden the construction:

Chart 8: Functional Landscape of Claim Construction Principles and Doctrines



The Appendix to this Article provides a chart illustrating cases that narrow or broaden ordinary meaning based upon the various doctrines in play.

c) Misuse of “Ordinary Meaning”

*Phillips*’ main contribution to claim construction law was reining in the *Texas Digital* line of cases.<sup>95</sup> *Texas Digital* and its progeny had put undue emphasis on dictionaries as defining the “ordinary meaning” of claim terms. *Texas Digital* established a “heavy presumption” that the “ordinary meaning”

95. See *Nystrom v. TREX Co.*, 424 F.3d 1136, 1142–43 (Fed. Cir. 2005); *Inverness Med. Switz. GmbH v. Warner Lambert Co.*, 309 F.3d 1373, 1378 (Fed. Cir. 2002); *Tex. Digital Sys., Inc. v. Telegenix, Inc.*, 308 F.3d 1193 (Fed. Cir. 2002).

from dictionaries applies, and that this presumption could only be overcome by explicit definitions in the specification, or by clear disavowals of claim scope.<sup>96</sup> Following *Texas Digital*, the Federal Circuit routinely referred to a “heavy presumption of ordinary meaning,” which became a mantra in the years leading up to *Phillips*.

*Phillips* explicitly rejected the language in *Texas Digital* that had been interpreted as elevating dictionary definitions above statements in the patent documents. This was an important clarification of claim construction law, and has largely succeeded in putting to rest *Texas Digital*'s over-emphasis on dictionaries. However, *Phillips* was perhaps not as clear as it could have been in silencing the *Texas Digital*-era statement that there is a “heavy presumption of ordinary meaning.” Lawyers and district courts have largely overlooked an important and fundamental shift in Federal Circuit law that has emerged since *Phillips*. Whereas the Federal Circuit routinely referred to this “heavy presumption of ordinary meaning” prior to *Phillips*, this “heavy presumption” is all but gone from the Federal Circuit's opinions. Indeed, since *Phillips* issued, the Federal Circuit has referred to this “heavy presumption of ordinary meaning” on only two occasions, which may be viewed as outliers, and which themselves rely on pre-*Phillips* law.<sup>97</sup> This appears to have been a deliberate shift by the Federal Circuit to drop a powerful presumption from claim construction law. This important change in Federal Circuit law has gone largely unnoticed.

It is unfortunate that the Federal Circuit has failed to expressly disavow the “heavy presumption of ordinary meaning.” Lawyers have persisted in citing pre-*Phillips* case law to argue this standard, and district courts have all-too-frequently adopted this obsolete rule. The result is that many district courts are unduly wedded to what they perceive to be the “ordinary meaning” of a claim term. As the Federal Circuit's post-*Phillips* case law makes clear, courts may depart from ordinary meaning in arriving at the proper construction. It is appropriate to depart from the “ordinary” meaning where the intrinsic evidence persuasively demonstrates “what the inventors actually invented and intended to envelop with the claim.”<sup>98</sup> In sum, “[t]he construction that stays true to the claim language and most naturally aligns with the patent's description of the invention will be, in the end, the correct

---

96. 308 F.3d at 1202.

97. *Epistar Corp. v. U.S. Int'l Trade Comm'n*, 566 F.3d 1321, 1334 (Fed. Cir. 2009); *Elbex Video, Ltd. v. Sensormatic Elecs. Corp.*, 508 F.3d 1366, 1371 (Fed. Cir. 2007).

98. *Phillips*, 415 F.3d at 1316 (quoting *Renishaw PLC v. Marposs Societa' Per Azioni*, 158 F.3d 1243, 1250 (Fed. Cir. 1998)).

construction.”<sup>99</sup> This standard is lower than the “explicit definition” or “clear disavowal” standard that the court used to insist upon for deviating from ordinary meaning.

d) Interpreting Claim Language in Light of the Specification

A fundamental challenge in patent law is how to construe claims “in view of the specification.”<sup>100</sup> Tension arises from the competing principles that provide, on the one hand, that “the claims made in the patent are the sole measure of the grant,”<sup>101</sup> and, on the other hand, that a claim term “can be defined only in a way that comports with the instrument as a whole.”<sup>102</sup> When, and to what extent, the terse wording of patent claims should be interpreted in light of the inventor’s other statements in the specification gives rise to a common tension in patent litigation. Indeed, *Phillips* arose out of precisely this type of dispute. And since *Phillips*, the Federal Circuit continues to acknowledge the “tightrope” that district courts must walk when construing claims in light of the specification.<sup>103</sup>

There are several common sources of meaning for claim construction: the preferred embodiments; the manner in which the patentee distinguishes the prior art; the usage of the claim term elsewhere in the patent document (including other claims); disclaimers within the prosecution history; and the preamble. Furthermore, as explored in Section II.B.3, *supra*, some commonly used claim terms have developed greater clarity through patent drafting convention and judicial decisions.

i) The Role of Preferred Embodiments in Claim Construction

Patent specifications typically describe the claimed invention through the use of illustrations or examples. In the terminology of patent law, they are characterized as “preferred embodiments.” Often the specification will recite a few or even many preferred embodiments of an invention. Claim construction disputes often center on the importance of such illustrations: (1) Must each claim encompass the preferred embodiments?; (2) Are the claims limited to the preferred embodiments?; (3) Does the number or range of

---

99. *Id.*

100. *Id.* at 1315 (quoting *Markman v. Westview Instruments, Inc.*, 52 F.3d 967, 973 (Fed. Cir. 1995) (en banc), *aff’d*, 517 U.S. 370 (1996)).

101. *Id.* at 1312 (quoting *Aro Mfg. Co. v. Convertible Top Replacement Co.*, 365 U.S. 336, 339 (1961)).

102. *Id.* at 1316 (quoting *Markman v. Westview Instruments, Inc.*, 517 U.S. 370, 389 (1996)).

103. *Andersen Corp. v. Fiber Composites, LLC*, 474 F.3d 1361, 1373 (Fed. Cir. 2007).

embodiments affect the breadth of the claims?; (4) Does ambiguity in a claim term limit its scope to the preferred embodiments?; (5) Do characterizations of embodiments as “the invention” or “the present invention” limit the patent accordingly?; (6) Does the patent distinguish over the prior art in a way that defines the invention?; and (7) Does the patent provide a consistent usage of claim terms to clarify their meaning?

(1) *Claim Scope Generally Includes Preferred Embodiments*

The patent claims should generally be construed to encompass the preferred embodiments described in the specification, and it is generally error to adopt a construction that excludes them.<sup>104</sup> Important exceptions to this oft-cited rule apply: where there is a disclaimer in the specification or prosecution history;<sup>105</sup> an embodiment is directed to only a subset of claims;<sup>106</sup> the claims evolved significantly during prosecution; or the ordinary meaning simply cannot be stretched to encompass the embodiment.<sup>107</sup>

There are two primary scenarios in which a claim can properly be construed in a way that excludes an embodiment: (1) where a change occurs in the file history—i.e., the specification remains static during prosecution but the applicant disclaims some claim scope that she originally sought during prosecution; and (2) where the specification contains and claims multiple embodiments, a particular claim may not cover a particular embodiment because other claims do.

(2) *Is the Patent Limited to the Preferred Embodiments?*

A common dispute is whether the claim scope should be limited to the embodiments. The mere fact of a particular embodiment being taught (or even “preferred”) is generally not sufficient to justify limiting otherwise broad claim scope to the particular embodiment taught.<sup>108</sup> The mere fact that

---

104. *See* On-Line Techs., Inc. v. Bodenseewerk Perkin-Elmer GmbH, 386 F.3d 1133, 1138 (Fed. Cir. 2004) (“[A] claim interpretation that excludes a preferred embodiment from the scope of the claim ‘is rarely, if ever, correct.’” (quoted by MBO Labs., Inc. v. Becton, Dickinson & Co., 474 F.3d 1323, 1333 (Fed. Cir. 2007))).

105. *See* Oatey Co. v. IPS Corp., 514 F.3d 1271, 1277 (Fed. Cir. 2008); N. Am. Container, Inc. v. Plastipak Packaging, Inc., 415 F.3d 1335, 1345–46 (Fed. Cir. 2005); SciMed Life Sys., Inc. v. Advanced Cardiovascular Sys., Inc., 242 F.3d 1337, 1344 (Fed. Cir. 2001); *see also infra* Section II.B.2.e.

106. *See* Helmsderfer v. Bobrick Washroom Equip., Inc., 527 F.3d 1379, 1383 (Fed. Cir. 2008).

107. *See id.*

108. *See, e.g.*, Acumed LLC v. Stryker Corp., 483 F.3d 800, 807–08 (Fed. Cir. 2007) (finding that a claimed “transverse” hole in a bone nail was not limited to the particular

the disclosed embodiments of a patented invention have a certain feature does not, by itself, justify limiting the scope of the claims to what is disclosed in the specification. Rather, the fact that the preferred embodiment teaches a certain configuration is just one factor that must be weighed, along with other factors such as the clarity of the claim language, the specification's descriptions of the claimed invention, its statements distinguishing the invention from the prior art, and the consistent and uniform usage of claim terms. Other contributing factors include the applicant's statements to the Patent Office during patent prosecution and the doctrine of claim differentiation. Depending on the strength of these other factors, the scale may tip so that the claim is limited to the embodiment disclosed in the specification.

The *Phillips* court acknowledged that “there is sometimes a fine line between reading a claim in light of the specification, and reading a limitation into the claim from the specification.”<sup>109</sup> The Federal Circuit suggested that courts can reasonably and predictably discern this line by focusing on how a person of ordinary skill in the art would understand the claim terms.<sup>110</sup> The Federal Circuit has specifically rejected the contention that a court interpreting a patent with only one embodiment must limit the claims of that patent to that embodiment.<sup>111</sup>

The patentee may use the specification in two different ways: (1) illustration—to set out specific examples of the invention to disclose how to make and use it; or (2) limitative—to indicate that the claims and embodiments are strictly coextensive.<sup>112</sup> Nonetheless, contrary to the suggestion in *Phillips*,<sup>113</sup> claim drafters routinely avoid providing a clear distinction between embodiments that define the invention as opposed to

---

“perpendicular” orientation shown in the specification); *Ormco Corp. v. Align Tech., Inc.*, 463 F.3d 1299, 1306–07 (Fed. Cir. 2006) (finding that a claimed “geometry” of orthodontic teeth was not limited to the geometries of orthodontics shown in the specification); *Agfa Corp. v. Creo Prods., Inc.*, 451 F.3d 1366, 1375–76 (Fed. Cir. 2006) (finding that a claimed “stack” of printing plates was not limited to the particular horizontal stack shown in the specification).

109. *Phillips v. AWH Corp.*, 415 F.3d 1303, 1323 (Fed. Cir. 2005) (en banc) (quoting *Comark Commc'ns, Inc. v. Harris Corp.*, 156 F.3d 1182, 1186–87 (Fed. Cir. 1998)).

110. *Id.*

111. *Id.* (citing *Gemstar-TV Guide Int'l, Inc. v. Int'l Trade Comm'n*, 383 F.3d 1352, 1366 (Fed. Cir. 2004)).

112. *Id.* at 1323.

113. *Id.* (“Much of the time, upon reading the specification in that context, it will become clear whether the patentee is setting out specific examples of the invention to accomplish those goals, or whether the patentee instead intends for the claims and the embodiments in the specification to be strictly coextensive.”).

merely illustrating it so as to preserve later flexibility regarding patent scope. In doing so, they hope to get the benefit of a narrow interpretation during prosecution (which may enhance the chances of allowance) while preserving the option of asserting a broad interpretation after the patent issues in enforcement actions. Thus, the “fine line” to which the Federal Circuit refers is often blurred.

(3) *Does the Number or Range of Embodiments Affect the Scope of the Claims?*

The patent drafter’s choice of language gives rise to disputes over how broadly to construe claims in light of the specification. The patent drafter is the “least cost avoider” in terms of creating a document that can be readily understood and relied on by the public and any courts that may have to interpret it.<sup>114</sup> Scant descriptions of the invention may not necessarily be limiting, but it is uniquely in the power of the patentee to avoid close calls of claim interpretation by clear descriptions, backed by multiple embodiments, of the full scope of the claimed invention. Just as empirical scientists will provide multiple data points so as to gauge the limits or reach of their theories, it might reasonably be expected that patentees should likewise express inventions of an empirical nature in a number and range of embodiments to convey fully the scope of the claimed invention to the public. Where the patentee provides but one or a few closely situated embodiments, courts have relatively little basis for determining boundaries of a claim. Even though a claim is not ordinarily limited to a particular disclosed embodiment, the number and range of embodiments ultimately affects the scope that can be supported. Proper claim drafting will reduce the burden of, uncertainty surrounding, and need for claim construction, but claim drafters do not always perceive this to be to their advantage.

---

114. Cf. Joseph S. Miller, *Enhancing Patent Disclosure for Faithful Claim Construction*, 9 LEWIS & CLARK L. REV. 177, 183–84 (2005). Miller suggests that the Patent Office could improve claim construction through enhanced disclosure requirements, including that

every applicant state on the face of any patent (a) the field of art to which the claimed invention pertains; (b) all problems that the claimed invention helps solve; (c) a lexicon of all claim terms to which the applicant gives a meaning other than its accustomed meaning to people having ordinary skill in the pertinent art; and (d) a list of preferred objective reference sources, such as technical treatises and dictionaries (general or specialized), to which an interested reader should refer to learn about the ordinary meaning of the remaining claim terms to a person having ordinary skill in the art.

*Id.*

It may be somewhat ironic, therefore, that claim construction often affords patents supported by just a few, or maybe even a single, embodiment with potentially broader scope than more fully illustrated patents. Without much to go on, the court in the former case is often left with simply the plain language. The principal countervailing force confronting the patentee—the risk that the claim will fail the written description requirement—does not exert much effect, as it is often difficult to prove this basis for invalidity. (The written description doctrine is particularly subtle and, as a jury issue, it is fraught with uncertainty.)<sup>115</sup> By contrast, patents that are more fully illustrated provide a clearer basis for construing (and, in some cases, circumscribing) the scope of the claims. A more balanced middle ground is to consider the lack of any significant range of illustrative embodiments to be a factor in construing claims based on an empirical foundation. Just as an empirical theory supported by just a single or few examples will be narrow, so a patent supported by a single or narrow range of embodiments should, all other factors the same, be understood more narrowly. Such an approach would have the benefit of providing patent drafters with greater incentive to articulate the boundaries of the claimed invention. By contrast, claims based upon a conceptual or theoretical foundation may not require disclosure of multiple embodiments to prove their validity or delineate their scope.

(4) *Does Ambiguity in a Claim Term Limit Its Scope to Preferred Embodiment(s)?*

When the claim language is ambiguous, courts look to the specification to determine a reasonable interpretation.<sup>116</sup> In *Comark Communications, Inc. v. Harris Corp.*, the Federal Circuit observed that “interpreting claim language in light of the specification” is proper when a term is “so amorphous that one of skill in the art can only reconcile the claim language with the inventor’s disclosure by recourse to the specification.”<sup>117</sup> At the same time, the court cautioned against reading limitations from the specification into the claims

---

115. See Mark D. Janis, *On Courts Herding Cats: Contending with the “Written Description” Requirement (and Other Unruly Patent Disclosure Doctrines)*, 2 WASH. U. J.L. & POL’Y 55, 68 (2000).

116. See *Rexnord Corp. v. Laitram Corp.*, 274 F.3d 1336, 1343 (Fed. Cir. 2001). The *Rexnord* court stated:

[I]f the term or terms chosen by the patentee so deprive the claim of clarity that there is no means by which the scope of the claim may be ascertained by one of ordinary skill in the art from the language used, a court must look to the specification and file history to define the ambiguous term in the first instance.

*Id.* (internal quotations omitted).

117. 156 F.3d 1182, 1187 (Fed. Cir.1998).

(as opposed to interpreting claim language in light of the specification) and declined to do so in that case.<sup>118</sup> Nonetheless, courts have on occasion limited claim terms to the preferred embodiments where there is no other way of grounding the ambiguous language.<sup>119</sup>

ii) Characterizations of “The Invention” or “The Present Invention”

When the patentee uses descriptive terms such as “the invention” or “the present invention” to describe what is claimed, then those descriptive embodiments may be definitional. For example, *Honeywell International, Inc. v. ITT Industries, Inc.* concerned claims to a “fuel injection system component.”<sup>120</sup> Even though the ordinary and customary meaning of a “fuel injection system component” is not limited to a fuel filter, the Federal Circuit found that the proper construction was narrower than that customary meaning and should be limited to a fuel filter because all the disclosed embodiments disclosed only fuel filters and the specification repeatedly described the fuel filter as “this invention” and “the present invention.” Applying *Phillips*, the court found that there was no need to show that the inventor had “disavowed or disclaimed scope of coverage,” the standard previously set by *Texas Digital*.<sup>121</sup> Rather, the Federal Circuit noted, given the repeated descriptions in the patent specification of “the invention,” that “[t]he public is entitled to take the patentee at his word and the word was that the invention is a fuel filter.”<sup>122</sup> The fact that a specification discloses only a single embodiment does not, by itself, compel limiting the claim’s scope to that embodiment.<sup>123</sup> There must be additional evidence beyond the disclosure of a single embodiment to justify narrowing a construction to that

---

118. *Id.*

119. *See, e.g., Rhodia Chimie, Inc. v. PPG Indus., Inc.*, 402 F.3d 1371 (Fed. Cir. 2005).

120. 452 F.3d 1312, 1318 (Fed. Cir. 2006).

121. *See id.*

122. *Id.*; *see also Andersen Corp. v. Fiber Composites, LLC*, 474 F.3d 1361, 1367–68 (Fed. Cir. 2007) (limiting claim term “composite composition” to pellets in light of statements in specification that are “not descriptions of particular embodiments, but are characterizations directed to the invention as a whole”); *Microsoft Corp. v. Multi-Tech. Sys., Inc.*, 357 F.3d 1340, 1348 (Fed. Cir. 2004) (finding that statements in common specification serve to limit claim language because they “are not limited to describing a preferred embodiment, but more broadly describe the overall inventions of all three patents”); *Alloc, Inc. v. Int’l Trade Comm’n*, 342 F.3d 1361, 1370 (Fed. Cir. 2003) (“[T]his court looks to whether the specification refers to a limitation only as a part of less than all possible embodiments or whether the specification read as a whole suggests that the very character of the invention requires the limitation be a part of every embodiment.”).

123. *Phillips v. AWH Corp.*, 415 F.3d 1303, 1323 (Fed. Cir. 2005) (en banc).

embodiment.<sup>124</sup> When taken into consideration with the patentee's description of the invention, the fact that only a single embodiment is shown is a factor that may show that the inventor only intended to claim a particular feature as his invention.<sup>125</sup>

iii) Distinctions Over the Prior Art

As with descriptions of “the invention,” the patentee's manner of distinguishing her invention over the prior art may be definitional. That is, the specification's emphasis on the importance of a particular feature in solving the problems of the prior art is an important factor in defining the claims. These statements distinguishing the claimed invention from the prior art go to the heart of *Phillips'* instruction to construe claims consistent with a “full understanding of what the inventors actually invented.”<sup>126</sup> For example, in *Inpro II Licensing, S.A.R.L. v. T-Mobile USA, Inc.*,<sup>127</sup> the Federal Circuit affirmed the construction of “host interface” as a “direct parallel bus interface.” Among the dispositive factors in this narrow construction were that the only embodiment disclosed was a direct parallel bus interface and that “the specification emphasizes the importance of a parallel connection in solving the problems of the previously used serial connection.”<sup>128</sup> Since under *Phillips*, there was no need to show that the inventor had disclaimed scope of coverage, T-Mobile obtained a narrowing construction by demonstrating “what the inventor has described as the invention.”<sup>129</sup>

Statements distinguishing the prior art must be sufficiently clear to warrant a narrowing construction. *Ventana Medical Systems, Inc. v. Biogenex Laboratories, Inc.*,<sup>130</sup> concerned claims to a method of “dispensing” reagents onto a microscope slide. The question was whether “dispensing” was limited to “direct dispensing” (i.e., where the reagent container directly dispenses reagents onto the slide without an intermediary), or whether the claims encompassed the use of an intermediary device to “sip and spit” the reagents from the reagent container onto the slide. The specification contained

---

124. *Agfa Corp. v. Creo Prods., Inc.*, 451 F.3d 1366, 1376–77 (Fed. Cir. 2006).

125. *See Honeywell Int'l*, 452 F.3d at 1318 (limiting scope of “fuel injection system component” to a “fuel filter” because “[t]he written description's detailed discussion of the prior art problem addressed by the patented invention, *viz.*, leakage of non-metal fuel filters in EFI systems, further supports the conclusion that the fuel filter is not a preferred embodiment, but an only embodiment”).

126. *Phillips*, 415 F.3d at 1316.

127. 450 F.3d 1350, 1354–55 (Fed. Cir. 2006).

128. *Id.*

129. *Id.* at 1355 (quoting *Netword, LLC v. Centraal Corp.*, 242 F.3d 1347, 1352 (Fed. Cir. 2001)).

130. 473 F.3d 1173, 1180–81 (Fed. Cir. 2006).

general criticisms of prior art dispensers, including those using “sip and spit” approaches, as well as those using “direct dispensing” approaches. Because the specification equally criticized both types of prior art dispensers, there was nothing to suggest that the inventor was describing the invention to be the use of “direct” instead of “sip and spit” dispensing. Therefore, the Federal Circuit found it was inappropriate to limit the claim scope.<sup>131</sup>

iv) Consistent Usage of Claim Terms

Another claim construction principle is that the consistent and uniform usage of a claim term in a certain way in the specification may be definitional, showing the “ordinary meaning” of the claim term in the context of the invention. In such circumstances, otherwise broad language in the claim may be limited by the specification’s description of the invention. Consistent usage of a claim term in the specification can be definitional even without a showing that there is an “express definition” of the term or a “disclaimer,” which the now-overruled *Texas Digital* would have required. For example, the claim term “board” was found to be limited to wooden boards (as opposed to plastic lumber) in light of consistent statements in the specification and prosecution history describing the claimed “boards” as made from wood.<sup>132</sup>

e) Prosecution Disclaimers

Beyond using the prosecution history to ascertain the ordinary meaning of claim terms, the prosecution history can also be used to determine whether there was a “disclaimer” of claim scope. In order to convince the Patent Office to issue patent claims that have been rejected in light of the prior art, patent applicants frequently represent that their patent claims do *not* cover certain technologies. These statements are important limitations on claim scope.<sup>133</sup> The legal standard for finding a prosecution history disclaimer requires “a clear and unmistakable disavowal of scope during prosecution.”<sup>134</sup> For example, in *Atofina v. Great Lakes Chemical Corp.*,<sup>135</sup> the Federal Circuit found a prosecution disclaimer to apply, and construed “chromium catalyst” as a catalyst where the only catalytically active material is chromium without the addition of metal oxides or non-inert additives. The court based their construction on the applicants’ statements in the prosecution history which distinguished the claimed invention from the prior art’s use of metal oxides and non-inert additives, and which emphasized the “criticality of utilizing

---

131. *Id.* at 1181.

132. *Nystrom v. TREX Co.*, 424 F.3d 1136, 1145 (Fed. Cir. 2005).

133. *Phillips v. AWH Corp.*, 415 F.3d 1303, 1317 (Fed. Cir. 2005) (en banc).

134. *Purdue Pharm. L.P. v. Endo Pharms. Inc.*, 438 F.3d 1123, 1136 (Fed. Cir. 2006).

135. 441 F.3d 991, 996–97 (Fed. Cir. 2006).

chromium catalyst alone rather than in combination with other metal components.”<sup>136</sup>

By contrast, ambiguous statements in the prosecution history do not warrant a disclaimer, particularly when the applicant’s statements are subject to multiple interpretations.<sup>137</sup> For example, in *Golight, Inc. v. Wal-Mart Stores, Inc.*, a claim to a “rotating” spotlight was not found to have been disclaimed where statements in the prosecution history referring to the spotlight rotating “through 360°” were attributable to other claims, not the claim at issue.<sup>138</sup>

f) Looking to Other Claims: The Doctrine of Claim Differentiation

Patents typically contain multiple claims, with variations among the claims describing the patented invention. The doctrine of “claim differentiation” provides that “each claim in a patent is presumptively different in scope.”<sup>139</sup> The doctrine is based on “the common sense notion that different words or phrases used in separate claims are presumed to indicate that the claims have different meanings and scope.”<sup>140</sup> It also reflects the economic reality that patent fees depend on the number of claims in the patent. Patentees would be disinclined to purchase additional claims if they did not offer different scope. But it is important to recognize that the uncertainties of claim interpretation lead all but the most financially sensitive patent drafters to seek multiple overlapping claims.<sup>141</sup> Additional claims do not always cover different subject matter. Claim differentiation gives rise to a rebuttable presumption for claim construction purposes, especially when comparing the scope of an independent claim in view of its dependent claims: “[T]he presence of a dependent claim that adds a particular limitation

---

136. *Id.* at 997.

137. *SanDisk Corp. v. Memorex Prods., Inc.*, 415 F.3d 1278, 1287 (Fed. Cir. 2005).

138. 355 F.3d 1327, 1332 (Fed. Cir. 2004); *see also* *LG Elecs., Inc. v. Bizcom Elecs., Inc.*, 453 F.3d 1364, 1373–74 (Fed. Cir. 2006) (finding that prosecution history statements that the prior art did not teach accessing data signals “over a system bus” were not sufficiently clear to justify limiting claims to require claimed signals to travel over a system bus), *reversed on other grounds*, *Quanta Computer, Inc. v. LG Elecs., Inc.*, 553 U.S. 617 (2008).

139. *RF Del., Inc. v. Pac. Keystone Techs., Inc.*, 326 F.3d 1255, 1263 (Fed. Cir. 2003) (quoting *Wenger Mfg., Inc. v. Coating Mach. Sys., Inc.*, 239 F.3d 1225, 1233 (Fed. Cir. 2001)).

140. *Andersen Corp. v. Fiber Composites, LLC*, 474 F.3d 1361, 1369 (Fed. Cir. 2007) (quoting *Karlin Tech., Inc. v. Surgical Dynamics, Inc.*, 177 F.3d 968, 971–72 (Fed. Cir. 1999)).

141. *See generally* Mark A. Lemley, *The Limits of Claim Differentiation*, 22 *BERKELEY TECH. L.J.* 1389 (2007).

gives rise to a presumption that the limitation in question is not present in the independent claim.”<sup>142</sup>

“Pure” claim differentiation refers to the situation where there is no meaningful difference between an independent claim and its dependent claim, except for the presence of an added limitation in the dependent claim. In that situation, the presumption is especially strong that the independent claim is *not* restricted by the added limitation in the dependent claim.<sup>143</sup> In such situations, construing the independent claim to share that limitation would render the dependent claim “superfluous.”<sup>144</sup>

The doctrine of claim differentiation has less force when there are additional differences between the independent claim and its dependent claim, such that the dependent claim would not be rendered “superfluous” by limiting the independent claim.<sup>145</sup>

In the case of two independent claims, the doctrine of claim differentiation is generally not applicable because patent drafters are free to, and commonly do, claim an invention using multiple linguistic variations in multiple independent claims.<sup>146</sup> Even in cases of “pure” claim differentiation where the presumption would apply most strongly, the doctrine can be trumped by other considerations. Claim differentiation “can not broaden claims beyond their correct scope.”<sup>147</sup> That is, “the written description and prosecution history over come [sic] any presumption arising from the doctrine of claim differentiation.”<sup>148</sup> For example, where the patent applicant disclaimed subject matter during prosecution in order to obtain the patent, the patentee cannot attempt to recapture that subject matter through the

---

142. *Phillips v. AWH Corp.*, 415 F.3d 1303, 1315 (Fed. Cir. 2005) (en banc).

143. *Acumed LLC v. Stryker Corp.*, 483 F.3d 800, 806 (Fed. Cir. 2007).

144. *Andersen*, 474 F.3d at 1369–70 (Fed. Cir. 2007).

145. *See, e.g., SRAM Corp. v. AD-II Eng'g, Inc.*, 465 F.3d 1351, 1357–58 (Fed. Cir. 2006) (restricting independent claim to use of “precision index downshifting” even though this term was present in dependent claim, when additional differences existed between the independent and dependent claim).

146. *See, e.g., Andersen*, 474 F.3d at 1370 (declining to apply claim differentiation to separate groups of claims to “pellets,” “linear extrudates,” and “composite compositions” where there were other differences varying the scope of the claims); *Curtiss–Wright Flow Control Corp. v. Velan, Inc.*, 438 F.3d 1374, 1380–81 (Fed. Cir. 2006) (recognizing that “[c]laim drafters can also use different terms to define the exact same subject matter”); *Hormone Research Found. v. Genentech, Inc.*, 904 F.2d 1558, 1567 n.15 (Fed. Cir. 1990) (“It is not unusual that separate claims may define the invention using different terminology, especially where (as here) independent claims are involved.”).

147. *Curtiss–Wright Flow Control*, 438 F.3d at 1380–81.

148. *Andersen*, 474 F.3d at 1369–70 (quoting *Kraft Foods, Inc. v. Int'l Trading Co.*, 203 F.3d 1362 (Fed. Cir. 2000)).

doctrine of claim differentiation.<sup>149</sup> Given the wide variety of situations where the doctrine of claim differentiation does not apply, the Federal Circuit has cautioned that “[c]laim differentiation is a guide, not a rigid rule.”<sup>150</sup>

Limiting statements in the specification or prosecution history can rebut a broad claim term interpretation, even if the breadth of that term is reinforced by the doctrine of claim differentiation.<sup>151</sup> For example, in *Regents of the University of California v. Dakocytomation California, Inc.*,<sup>152</sup> the Federal Circuit approved of a limiting construction on the independent claim term “heterogenous mixture” to exclude repetitive sequences, notwithstanding the presence of dependent claims that do not exclude them.

As discussed more fully below, means-plus-function claims are limited to the corresponding structures, and their equivalents under § 112 para. 6. The statutorily-mandated scope of these claims cannot be stretched through resort to claim differentiation.<sup>153</sup>

#### g) Significance of the “Preamble” in Claim Construction

Patent claims commonly have a “preamble” that introduces the claimed invention. Some preambles may be just a few words, while others may be lengthy and detailed. A common dispute is whether or not the wording of the preamble is a limitation on the scope of the patent. A famously vague standard governs this inquiry: terms in the preamble are limiting when they are “necessary to give life, meaning, and vitality to the claims.”<sup>154</sup> The following principles are used in applying this standard.

Where the preamble is grammatically essential to the claim, the general rule is that it is limiting.<sup>155</sup> For example, where other terms in the body of the

---

149. *See* *Fantasy Sports Props., Inc. v. Sportsline.com, Inc.*, 287 F.3d 1108, 1115–16 (Fed. Cir. 2002).

150. *Laitram Corp. v. Rextord, Inc.*, 939 F.2d 1533, 1538 (Fed. Cir. 1991).

151. *See* *Seachange Int’l, Inc. v. C-COR Inc.*, 413 F.3d 1361, 1369 (Fed. Cir. 2005) (noting that claim differentiation is “not a hard and fast rule and will be overcome by a contrary construction dictated by the written description or prosecution history” (quoting *Kraft Foods, Inc. v. Int’l Trading Co.*, 203 F.3d 1362, 1368 (Fed. Cir. 2000))).

152. 517 F.3d 1364, 1375 (Fed. Cir. 2008).

153. *See, e.g.*, *Cross Med. Prods., Inc. v. Medtronic Sofamor Danek, Inc.*, 424 F.3d 1293, 1304 (Fed. Cir. 2005) (“[A]lthough the doctrine of claim differentiation suggests that claim 5 should be broader than claim 1, any presumption that the claims differ with respect to this feature may be overcome by a contrary construction mandated by the application of § 112 [para.] 6.”); *Laitram Corp.*, 939 F.2d at 1538.

154. *Kropa v. Robie*, 187 F.2d 150, 152 (C.C.P.A. 1951).

155. *See* *Catalina Mktg. Int’l, Inc. v. Coolsavings.com, Inc.*, 289 F.3d 801, 808–09 (Fed. Cir. 2002).

claim derive “antecedent basis” from the preamble, then the preamble is commonly found to be limiting.<sup>156</sup> Likewise, where the preamble is “essential to understand limitations or terms in the claim body,” it is similarly limiting.<sup>157</sup>

If a preamble term is a “necessary and defining aspect of the invention” the preamble is limiting.<sup>158</sup> This principle applies with special force where the language of the preamble was used during prosecution history to distinguish the claimed invention from the prior art.<sup>159</sup>

The countervailing principle is that a preamble is not limiting when the body of the claim “describes a structurally complete invention.”<sup>160</sup> Statements of an invention’s intended uses are generally *not* limiting.<sup>161</sup> This is because “the patentability of apparatus or composition claims depends on the claimed structure, not on the use or purpose of that structure.”<sup>162</sup> Thus, many cases turn on the question of whether a statement in the preamble describing the purpose of an invention is deemed to describe a “necessary and defining aspect of the invention” (which is limiting), or is simply a statement of intended use (which is not limiting).<sup>163</sup> A review of the Federal Circuit’s cases over the past ten years that litigated the issue of whether to construe the preamble reveals that the dominant approach in the close cases is to construe the preamble as a limitation.<sup>164</sup>

---

156. *Id.* at 808; *see also* *Bicon, Inc. v. Strauman Co.*, 441 F.3d 945, 952 (Fed. Cir. 2006).

157. *Catalina*, 289 F.3d at 808.

158. *On Demand Mach. Corp. v. Ingram Indus., Inc.*, 442 F.3d 1331, 1343 (Fed. Cir. 2006); *see also* *MBO Labs., Inc. v. Becton, Dickinson & Co.*, 474 F.3d 1323, 1330 (Fed. Cir. 2007) (interpreting the preamble term “immediately” as limiting, because “[t]he patentee here has clearly indicated via the specification and the prosecution history that the invention provides as an essential feature, immediate needle safety upon removal from the patient”).

159. *Catalina*, 289 F.3d at 808; *see also In re Cruciferous Sprout Litig.*, 301 F.3d 1343, 1347–48 (Fed. Cir. 2004) (finding the preamble phrase “rich in glucosinolates” limiting because the patentee relied on the preamble to distinguish the prior art in prosecution).

160. *Catalina*, 289 F.3d at 809; *see also Intertool, Ltd. v. Texar Corp.*, 369 F.3d 1289, 1295 (Fed. Cir. 2004) (finding the preamble non-limiting where the body of the claim described the invention in “complete and exacting structural detail”).

161. *Catalina*, 289 F.3d at 809.

162. *Id.*

163. *See* *Computer Docking Station Corp. v. Dell, Inc.*, 519 F.3d 1366, 1375 (Fed. Cir. 2008).

164. *See, e.g., TIP Sys., LLC v. Phillips & Brooks/Gladwin, Inc.*, 529 F.3d 1364, 1370 (Fed. Cir. 2008) (interpreting “handle” to be a structural limitation of the claim at issue); *Bass Pro Trademarks, LLC v. Cabela’s, Inc.*, 485 F.3d 1364, 1369 (Fed. Cir. 2007) (reversing the district court and noting that the term “vest” in the preamble of the claim at issue was stressed during patent prosecution and was thus limiting); *MBO Labs., Inc. v. Becton, Dickinson & Co.*, 474 F.3d 1323, 1330 (Fed. Cir. 2007) (holding that the specification and the prosecution history clearly indicated that the term “immediately” in the preamble was a

### 3. *Claim Terms Having Conventional, Presumed, or Established Meanings*

Claim terms generally take their meaning from the language of the patent, the prosecution history, and the applicable extrinsic evidence. Some terms, however, derive their meanings from conventional usage in claim drafting or prior judicial construction. The case law in this area, however, is notoriously malleable. Take, for example, the term “a” (or “an”). The Federal Circuit “has repeatedly emphasized that an indefinite article ‘a’ or ‘an’ in patent parlance carries the meaning of ‘one or more’ in open-ended claims containing the transitional phrase ‘comprising.’”<sup>165</sup> The court commented that this interpretation can be

best described as a rule, rather than merely as a presumption or even a convention. The exceptions to this rule are extremely limited: a patentee must “evinced [ ] a clear intent” to limit “a” or “an” to “one.” . . . An exception to the general rule that “a” or “an” means more than one only arises where the language of the claims themselves, the specification, or the prosecution history necessitate a departure from the rule.<sup>166</sup>

Just two weeks after stating this “rule,” the Federal Circuit found that the exception (singular meaning) applied based upon the claims and written description in *Tivo, Inc. v. Echostar Communications Corp.*<sup>167</sup> Thus, even for as simple and commonplace a word as “a,” the term can have divergent

limitation); *Seachange Int'l, Inc. v. C-COR Inc.*, 413 F.3d 1361, 1376 (Fed. Cir. 2005) (“The preamble provides the only antecedent basis and thus the context essential to understand the meaning . . .”); *NTP, Inc. v. Research In Motion, Ltd.*, 392 F.3d 1336, 1358 (Fed. Cir. 2004) (“[I]f the preamble helps to determine the scope of the patent claim, then it is construed as part of the claimed invention.”); *Eaton Corp. v. Rockwell Int'l Corp.*, 323 F.3d 1332, 1342 (Fed. Cir. 2003) (“[T]he inventor chose to use both the preamble and the body of the claim to define his invention. The preamble therefore limits the claimed invention.”). *But see* *Symantec Corp. v. Computer Assocs. Int'l, Inc.* 522 F.3d 1279, 1288–89 (Fed. Cir. 2008). The *Symantec* court stated:

[T]he purpose of a claim preamble is to give context for what is being described in the body of the claim; if it is reasonably susceptible to being construed to be merely duplicative of the limitations in the body of the claim (and was not clearly added to overcome a rejection), we do not construe it to be a separate limitation.

*Id.*

165. *Baldwin Graphic Sys., Inc. v. Siebert, Inc.* 512 F.3d 1338, 1342 (Fed. Cir. 2008).

166. *Id.* at 1342–43 (quoting *KCJ Corp. v. Kinetic Concepts, Inc.*, 223 F.3d 1351, 1356 (Fed. Cir. 2000)) (alterations in original).

167. 516 F.3d 1290, 1303 (Fed. Cir. 2008) (“The pertinent claim language refers to ‘assembl[ing] said video and audio components into an MPEG stream,’ which in context clearly indicates that two separate components are assembled into a single stream, not that the video components are assembled into one stream and the audio components into a second stream.”).

meanings based on the context of the patent (and despite the best efforts of the Federal Circuit to institute “rules” for its construction). Courts must remain sensitive to the context of patent claims, and avoid rigidly applying what may appear to be an established meaning.

“Transitional phrases” are terms that are used to link the various limitations in a claim. Transitional phrases govern, among other things, whether the claim is “open” or “closed” to the presence of additional elements. Restated, transitional phrases define whether a claim with defined limitations can be infringed by a device that has additional elements beyond what is specified in the claim. The term “consisting of” is a closed transitional phrase, while the term “comprising” is an open transitional phrase.<sup>168</sup> These terms have particularly established meanings based upon decades of consistent use in claim drafting.

Table B collects common terms that have been construed by the Federal Circuit. As the table reflects, some of these terms have been construed differently depending upon the context. Thus, courts should not woodenly adopt meanings from prior cases. Rather, they should be aware that the Federal Circuit has considered some terms in the past and has, in some cases, attributed general meanings. In every case, however, courts should carefully examine the claim term in context. Where a term does not have a clear meaning from the intrinsic evidence, then the jurisprudence may offer useful guidance.

**Table B: Common Terms Construed by the Federal Circuit**

Term	Meaning	Citation
<b>ARTICLES</b>		
a, an	Dominant meaning in open-ended claim: one or more.	Baldwin Graphic Sys., Inc. v. Siebert, Inc., 512 F.3d 1338, 1342 (Fed. Cir. 2008) (“That ‘a’ or ‘an’ can mean ‘one or more’ is best described as a rule, rather than merely as a presumption or even a convention.”); Lava Trading, Inc. v. Sonic Trading Mgmt., LLC, 445 F.3d 1348, 1354 (Fed. Cir. 2006); Free Motion Fitness, Inc. v. Cybex Int’l, Inc., 423 F.3d 1343, 1350 (Fed. Cir. 2005) (holding that “a” meant “one or more” where “references to a single cable in the specification are found in the description of the preferred embodiments, and do not evince a clear intent by the patentee to limit the article to the singular”); Collegenet, Inc. v. Applyyourself, Inc., 418 F.3d 1225, 1232

168. See AFG Indus., Inc. v. Cardinal IG Co., 239 F.3d 1239, 1244–45 (Fed. Cir. 2001).

		(Fed. Cir. 2005).
	However, sometimes means: only one.	Cat Tech LLC v. Tubemaster, Inc., 528 F.3d 871, 886 (Fed. Cir. 2008) (holding that even though “a” typically means “one or more,” the prosecution history trumped this conventional meaning and the patentee was playing “semantic antics”); Tivo, Inc. v. EchoStar Commc’ns Corp., 516 F.3d 1290, 1303 (Fed. Cir. 2008); Baldwin Graphic Sys., Inc. v. Siebert, Inc., 512 F.3d 1338, 1342–43 (Fed. Cir. 2008) (“An exception to the general rule . . . only arises where the language of the claims themselves, the specification, or the prosecution history necessitate a departure from the rule.”); Norian Corp. v. Stryker Corp., 432 F.3d 1356, 1359 (Fed. Cir. 2005) (“[T]he claim language ‘consisting of . . . a sodium phosphate,’ on its own, suggests the use of a single sodium phosphate.” (emphasis in original)).
at least one	There can be only one or more than one.	Z4 Techs., Inc. v. Microsoft Corp., 507 F.3d 1340, 1348–49 (Fed. Cir. 2007); Rhine v. Casio, Inc., 183 F.3d 1342, 1345 (Fed. Cir. 1999).
the, said	Indicates identity with a previously used claim term.	Baldwin Graphic Sys., Inc. v. Siebert, Inc., 512 F.3d 1338, 1342–43 (Fed. Cir. 2008).
plurality	At least one.	Verizon Servs. Corp. v. Vonage Holding Corp., 503 F.3d 1295, 1308–09 (Fed. Cir. 2007) (holding that a limitation in the specification requiring a “‘plurality’ may be satisfied by a single object”). <i>But see</i> York Prods., Inc. v. Cent. Tractor Farm & Family Ctr., 99 F.3d 1568, 1575 (Fed. Cir. 1996) (finding from the dictionary definition, “the state of being plural,” that a “‘plurality’” means “at least two”).
first, second	Distinguishes between repeated instances of an element or limitation.	Free Motion Fitness, Inc. v. Cybex Int’l, Inc., 423 F.3d 1343, 1348 (Fed. Cir. 2005); 3M Innovative Props. Co. v. Avery Dennison Corp., 350 F.3d 1365, 1371 (Fed. Cir. 2003).
<b>TRANSITIONAL PHRASES</b>		
comprising, comprised of	Is an “open” phrase and allows coverage of technologies that employ additional,	Predicate Logic, Inc. v. Distributive Software, Inc., 544 F.3d 1298, 1304 (Fed. Cir. 2008); CIAS, Inc. v. Alliance Gaming Corp., 504 F.3d 1356, 1360 (Fed. Cir. 2007)

	unrecited elements.	(“The usual and generally consistent meaning of ‘comprised of’ . . . is, like ‘comprising,’ that the ensuing elements or steps are not limiting.”); <i>AFG Indus. v. Cardinal IG Co.</i> , 239 F.3d 1239, 1245 (Fed. Cir. 2001). <i>But see Dippin’ Dots, Inc. v. Mosey</i> , 476 F.3d 1337, 1343 (Fed. Cir. 2007) (“‘[C]omprising’ is not a weasel word with which to abrogate claim limitations” and “[t]he presumption raised by the term ‘comprising’ does not reach into each of the six steps to render every word and phrase therein open-ended.”).
containing	Synonymous with “comprising.”	<i>Mars, Inc. v. H.J. Heinz Co.</i> , 377 F.3d 1369, 1377 (Fed. Cir. 2004).
including	Synonymous with “comprising.”	<i>Lucent Techs., Inc. v. Gateway, Inc.</i> , 525 F.3d 1200, 1214 (Fed. Cir. 2008) (“This court has consistently interpreted ‘including’ and ‘comprising’ to have the same meaning, namely, that the listed elements . . . are essential but other elements may be added.”); <i>Amgen Inc. v. Hoechst Marion Roussel, Inc.</i> , 314 F.3d 1313, 1344–45 (Fed. Cir. 2003). Note that in <i>Toro Co. v. White Consol. Indus., Inc.</i> , 199 F.3d 1295, 1300–02 (Fed. Cir. 1999), the term “including” was found to require permanency of the recited element—i.e., the claim phrase “cover including means for increasing the pressure” required the device’s restriction ring to be permanently affixed to and included as part of the air inlet cover, so claims were not literally infringed by device having separate restriction ring that was inserted and removed as a separate part.
having	May be “open” but does not convey an “open” meaning as strongly as “comprising.”	<i>Pieczenik v. Dyax Corp.</i> , 76 F. App’x 293, 296 (Fed. Cir. 2003); <i>Crystal Semiconductor Corp. v. TriTech Microelectronics Int’l Inc.</i> , 246 F.3d 1336, 1348 (Fed. Cir. 2001).
	May be closed, depending on the context of the patent.	<i>Lampi Corp. v. Am. Power Prods., Inc.</i> , 228 F.3d 1365, 1376 (Fed. Cir. 2000) (“Transitional phrases such as . . . ‘having’ . . . must be interpreted in light of the specification to determine whether open or closed language is intended.” (quoting U.S. PATENT & TRADEMARK OFFICE, MANUAL OF PATENT EXAMINING PROCEDURE § 2111.03 (7th ed. rev. 2000))).

consisting of	Is a “closed” phrase and excludes elements, steps, or ingredients not specified in the claims.	Immunocept, LLC v. Fulbright & Jaworski, L.L.P., 504 F.3d 1281, 1286 n.4 (Fed. Cir. 2007) (noting that “a competitor could design around a claim with this transitional phrase by adding any step or element not recited in the claim”); CIAS, Inc. v. Alliance Gaming Corp., 504 F.3d 1356, 1361 (Fed. Cir. 2007) (holding that even though “consisting of” limits the claimed invention to what is expressly set forth in the claim, “it does not limit aspects unrelated to the invention”); AFG Indus. v. Cardinal IG Co., 239 F.3d 1239, 1245 (Fed. Cir. 2001).
consisting essentially of	Occupies a middle ground between “open” and “closed” claims and is open to unlisted ingredients that do not materially affect the basic and novel properties of the invention.	PPG Indus. v. Guardian Indus. Corp., 156 F.3d 1351, 1354 (Fed. Cir. 1998); <i>see also</i> Ecolab, Inc. v. FMC Corp., 569 F.3d 1335, 1343–44 (Fed. Cir. 2009) (noting that “a patentee can alter [the] typical meaning” of “consisting essentially of” by making clear in the specification what it regarded as constituting a material change in the basic and novel properties of the invention); Atlas Powder Co. v. E.I. du Pont De Nemours & Co., 750 F.2d 1569, 1574 (Fed. Cir. 1984).
composed of	Synonymous with “consisting essentially of.”	AFG Indus. v. Cardinal IG Co., 239 F.3d 1239, 1245 (Fed. Cir. 2001).
<b>TERMS OF DEGREE</b>		
about	Avoids a strict numerical boundary.	Cohesive Techs., Inc. v. Waters Corp., 543 F.3d 1351, 1368 (Fed. Cir. 2008); Cent. Admixture Pharmacy Servs., Inc. v. Advanced Cardiac Solutions, P.C., 482 F.3d 1347, 1355–56 (Fed. Cir. 2007); Ortho–McNeil Pharm., Inc. v. Caraco Pharm. Labs., Ltd., 476 F.3d 1321, 1327 (Fed. Cir. 2007) (noting that in determining how far beyond the claimed range the term “about” extends the claim, a court “must focus . . . on the criticality of the [numerical limitation] to the invention”).
approximately	Serves only to expand the scope of literal infringement, not to enable application of the doctrine of equivalents.	U.S. Philips Corp. v. Iwasaki Elec. Co., 505 F.3d 1371, 1379 (Fed. Cir. 2007); <i>see also</i> Warner–Jenkinson Co. v. Hilton Davis Chem. Co., 520 U.S. 17, 25–28 (1997) (failing to even mention the patentees use of the term “approximately” in allowing consideration of the doctrine of equivalents).

effective amount	Any amount (or dosage) that can achieve therapeutic synergy.	Geneva Pharm., Inc. v. GlaxoSmithKline PLC, 349 F.3d 1373, 1383–84 (Fed. Cir. 2003) (“[E]ffective amount” is a common and generally acceptable term for pharmaceutical claims and is not ambiguous or indefinite, provided that a person of ordinary skill in the art could determine the specific amounts without undue experimentation.”); Abbott Labs v. Baxter Pharm. Prods., Inc., 334 F.3d 1274, 1280 (Fed. Cir. 2003).
essentially	Synonymous with “about.”	Eiselstein v. Frank, 52 F.3d 1035, 1039 (Fed. Cir. 1995).
substantially	Meaning is highly dependent on intrinsic evidence.	Deering Precision Instruments, LLC v. Vector Distrib. Sys., Inc., 347 F.3d 1314, 1322 (Fed. Cir. 2003) (construing the term “substantially in an imaginary plane”); Epcon Gas Sys., Inc. v. Bauer Compressors, Inc., 279 F.3d 1022 (Fed. Cir. 2002) (construing the terms “substantially constant” and “substantially below”); Zodiac Pool Care, Inc. v. Hoffinger Indus., Inc., 206 F.3d 1408 (Fed. Cir. 2000) (construing the term “substantially inward”); York Prods., Inc. v. Cent. Tractor Farm & Family Ctr., 99 F.3d 1568 (Fed. Cir. 1996) (construing the term “substantially the entire height thereof”); Tex. Instruments Inc. v. Cypress Semiconductor Corp., 90 F.3d 1558 (Fed. Cir. 1996) (construing the term “substantially in the common plane”).
up to about	May include or exclude the endpoint, depending on the context. Where the endpoint is numeric (e.g., up to about 10%), the endpoint may be included; whereas, where the endpoint is physical (e.g., painting the wall up to about the door), the endpoint may be excluded.	AK Steel Corp. v. Sollac & Ugine, 344 F.3d 1234, 1241 (Fed. Cir. 2003).

<b>SPATIAL RELATIONSHIPS</b>		
adjoining	Touching.	Int'l Rectifier Corp. v. IXYS Corp., 361 F.3d 1363, 1375 (Fed. Cir. 2004) (holding that as a matter of law, "adjoining" means "touching").
surround	To encircle on all sides simultaneously.	Libman Co. v. Quickie Mfg. Corp., 74 F. App'x 900, 904–05 (2003) (unpublished) (relying heavily on a dictionary definition).
in, between, within	Not required to be completely or continuously in, between or within; between may be satisfied even if extension beyond boundaries.	Foster v. Hallco Mfg. Co., No. 96-1399, 1997 U.S. App. LEXIS 18989, at *20 (Fed. Cir. July 14, 1997) (relying on dictionary definition).
to	When A travels "to" B, it is sufficient to travel on a pathway with B as a destination, possibly visiting intervening components.	Cybor Corp. v. FAS Techs., Inc., 138 F.3d 1448, 1458–59 (Fed. Cir. 1998).
defined	Can be used to mean that one element creates or forms the outline or shape of another element.	Rival Co. v. Sunbeam Corp., Nos. 98-1198 & 98-1199, 1999 WL 96416, at *4 (Fed. Cir. Feb. 23, 1999) (unpublished table decision).
<b>OTHER</b>		
whereby	"A 'whereby' clause that merely states the result of the limitations in the claim adds nothing to the patentability or substance of the claim."	Hoffer v. Microsoft Corp., 405 F.3d 1326, 1329 (Fed. Cir. 2005); Tex. Instruments Inc. v. U.S. Int'l Trade Comm'n, 988 F.2d 1165, 1172 (Fed. Cir. 1993).
	However, a "whereby" clause that "sets forth a structural limitation," and not merely the results achieved by the claimed structure, is a positive limitation of the claim.	Scheinman v. Zalkind, 112 F.2d 1017, 1019 (C.C.P.A. 1940).
standard, normal,	Time-dependent terms that are limited	PC Connector Solutions LLC v. SmartDisk Corp., 406 F.3d 1359, 1363 (Fed. Cir.

conventional, traditional	to technologies existing at the time of the invention.	2005).
mixture	Open ended and “does not exclude additional, unnamed ingredients.”	Mars, Inc. v. H.J. Heinz Co., L.P., 377 F.3d 1369, 1376 (Fed. Cir. 2004).
such as, may	“[O]f a kind or character about to be indicated, suggested, or exemplified; for instance.”	<i>In re</i> Johnston, 435 F.3d 1381, 1384 (Fed. Cir. 2006) (“[O]ptional elements do not narrow the claim because they can always be omitted.”); Catalina Mktg. Int’l, Inc. v. Coolsavings.com, Inc., 289 F.3d 801, 811 (Fed. Cir. 2002).
adapted	Made fit for a purpose; capable of a purpose.	Mattox v. Infotopia, Inc., 136 F. App’x 366, 369 (Fed. Cir. 2005).
assembly	“[A] collection of parts . . . to form a . . . structure.”	Kegel Co. v. AMF Bowling, Inc., 127 F.3d 1420, 1427 (Fed. Cir. 1997) (quoting WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY 131 (1986)).
uniform	“[H]aving always the same form.”	Middleton, Inc. v. Minn. Mining & Mfg. Co., 311 F.3d 1384, 1387 (Fed. Cir. 2002) (relying on <i>Tex. Digital Sys., Inc. v. Telegenix, Inc.</i> , 308 F.3d 1193 (Fed. Cir. 2002) and the “heavy presumption” rule).
predetermined	“[D]etermined beforehand.”	Koito Mfg. Co. v. Turn-Key-Tech, LLC, 381 F.3d 1142, 1147–48 (Fed. Cir. 2004).

#### 4. *Interpreting Terms to Preserve Validity*

Construing claims to preserve validity is a doctrine with a long and conflicted past. The Supreme Court has held that “if the claim were fairly susceptible of two constructions, that should be adopted which will secure to the patentee his actual invention.”<sup>169</sup> The doctrine arises from the presumption that the Patent Office has properly examined claims, and if those could be interpreted in two ways consistent with the patent documents, then the presumption of validity should drive the construction to maintain the patent’s validity. *Phillips* reaffirmed the doctrine (and given the doctrine’s Supreme Court roots, there was no choice), but simultaneously limited it to all but a rarity.<sup>170</sup>

169. *Smith v. Snow*, 294 U.S. 1, 14 (1935).

170. *See Phillips v. AWH Corp.*, 415 F.3d 1303, 1328 (Fed. Cir. 2005) (en banc). The *Phillips* court stated:

While we have acknowledged the maxim that claims should be construed to preserve their validity, we have not applied that principle broadly, and

There is a fundamental tension between this doctrine and the basic canons for construing claims. Claims are to be construed in light of the intrinsic and the pertinent extrinsic evidence that bears on the meaning of terms as they are used in the patent claims. That basic framework does not accommodate further modifications of claim language based on other prior art disclosures. Indeed, the public notice function of patents would suffer if untold prior art references were used in litigation to limit claim scope in order to rescue claims that would otherwise be invalid. Thus, when the Federal Circuit mentions the doctrine of construing claims to preserve validity, it commonly does so in the context of reversing district courts that improperly relied on the doctrine.<sup>171</sup>

The limited circumstances where the doctrine does have applicability are when two constructions are equally plausible, and a strong inference can be shown “that the PTO would have recognized that one claim interpretation would render the claim invalid, and that the PTO would not have issued the patent assuming that to be the proper construction of the term.”<sup>172</sup> This is a rare circumstance, and the best course will usually be to construe the claim language in view of the pertinent intrinsic and extrinsic evidence, and let the validity chips fall where they may.

### C. SPECIAL CASE: MEANS-PLUS-FUNCTION CLAIMS

A special class of claim language is construed as “means-plus-function” claim terms. When a party seeks to have a term construed as a “means-plus-function” term, the analysis is governed by § 112 para. 6:

An element in a claim for a combination may be expressed as a means or a step for performing a specified function without the recital of structure, material, or acts in support thereof, and such claim shall be construed to cover the corresponding structure,

---

we have certainly not endorsed a regime in which validity analysis is a regular component of claim construction. Instead, we have limited the maxim to cases in which “the court concludes, after applying all the available tools of claim construction, that the claim is still ambiguous.” In such cases, we have looked to whether it is reasonable to infer that the PTO would not have issued an invalid patent, and that the ambiguity in the claim language should therefore be resolved in a manner that would preserve the patent’s validity.

*Id.* (citations omitted).

171. *See, e.g.*, *Saunders Group, Inc. v. Comfortrac, Inc.*, 492 F.3d 1326, 1335 (Fed. Cir. 2007) (“[W]e hold only that the court’s validity analysis cannot be used as basis for adopting a narrow construction of the claims.”).

172. *Phillips*, 415 F.3d at 1328.

material, or acts described in the specification and equivalents thereof.<sup>173</sup>

When § 112 para. 6 is found to apply to claim language, the court construes the claim term by identifying the “function” associated with the claim language, and then identifying the corresponding “structure” in the specification associated with that function. The claim is construed to be limited to those corresponding structures and their equivalents. Thus, parties frequently attempt to invoke § 112 para. 6 as a way to narrow the scope of a patent to the particular technologies disclosed in the specification. Chart 9 sets forth the framework for construing functional claims terms. The court addresses Steps 1, 2A, and 2B as part of claim construction. Step 2C—determining whether the accused device is an “equivalent thereof”—is a question of fact for the jury.

**Chart 9: Framework for Construing Means-Plus-Function Claims**

Step 1: Is term in question “means-plus-function?”

- Rebuttable presumption: inclusion of “means”
- Rebutted if claim includes sufficient structure to perform recited function

Step 2: Interpretation Process

- A. Identify function of term (based on claim term language; not embodiments)
- B. Identify corresponding structure, material, or act based on disclosed embodiments
- C. Infringement Stage (Question of Fact): Determine whether accused device is the corresponding structure or “equivalents thereof” (as of time of issuance)

1. *Step 1: Is the Term in Question “Means-Plus-Function”?*

When presented with a request to invoke § 112 para. 6, the court must first determine if that section applies. Means-plus-function claiming applies only to “purely functional limitations that do not provide the structure that performs the recited function.”<sup>174</sup> There is a rebuttable presumption that § 112 para. 6 applies “[i]f the word ‘means’ appears in a claim element in association with a function.”<sup>175</sup> The use of the term “means” or

173. 35 U.S.C. § 112 para. 6 (2006).

174. *Depuy Spine, Inc. v. Medtronic Sofamor Danek, Inc.*, 469 F.3d 1005, 1023 (Fed. Cir. 2006) (quoting *Phillips*, 415 F.3d at 1328 (citing *Watts v. XL Sys. Inc.*, 232 F.3d 877, 880–81 (Fed. Cir. 2000))).

175. *Callicrate v. Wadsworth Mfg.*, 427 F.3d 1361, 1368 (Fed. Cir. 2005) (quoting *Micro Chem., Inc. v. Great Plains Chem. Co.*, 194 F.3d 1250, 1257 (Fed. Cir. 1999) (citing *Al-Site Corp. v. VSI Int’l, Inc.*, 174 F.3d 1308, 1314, 1318 (Fed. Cir. 1999))).

“mechanism”<sup>176</sup> in a claim limitation typically implies that the inventor used the “means-plus-function” claim format, which invokes the associated statutory limits on the literal scope of that claim limitation.<sup>177</sup> Nonetheless, this implication does not apply where the claim language itself provides the structure that performs the recited function.<sup>178</sup>

Conversely, a “claim term that does not use ‘means’ will trigger the rebuttable presumption that [35 U.S.C.] § 112 ¶ 6 does not apply.”<sup>179</sup> Disputes commonly arise over whether terms should be construed as means-plus-function language despite lacking an explicit “means” format. The presumption that such terms are *not* means-plus-function terms “can be rebutted ‘by showing that the claim element recite[s] a function without reciting sufficient structure for performing that function.’”<sup>180</sup> Whether a claim invokes § 112 para. 6 is decided on a limitation-by-limitation basis looking to the patent and the prosecution history.<sup>181</sup>

For example, the Federal Circuit applied § 112 para. 6 to the term “colorant selection mechanism,” explaining that “[t]he term ‘mechanism’ standing alone connotes no more structure than the term ‘means,’” and “the term ‘colorant selection’ . . . is not defined in the specification and has no dictionary definition, and there is no suggestion that it has a generally understood meaning in the art.”<sup>182</sup> By contrast, the Federal Circuit found § 112 para. 6 inapplicable to the term “compression member” because

---

176. *See Welker Bearing Co. v. PHD, Inc.*, 550 F.3d 1090, 1095–97 (Fed. Cir. 2008); *Mass. Inst. of Tech. v. Abacus Software*, 462 F.3d 1344, 1354 (Fed. Cir. 2006) (noting that “[t]he generic terms ‘mechanism,’ ‘means,’ ‘element,’ and ‘device,’ typically do not connote sufficiently definite structure [to avoid means-plus-function treatment] . . . . The term ‘mechanism’ standing alone connotes no more structure than the term ‘means.’”).

177. *See Greenberg v. Ethicon Endo-Surgery, Inc.*, 91 F.3d 1580, 1584 (Fed. Cir. 1996).

178. *See Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005) (en banc) (finding that a claim limitation stating “means disposed inside the shell for increasing its load bearing capacity comprising internal steel baffles” provides the relevant structure (“internal steel baffles”) and hence is not limited to the embodiments in the specification and equivalents thereof); *Cole v. Kimberly–Clark Corp.*, 102 F.3d 524, 531 (Fed. Cir. 1996) (finding that use of the phrase “perforation means . . . for tearing” does not invoke § 112 para. 6 because “perforation” provides the means for accomplishing the tearing function).

179. *Depuy Spine*, 469 F.3d at 1023 (quoting *CCS Fitness v. Brunswick Corp.*, 288 F.3d 1359, 1369 (Fed. Cir. 2002)) (brackets in original).

180. *Id.* (citation omitted); *see also Mas–Hamilton Group v. LaGard, Inc.*, 156 F.3d 1206, 1213–15 (Fed. Cir. 1998) (finding that “lever moving element” was not a known structure in the lock art and hence should be read to invoke the specific embodiments in the specification and equivalents thereof); *Raytheon Co. v. Roper Corp.*, 724 F.2d 951, 957 (Fed. Cir. 1983) (construing functional language introduced by “so that” to be equivalent to “means for” claim language).

181. *See Cole*, 102 F.3d at 531.

182. *Mass. Inst. of Tech. v. Abacus Software*, 462 F.3d 1344, 1354 (Fed. Cir. 2006).

“dictionary definitions and experts on *both* sides confirm that ‘compression member’ is an expression that was understood by persons of ordinary skill in the art to describe a kind of structure.”<sup>183</sup>

2. *Step 2: Interpretation of Means-Plus-Function Claim Terms*
  - a) Step 2A: Identify Claim Term Function

If the court concludes that § 112 para. 6 applies to a claim term, then the court must first identify the function of that term. It is important to identify the function associated with means-plus-function claim language before identifying the corresponding structure, material, or acts, and not to confuse these two analytically separate steps.<sup>184</sup> Errors arise when courts attempt to identify the function of a claimed invention in reference to a working embodiment, rather than by identifying function solely based on the claim language.<sup>185</sup> Attributing functions to a working device, rather than focusing on the claim language, may wrongly sweep additional functions into the claim.<sup>186</sup>

- b) Step 2B: Identify “Structure, Material, or Acts”

After identifying the claimed function, the court must identify the corresponding structure in the specification. This step is a frequent source of disputes. As a preliminary matter, if there is no structure in the specification corresponding to the claimed function, the claim is deemed to be indefinite, and is therefore invalid.<sup>187</sup> To find a claim invalid due to lack of a corresponding structure, clear and convincing evidence must be shown in order to overcome the presumption of validity (which is one of the few instances where there is a burden of proof in *Markman* proceedings).<sup>188</sup> Material incorporated by reference in a specification cannot serve as “corresponding structure.”<sup>189</sup>

If there is some structure identified, the next question is how much structure is “corresponding structure.” Where there are multiple

183. *Deputy Spine*, 469 F.3d at 1023.

184. *See* *JVW Enters., Inc. v. Interact Accessories, Inc.*, 424 F.3d 1324, 1330 (Fed. Cir. 2005) (“Determining a claimed function and identifying structure corresponding to that function involve distinct, albeit related, steps that must occur in a particular order.”).

185. *Id.* at 1330–31.

186. *Id.* at 1330.

187. *See* *Budde v. Harley–Davidson, Inc.*, 250 F.3d 1369, 1376 (Fed. Cir. 2001).

188. *Id.* at 1381–82 (finding that the disclosure of “commercially available units” was sufficient disclosure of vacuum sensors, especially in the face of weak expert testimony to show how persons of skill in the art would interpret the specification).

189. *Default Proof Credit Card Sys. Inc., v. Home Depot U.S.A., Inc.*, 412 F.3d 1291, 1301 (Fed. Cir. 2005).

embodiments of structures corresponding to the claimed function, all of those embodiments are deemed to be “corresponding.”<sup>190</sup> Thus, the claim would be infringed by an accused product using any of those corresponding structures.

A closely related question, however, is the extent of the structures that should be swept into the analysis. Any structures “necessary” to the claimed function must be disclosed.<sup>191</sup> However, the range of “necessary” structures can be pushed to the absurd. For example, when a claimed function is a means for computing, there is no need to disclose the power plant that provides the electricity to run the computer. And similarly when patents disclose some of the underlying infrastructure for carrying out the invention, there is no need to sweep in all that underlying structure when identifying the corresponding structure. Rather, “structure disclosed in the specification is ‘corresponding’ structure only if the specification or prosecution history clearly links or associates that structure to the function recited in the claim.”<sup>192</sup> Relatedly, where a specification’s disclosed structure has multiple components, only some of which perform a claimed function, the “necessary structure” is limited to the components that perform the claimed function.<sup>193</sup>

c) Step 2C: “Equivalents Thereof”

In addition to structures, materials, or acts of the embodiments described in the patent’s specification, the patentee is entitled to “equivalents thereof” as of the time the patent issued. Unlike the determination of function and corresponding structure, material, or acts which are clearly part of claim construction, the “equivalents” issue arises in the context of the infringement determination. The fact-finder must determine whether the means in the accused device or method performs the function stated in the claim in the same or an equivalent manner as the corresponding structures, materials, or acts set forth in the specification.<sup>194</sup>

---

190. *See Callicrate v. Wadsworth Mfg., Inc.*, 427 F.3d 1361, 1369 (Fed. Cir. 2005).

191. *See In re Dossel*, 115 F.3d 942, 946 (Fed. Cir. 1997).

192. *Minks v. Polaris Indus.*, 546 F.3d 1364, 1377 (Fed. Cir. 2008) (citing *Texas Digital Sys., Inc. v. Telegenix, Inc.*, 308 F.3d 1193, 1208 (Fed. Cir. 2002) (quoting *B. Braun Med., Inc. v. Abbott Lab.*, 124 F.3d 1419, 1424 (Fed. Cir. 1997))).

193. *See, e.g., Clearwater Sys. Corp. v. Evapco, Inc.*, 553 F. Supp. 2d 173, 179–80 (D. Conn. 2008) (differentiating between disclosed circuitry components that perform the claimed function and those that do not, and excluding other components from construction).

194. *See Palumbo v. Don–Joy Co.*, 762 F.2d 969, 974–75 (Fed. Cir. 1985).

d) Specific Rule for Means-Plus-Function Claims in the Computer Software Context

Merely pointing to a “computer” may not be sufficient to provide sufficient structure to a software or computer patent. Rather, the particular algorithms that carry out the invention may be the necessary “structure” to fulfill § 112 para. 6. In *WMS Gaming Inc. v. International Game Technology*,<sup>195</sup> the Federal Circuit ruled that the structure in the specification supporting the claim language, “means for assigning,” was *not* merely an algorithm executed by a computer, but was rather the particular algorithms taught in the specification. “In a means-plus-function claim in which the disclosed structure is a computer, or microprocessor, programmed to carry out an algorithm, the disclosed structure is not the general purpose computer, but rather the special purpose computer programmed to perform the disclosed algorithm.”<sup>196</sup>

D. DYSFUNCTIONAL CLAIMS: MISTAKES AND INDEFINITENESS

Courts must occasionally deal with dysfunctional claims, falling into two principal categories: (1) claims that contain obvious typographical, grammatical, or other errors that render the claim unworkable; and (2) claims that may be indefinite (possibly depending on how they are construed), raising the possibility that the claims are invalid under § 112 para. 2. The former may be obvious from the context and quite possibly can be due to the Patent Office’s oversight. Some mistakes are more intractable, and go to the heart of the claimed invention. Deciding whether these mistakes can be fixed at all, who should fix them (the court or the PTO), and what the consequences of changing the claims are, can be challenging.

1. *Mistakes*

When issues of mistaken claim language arise, the parties often call into question the power of courts to correct mistakes in patents through the claim construction process. Attempts to correct patents raise the threshold question of whether the district court has legal authority to correct the alleged error or omission or whether such an issue must be brought to the PTO. The somewhat ambiguous answer is that “courts can continue to correct obvious minor typographical and clerical errors in patents,” whereas “major errors are subject only to correction by the PTO.”<sup>197</sup>

---

195. 184 F.3d 1339, 1348–49 (Fed. Cir. 1999).

196. *Id.* at 1349.

197. *Novo Indus., L.P. v. Micro Molds Corp.*, 350 F.3d 1348, 1357 (Fed. Cir. 2003).

The general rule is that “the district court can correct an error only if the error is evident from the face of the patent.”<sup>198</sup> In order to permit correction, two requirements must be met: “A district court can correct a patent only if (1) the correction is not subject to reasonable debate based on consideration of the claim language and the specification *and* (2) the prosecution history does not suggest a different interpretation of the claims.”<sup>199</sup> Another general rule limiting the corrective power of courts is that “courts may not redraft claims, whether to make them operable or to sustain their validity.”<sup>200</sup>

Whether an error is “evident from the face of the patent” is a matter of frequent dispute. Where the applicant uses an inapt claim term, the applicant is typically held to the wording, even if the intended meaning is abundantly clear. For example, in *Chef America, Inc. v. Lamb–Weston, Inc.*,<sup>201</sup> in a patent which dealt with a process for cooking dough, the claim language required “heating the resulting batter-coated dough *to* a temperature in the range of about 400°F to 850°F.”<sup>202</sup> If the dough is heated “to” that temperature range, it would be burned to a crisp. Heating the dough “at” that temperature range supposedly results in a light, flaky, crispy texture, according to the patent’s specification.<sup>203</sup> Even though it would be nonsensical to require heating the dough “to” 400°F, the court refused to construe the claims otherwise, and the Federal Circuit affirmed, which rendered the claims non-infringed.<sup>204</sup>

Courts have somewhat greater leeway to correct administrative errors attributable to the Patent Office. Minor errors can be corrected by a district court, even if the prosecution history must be consulted in order to determine how to fix the error. For example, in *Hoffer v. Microsoft Corp.*,<sup>205</sup> the Federal Circuit ruled that the district court could have fixed an error in patent claim numbering that left a dependent claim without a reference to its independent claim, where the appropriate reference was easily determined by reference to the prosecution history. However, where the PTO printing office omitted a block of claim text from a patent, that error was found to be beyond the district court’s corrective powers.<sup>206</sup>

---

198. *See* *Group One, Ltd. v. Hallmark Cards, Inc.*, 407 F.3d 1297, 1303 (Fed. Cir. 2005).

199. *Id.* (quoting *Novo Indus.*, 350 F.3d at 1357).

200. *Chef Am., Inc. v. Lamb–Weston, Inc.*, 358 F.3d 1371, 1374 (Fed. Cir. 2004).

201. *Id.* at 1374.

202. *Id.* at 1371 (emphasis added).

203. *See id.* at 1372.

204. *See id.* at 1373–74.

205. 405 F.3d 1326, 1331 (Fed. Cir. 2005).

206. *See* *Group One, Ltd. v. Hallmark Cards, Inc.*, 407 F.3d 1297, 1303 (Fed. Cir. 2005) (“The prosecution history discloses that the missing language was required to be added by the examiner as a condition for issuance, but one cannot discern what language is missing

When a district court construes a patent claim to correct an error, the construction generally has a retroactive effect. Conversely, corrections by the Patent Office are prospective.<sup>207</sup> Thus, litigants have a strong incentive to fix errors through judicial construction as opposed to petitioning the Patent Office for a certificate of correction. However, the risk is that if the district court declines to fix the correction, the defective claims may be held invalid for indefiniteness, or may fail for other reasons such as non-infringement.<sup>208</sup>

## 2. *Indefiniteness*

The potentially dispositive issue of “indefiniteness” is frequently intertwined with the claim construction process. “Indefiniteness” is an invalidity defense based on § 112 para. 2, which requires that the claims of a patent “particularly point[] out and distinctly claim[] the subject matter which the applicant regards as his invention.”<sup>209</sup>

The primary purpose of the definiteness requirement is to ensure that the claims are written in such a way that they give notice to the public of the extent of the legal protection afforded by the patent, so that interested members of the public, e.g., competitors of the patent owner, can determine whether or not they infringe.<sup>210</sup>

When a claim cannot be construed, it is indefinite, and therefore invalid.<sup>211</sup> Some authority suggests that all indefiniteness issues boil down to an issue of claim construction.<sup>212</sup> However, there are instances where a claim *can* be construed, but cannot be meaningfully applied, in which case the claim is also invalid for indefiniteness.

Indefiniteness is unique among claim construction issues in that it carries a burden of proof. Under § 282 of the Patent Act, issued patents carry a

---

simply by reading the patent. The district court does not have authority to correct the patent in such circumstances.”).

207. *See* *Novo Indus., L.P. v. Micro Molds Corp.*, 350 F.3d 1348, 1356 (Fed. Cir. 2003) (noting that a certificate of correction from the Patent Office is “only effective for causes of action arising after it was issued” (quoting *Southwest Software, Inc. v. Harlequin Inc.*, 226 F.3d 1280 (Fed. Cir. 2000))).

208. *See, e.g., id.* at 1358 (refusing to correct patent, and holding claim indefinite).

209. 35 U.S.C. § 112 (2006).

210. *All Dental Prodx, LLC v. Advantage Dental Prods., Inc.*, 309 F.3d 774, 779 (Fed. Cir. 2002).

211. *Aero Prods. Int'l, Inc. v. Intex Recreation Corp.*, 466 F.3d 1000, 1016 (Fed. Cir. 2006).

212. *See id.* (“If a claim is amenable to construction, ‘even though the task may be formidable and the conclusion may be one over which reasonable persons will disagree,’ the claim is not indefinite.” (quoting *Exxon Res. & Eng'g Co. v. United States*, 265 F.3d 1371, 1375 (Fed. Cir. 2001))).

presumption of validity that can only be rebutted by clear and convincing evidence.<sup>213</sup> Therefore, because it invalidates a patent, a claim construction finding the claim indefinite must be supported by clear and convincing evidence.

Indefiniteness issues can arise from the wide variety of inadvertent mistakes and nonsensical statements that pervade patents. Courts must decide if the claims are so “insolubly ambiguous” that they are not amenable to construction or application to an infringement determination.<sup>214</sup> Some indefiniteness disputes arise in the context of typographical and printing errors that make a claim impossible to read or interpret. Minor errors are commonly overlooked so long as persons of skill in the art can still understand the claims.<sup>215</sup> However, where entire blocks of text are missing from claims, then the public cannot reasonably be expected to appreciate their scope, and the claims are invalid.<sup>216</sup>

Another type of indefiniteness issue arises in the context of means-plus-function claims, where there is no structure in the specification corresponding to the claimed function. In such circumstances, the claim cannot be construed.<sup>217</sup>

Claims may also be invalid for indefiniteness where the claim language is so inherently standardless that it cannot be meaningfully applied. These matters are often treated as “claim construction” questions, although they might more aptly be considered a question of whether the claims are indefinite as applied. For example, a claim requiring an “aesthetically pleasing” interface screen was found indefinite where even the patentee’s expert could not articulate how to determine infringement.<sup>218</sup> Another example is a claim directed to *both* a system and a method of using that system, which is invalid because the public cannot determine the acts that constitute infringement.<sup>219</sup> These latter examples are not so much “claim

---

213. *See* Bancorp Servs., LLC v. Hartford Life Ins. Co., 359 F.3d 1367, 1372 (Fed. Cir. 2004).

214. *Star Scientific, Inc. v. R.J. Reynolds Tobacco Co.*, 537 F.3d 1357, 1371 (Fed. Cir. 2008).

215. *See* Energizer Holdings, Inc. v. Int’l Trade Comm’n, 435 F.3d 1366, 1369–70 (Fed. Cir. 2006) (refusing to invalidate claim where phrase “said zinc anode” lacked an antecedent basis).

216. *See, e.g.*, *Group One, Ltd. v. Hallmark Cards, Inc.*, 407 F.3d 1297, 1302 (Fed. Cir. 2005).

217. *See* *Default Proof Credit Card Sys., Inc. v. Home Depot U.S.A., Inc.*, 412 F.3d 1291, 1302–03 (Fed. Cir. 2005) (invalidating claim for indefiniteness for lack of a structure in the specification corresponding to the claimed function).

218. *Datamize, LLC v. Plumtree Software, Inc.*, 417 F.3d 1342, 1354 (Fed. Cir. 2005).

219. *IPXL Holdings, LLC v. Amazon.com, Inc.*, 430 F.3d 1377, 1383–84 (Fed. Cir.

construction” issues, but rather are fundamental flaws in patent claims that make them impossible to apply. Nonetheless, these matters are commonly briefed during the claim construction process and, depending on the case, it may be appropriate to handle them along with other claim construction matters.

#### E. DEFERENCE TO PRIOR CLAIM CONSTRUCTION RULINGS

Where a claim term has been construed in a prior judicial proceeding, it is not uncommon for one or more of the litigants to assert that the court is bound by or, at a minimum, should accord substantial deference to that prior ruling. The Supreme Court’s *Markman* decision ostensibly encourages deference to prior claim construction in noting “the importance of uniformity in the treatment of a given patent as an independent reason to allocate all issues of construction to the court.”<sup>220</sup> The Supreme Court acknowledged in the next paragraph, however, that “issue preclusion could not be asserted against new and independent infringement defendants even within a given jurisdiction.”<sup>221</sup>

Determining the standards for according deference to prior *Markman* orders, as well as the application of such standards, has proven to be complicated in practice. Parties, sometimes uncritically, invoke a variety of doctrines—claim preclusion, res judicata, issue preclusion, collateral estoppel, judicial estoppel, or stare decisis—in efforts to constrain or obviate *Markman* determinations. The intermediate nature of *Markman* rulings makes it all the more complicated to apply such doctrines. *Markman* rulings are a means (construing claim terms) to an end (adjudicating patent validity and infringement or, more commonly, reaching a settlement agreement), not final judgments in and of themselves. Even though *Markman* orders often serve as the basis for summary judgment rulings, they are not always vital to the outcome and might be vacated as part of a settlement agreement. An additional complicating factor is the characterization of *Markman* rulings as questions of law. As a result, determining the preclusive effect of such orders requires navigation of overlapping and not entirely cohesive civil procedure doctrines.

Before turning to the particular legal standards for according deference to prior *Markman* determinations, it will be useful to clarify the relevant terminology. There are four distinct concepts: (1) claim preclusion (and the related concept of res judicata); (2) issue preclusion (and the related concepts

---

2005).

220. *Markman v. Westview Instruments, Inc.*, 517 U.S. 370, 390 (1996).

221. *Id.* at 391.

of collateral and direct estoppel); (3) judicial estoppel; and (4) stare decisis. Issue preclusion, judicial estoppel, and stare decisis are pertinent to the appropriate deference to be accorded prior claim construction rulings; claim preclusion generally does not come into play in claim construction.

1. *Distinguishing Among Preclusion and Estoppel Doctrines*

Although *res judicata* has historically been interpreted broadly to encompass the binding effect of a judgment in a prior case on claims asserted in pending litigation (and hence encompassing both claim and issue preclusion), the modern trend limits *res judicata* to claim preclusion.<sup>222</sup> “Claim preclusion refers to the effect of a judgment in foreclosing litigation of a matter that never has been litigated, because of a determination that it should have been advanced in an earlier suit. Claim preclusion therefore encompasses the law of merger and bar.”<sup>223</sup> When a plaintiff prevails in a lawsuit arising from a particular transaction, all of the claims that the plaintiff raised *or could have raised* “merge” into that judgment and are “barred” from further litigation.<sup>224</sup> If the plaintiff attempts to litigate any of those claims again, the judgment itself will serve as a defense. Since *Markman* rulings do not themselves resolve claims to relief (they merely interpret patent claim terms), they cannot be said to constitute “claim preclusion” judgments as that technical terminology is used in civil procedure.<sup>225</sup>

By contrast, the related doctrine of issue preclusion arises with some frequency in *Markman* proceedings. “Issue preclusion refers to the effect of a judgment in foreclosing relitigation of a matter that has been litigated and decided. . . . This effect also is referred to as direct or collateral estoppel.”<sup>226</sup> Where a patentee (including those in privity with her) has previously litigated the scope of a patent claim term, a defendant in a subsequent lawsuit relating

---

222. See 18 JAMES W. MOORE ET AL., MOORE’S FEDERAL PRACTICE ¶ 131.10[1][b] (3d ed. 2010).

223. *Migra v. Warren City Sch. Dist. Bd. of Educ.*, 465 U.S. 75, 77 n.1 (1984). *The Restatement (Second) of Judgments* adheres to the broader definition of *res judicata* as encompassing both claim and issue preclusion. See RESTATEMENT (SECOND) OF JUDGMENTS, Ch. 3 introductory note (1982).

224. See *Waid v. Merrill Area Pub. Sch.*, 91 F.3d 857, 863 (7th Cir. 1996).

225. Moreover, decisions by the International Trade Commission cannot have claim preclusive effect in district courts because the commission cannot award damages. See *Tex. Instruments Inc. v. Cypress Semiconductor Corp.*, 90 F.3d 1558, 1569 (Fed. Cir. 1996); *Bio-Technology Gen. Corp. v. Genentech, Inc.*, 80 F.3d 1553, 1563–64 (Fed. Cir. 1996). Nonetheless, a district court can “attribute whatever persuasive value to the prior ITC decision that it considers justified.” *Tex. Instruments*, 90 F.3d at 1569.

226. *Migra*, 465 U.S. at 77 n.1; see also *Pharmacia & Upjohn Co. v. Mylan Pharm., Inc.*, 170 F.3d 1373, 1379 (Fed. Cir. 1999).

to the same patent claim term might assert issue preclusion to foreclose relitigation of that matter.<sup>227</sup> The test for issue preclusion, however, is relatively strict and authority is split on its role in the context of prior *Markman* rulings.

Judicial estoppel is an equitable doctrine that precludes a party from adopting a position that is inconsistent with a position taken in prior lawsuit, whether or not that issue had been actually litigated in the prior proceeding party.<sup>228</sup>

Where a party assumes a certain position in a legal proceeding, and succeeds in maintaining that position, he may not thereafter, simply because his interests have changed, assume a contrary position, especially if it be to the prejudice of the party who has acquiesced in the position formerly taken by him.<sup>229</sup>

The purpose of the doctrine is “‘to protect the integrity of the judicial process’ by ‘prohibiting parties from deliberately changing positions according to the exigencies of the moment.’”<sup>230</sup>

The doctrine of *stare decisis* promotes adherence to decided matters of law so as to foster stability and equal treatment. It takes its name from the Latin maxim, “*stare decisis et non quieta movere*” or “to stand by things decided, and not to disturb settled points.”<sup>231</sup> The strength of such adherence depends on the source of the prior decision. *Stare decisis* compels lower courts to follow the decisions of higher courts on questions of law, whether applied to parties (or those in privity) or complete strangers to the prior proceeding. The decision of a district court is not binding precedent on a different judicial district, the same judicial district, or even the same judge in a different case under the doctrine of *stare decisis*. Rather, *stare decisis* requires only that the later court encountering the issue give consideration and careful analysis to that sister court’s decision where applicable to a similar fact pattern.<sup>232</sup>

---

227. A patentee cannot use issue preclusion offensively to foreclose a defendant who was not party to that prior litigation from litigating the scope of the patent claim. *See Tex. Instruments*, 182 F. Supp. 2d at 589–90. Had the Federal Circuit construed that claim term, however, the defendant might be bound under the doctrine of *stare decisis*.

228. *See generally* 18 MOORE ET AL., *supra* note 222, ¶ 18-134.30.

229. *New Hampshire v. Maine*, 532 U.S. 742, 749 (2001) (quoting *Davis v. Wakelee*, 156 U.S. 680, 689 (1895)).

230. *Id.* at 749–50 (quoting *Edwards v. Aetna Life Ins. Co.*, 690 F.2d 595, 598 (6th Cir. 1982) and *United States v. McCaskey*, 9 F.3d 368, 378 (5th Cir. 1993)).

231. *See* BLACK’S LAW DICTIONARY 1537 (9th ed. 2009).

232. *See United States v. Rodriguez-Pacheco*, 475 F.3d 434, 441 (1st Cir. 2007).

## 2. *Issue Preclusion and Collateral Estoppel*

Issue preclusion most commonly arises in the context of claim construction where a patentee who has previously litigated a patent through a *Markman* ruling seeks a fresh opportunity to construe a claim and an opposing party argues that the prior construction should govern interpretation of the term in question.<sup>233</sup> The previous litigation might have ended in a settlement agreement, including possibly an order vacating the claim construction ruling. The courts have divided on what effect, if any, to accord prior claim construction rulings.

The general standard for issue preclusion requires the party seeking to foreclose relitigation of an issue to prove: (a) the issue sought to be precluded is identical to the issue decided in the prior action; (b) the issue was actually litigated in that action; (c) the party against whom collateral estoppel is sought had a full and fair opportunity to litigate the issue in the prior action; and (d) the determination was essential to the final judgment of the prior action.<sup>234</sup> Courts apply the collateral estoppel standard of the regional circuit since issue preclusion is a procedural matter.<sup>235</sup>

### a) Identity of Issues

The first prong of the issue preclusion test is satisfied where the patent claims (and claim terms) at issue in the *Markman* proceeding were interpreted in the prior case.<sup>236</sup> When new claim terms are at issue, then collateral estoppel does not apply.<sup>237</sup> Since different claims within the same patent may use the same language, the “identity of issues” prong may nonetheless be satisfied if the language and context of the language are identical.<sup>238</sup> Similarly, since different patents may emanate from the same specification, as in the case of divisional and continuation applications, the “identity of issues”

---

233. *Cf. Blonder-Tongue Labs., Inc. v. Univ. of Ill. Found.*, 402 U.S. 313, 333 (1971) (holding that a patentee whose patent is invalidated after “a full and fair” opportunity to litigate its validity is collaterally estopped from relitigating the validity of the patent).

234. *See Innovad Inc. v. Microsoft Corp.*, 260 F.3d 1326, 1334 (Fed. Cir. 2001) (citing *In re Freeman*, 30 F.3d 1459, 1465 (Fed. Cir. 1994)).

235. *See RF Del., Inc. v. Pac. Keystone Techs., Inc.*, 326 F.3d 1255, 1261 (Fed. Cir. 2003).

236. *See, e.g., Dynacore Holdings Corp. v. U.S. Philips Corp.*, 243 F. Supp. 2d 31, 35 (S.D.N.Y. 2003) (same patent claims at issue); *Kollmorgen Corp. v. Yaskawa Elec. Corp.*, 147 F. Supp. 2d 464, 466 (W.D. Va. 2001); *Abbott Labs. v. Dey, L.P.*, 110 F. Supp. 2d 667, 669 (N.D. Ill. 2000) (“The claim construction issues disputed in this case are the same issues litigated in the [first] case . . .”).

237. *See, e.g., P.A.T., Co. v. Ultrak, Inc.*, 948 F. Supp. 1518, 1520–21 (D. Kan. 1996).

238. *See, e.g., In re Freeman*, 30 F.3d 1459, 1465 n.4 (Fed. Cir. 1994).

prong may nonetheless be satisfied if the language and context of the language are identical.<sup>239</sup>

b) Actual Litigation

To satisfy the “actual litigation” prong, the parties to the original litigation must have disputed the claim term at issue and it must have been adjudicated by the court.<sup>240</sup> The “actual litigation” test is not satisfied where: an issue was raised but later abandoned,<sup>241</sup> the court in the earlier proceeding declined to rule on the issue,<sup>242</sup> or there is ambiguity as to what was actually litigated and decided.<sup>243</sup> Courts usually do not consider matters resolved by stipulation to have been actually litigated.<sup>244</sup> An exception exists, however, where the parties intend to foreclose future litigation of the issue.<sup>245</sup>

c) Full and Fair Opportunity to Litigate

Issue preclusion requires the underlying proceeding to have afforded the party to be foreclosed from relitigation a full and fair opportunity to litigate. This means that issue preclusion can never be applied against a party not involved (or in privity with those involved) in the prior proceeding. In *Blonder-Tongue Laboratories, Inc. v. University of Illinois Foundation*, the Supreme Court identified a range of factors bearing on whether a patentee had a full and fair chance to litigate the validity of a patent: choice of forum; incentive to litigate; if the issue is obviousness, whether the first validity determination used the standards announced in *Graham v. John Deere Co.*,<sup>246</sup> whether opinions filed in the first case suggest that it was one of those rare instances where the court or jury failed to grasp the technical subject matter and issues; and whether, without fault of its own, the patentee was deprived of crucial

239. See *Masco Corp. v. United States*, 49 Fed. Cl. 337, 343–44 (Fed. Cl. 2001) (applying collateral estoppel to a continuation patent (employing identical claim language) relating back to the patent construed in the earlier litigation).

240. See, e.g., *Kollmorgen*, 147 F. Supp. 2d at 466 (stating that the “actually litigated” prong was met after a lengthy Markman hearing on the claim construction); *Abbott Labs.*, 110 F. Supp. 2d at 669–70 (stating the “actually litigated” prong was met because the parties “briefed and argued the issues” before the judge); *Freeman*, 30 F.3d at 1466; RESTATEMENT (SECOND) OF JUDGMENTS § 27 cmt. d (1980).

241. See 18 MOORE ET AL., *supra* note 222, ¶ 132.03[2][e].

242. See *id.* ¶ 132.03[4][g].

243. See *id.* ¶ 132.03[2][g].

244. See, e.g., *United States v. Young*, 804 F.2d 116, 118 (8th Cir. 1986) (“A fact established in prior litigation not by judicial resolution but by stipulation has not been ‘actually litigated’ . . .”).

245. See *Hartley v. Mentor Corp.*, 869 F.2d 1469, 1470 (Fed. Cir. 1989); 18 MOORE ET AL., *supra* note 222, ¶ 132.03[2][i][ii].

246. 383 U.S. 1, 12–24 (1966).

evidence or witnesses in the prior litigation.<sup>247</sup> The Court concluded that there is no “automatic formula” for assessing this prong and that “[i]n the end, decision will necessarily rest on the trial courts’ sense of justice and equity.”<sup>248</sup> Where the prior court has conducted a *Markman* hearing in which the parties were afforded the ability to present their positions and respond, the “full and fair opportunity to litigate” requirement has been satisfied.<sup>249</sup>

Decisions by the International Trade Commission (ITC) do not have preclusive effect on district courts, although district courts have discretion to attribute persuasive effect to ITC rulings. Congress passed the Trade Reform Act of 1974, amending the Tariff Act of 1930 to allow respondents in ITC proceedings to plead, and the ITC to consider, all legal and equitable defenses, including patent invalidity and unenforceability.<sup>250</sup> In authorizing the Commission to consider these defenses, Congress stated:

[I]n patent-based cases, the Commission considers, for its own purposes under section 337, the status of imports with respect to the claims of U.S. patents. The Commission’s findings neither purport to be, nor can they be, regarded as binding interpretations of the U.S. patent laws in particular factual contexts. Therefore, it seems clear that any disposition of a Commission action by a Federal Court should not have res judicata or collateral estoppel effect in cases before such courts.<sup>251</sup>

Based on this legislative history, the Federal Circuit determined that Congress did not intend decisions of the ITC on patent issues to have preclusive effect.<sup>252</sup>

#### d) Determination Was Essential to the Final Judgment

The final prong of the issue preclusion test has attracted the most controversy in the claim construction context. It can be divided into two

---

247. *Blonder-Tongue Labs., Inc. v. Univ. of Ill. Found.*, 402 U.S. 313, 329–34 (1971).

248. *Id.* at 334.

249. *See Kollmorgen Corp. v. Yaskawa Elec. Corp.*, 147 F. Supp. 2d 464, 466 (W.D. Va. 2001) (stating that a lengthy *Markman* hearing on the claim construction satisfied the requirement); *TM Patents, L.P. v. Int’l Bus. Machs. Corp.*, 72 F. Supp. 2d 370, 375 (S.D.N.Y. 1999) (noting that both parties agreed that there was a full and fair opportunity to litigate because a *Markman* hearing occurred).

250. *See Trade Act of 1974*, Pub. L. No. 93-618, 88 Stat. 1978 (1974).

251. S. REP. NO. 93-1298, at 196 (1974), *reprinted in* 1974 U.S.C.C.A.N. 7186, 7329.

252. *See Tex. Instruments Inc. v. Cypress Semiconductor Corp.*, 90 F.3d 1558, 1568 (Fed. Cir. 1996); *Tandon Corp. v. U. S. Int’l Trade Comm’n*, 831 F.2d 1017, 1019 (Fed. Cir. 1987) (“[O]ur appellate treatment of decisions of the Commission does not estop fresh consideration by other tribunals.”).

useful, separate inquires: whether (1) the prior ruling was “final”; and (2) the prior ruling was essential to the judgment.

i) Finality

The question of whether a prior claim construction constitutes a final judgment can be characterized along a spectrum. At the easier end of the spectrum, where the court in the prior proceeding interprets the pertinent claim language and issues a final, appealable judgment on validity or infringement, the finality requirement is satisfied.<sup>253</sup> The preclusive effect of prior summary judgment, preliminary injunction, and settlement dispositions are less clear.

(1) *Summary Judgment*

Issue preclusion can also arise out of a ruling that grants summary judgment,<sup>254</sup> although denial of summary judgment or a grant of partial summary judgment usually does not have preclusive effect.<sup>255</sup>

(2) *Preliminary Injunction*

The Federal Circuit held in *Transonic Systems, Inc. v. Non-Invasive Medical Technologies Corp.*<sup>256</sup> that claim constructions conducted for purposes of a preliminary injunction ruling are not binding, even in the same litigation. Drawing upon the Supreme Court’s statement in *University of Texas v. Camenisch*,<sup>257</sup> that “findings of fact and conclusions of law made by a court granting a preliminary injunction are not binding at trial on the merits,” the Federal Circuit views claim constructions reached during appeals from a grant of a preliminary injunction to be tentative and hence not binding on

---

253. See, e.g., *In re Freeman*, 30 F.3d 1459, 1466 (Fed. Cir. 1994) (“[J]udicial statements regarding the scope of patent claims are entitled to collateral estoppel effect in a subsequent infringement suit only to the extent that determination of scope was essential to a final judgment on the question of validity or infringement.” (quoting *A.B. Dick Co. v. Burroughs Corp.*, 713 F.2d 700, 704 (Fed. Cir. 1983))); *Home Diagnostics Inc. v. Lifescan, Inc.*, 120 F. Supp. 2d 864, 870 (N.D. Cal. 2000) (noting there must be a final judgment on validity or infringement for collateral estoppel to apply).

254. See *Stevenson v. Sears, Roebuck & Co.*, 713 F.2d 705, 712 (Fed. Cir. 1983); *Sec. People, Inc. v. Medeco Sec. Locks, Inc.*, 59 F. Supp. 2d 1040, 1044–45 (N.D. Cal. 1999), *aff’d mem.*, 243 F.3d 555 (Fed. Cir. 2000).

255. See *Syntex Pharms. Int’l, Ltd. v. K-Line Pharms., Ltd.*, 905 F.2d 1525, 1526 (Fed. Cir. 1990) (noting that an order granting summary judgment of infringement of a patent and denying the alleged infringer’s motion for summary judgment of invalidity did not present an appealable final judgment).

256. 75 F. App’x 765, 774 (Fed. Cir. 2003).

257. 451 U.S. 390, 395 (1981).

the district court in subsequent proceedings.<sup>258</sup> Therefore, claim constructions made in the context of preliminary injunction motions should not be considered final judgments, as the district court remains “at liberty to change the construction of a claim term as the record in a case evolves after a preliminary injunction appeal.”<sup>259</sup>

### (3) *Settlement*

Courts are deeply divided on the issue of finality when the outcome of the prior proceeding is a settlement. Several courts have interpreted the “finality” requirement liberally and functionally, looking to whether the previous judgment is sufficiently firm to be accorded preclusive effect. In *TM Patents, L.P. v. IBM Corp.*,<sup>260</sup> the defendant sought to hold the patentee to a claim construction rendered in a case resolved through settlement. While recognizing that the settlement did not result in a final appealable judgment, the court nonetheless determined that the prior claim construction was entitled to preclusive effect.<sup>261</sup> Seeking to elevate substance over form, the court focused upon the careful consideration of the issues during the prior litigation and drew upon the Supreme Court’s policy ruminations in *Markman* emphasizing the importance of “uniformity in the treatment of a given patent.”<sup>262</sup> The court recast “finality” for issue preclusion purposes as whether the prior litigation passed a stage for which there is “no really good reason for permitting [an issue] to be litigated again.”<sup>263</sup> The court noted as well that the patentee voluntarily entered into the settlement agreement and the *Markman* ruling was not vacated as part of the settlement.<sup>264</sup>

Although some other courts have since followed *TM Patents*’ application of collateral estoppel in the context of settlements following *Markman* rulings,<sup>265</sup> a contrary line of cases emerged holding that *Markman* rulings from cases that settled were not final and hence not properly entitled to preclusive

---

258. See *Jack Guttman, Inc. v. Kopykake Enters.*, 302 F.3d 1352, 1361 (Fed. Cir. 2002) (“District courts may engage in a rolling claim construction, in which the court revisits and alters its interpretation of the claim terms as its understanding of the technology evolves.”); *Transonic Sys.*, 75 F. App’x at 774.

259. *Transonic Sys.*, 75 F. App’x at 774.

260. 72 F. Supp. 2d 370, 375–77 (S.D.N.Y. 1999).

261. *Id.* at 378–79.

262. See *Markman v. Westview Instruments, Inc.*, 517 U.S. 370, 390 (1996).

263. *TM Patents*, 72 F. Supp. 2d at 376 (quoting *Lummus Co. v. Commonwealth Oil Ref. Co.*, 297 F.2d 80, 89 (2d Cir. 1961)).

264. *Id.* at 378.

265. See, e.g., *Edberg v. CPI-The Alternative Supplier, Inc.*, 156 F. Supp. 2d 190, 195 (D. Conn. 2001).

effect.<sup>266</sup> The cases read the Supreme Court's policy discussion in the *Markman* case as merely recognizing the importance of uniformity, not changing the fundamental principles for issue preclusion. In *Graco Children's Products, Inc. v. Regalo International*, the district court expressed concern that granting preclusive effect to cases settled after claim constructions might discourage settlement and encourage appeals by patentees who obtained favorable verdicts but nonetheless needed to correct what they believed to be unduly narrow or otherwise flawed claim constructions.<sup>267</sup>

The preclusive effect of claim construction rulings in cases resolved by settlement came before the Federal Circuit in *RF Delaware, Inc. v. Pacific Keystone Technologies, Inc.*<sup>268</sup> Without expressly resolving the district court conflict, the Federal Circuit, applying Eleventh Circuit law, applied a stringent standard to the question of finality: "if the parties to a suit enter into an extrajudicial settlement or compromise, there is no judgment, and future litigation is not barred by res judicata or collateral estoppel . . . ."<sup>269</sup> The Federal Circuit drew no implication from the Supreme Court's *Markman* language seized upon by the *TM Patents* court. Nonetheless, the court included some language inclining toward a functional approach to finality: "[f]or purposes of issue preclusion . . . , 'final judgment' includes any prior adjudication of an issue in another action that is determined to be sufficiently firm to be accorded conclusive effect."<sup>270</sup> Whether a decision is "sufficiently firm" depends on whether the parties were "fully heard."<sup>271</sup> The Federal Circuit noted that the Eleventh Circuit held that a prior district court order issued after an evidentiary hearing satisfied the finality standard because the district court notified the parties of possible preclusive effect, considered the findings final, and entered a final order approving the proposed settlement.<sup>272</sup> In *RF Delaware*, the Federal Circuit denied preclusive effect of the earlier *Markman* ruling on the grounds that there was no evidence that a *Markman* hearing had been conducted in the earlier case, the parties did not have

---

266. See *Kollmorgen Corp. v. Yaskawa Elec. Corp.*, 147 F. Supp. 2d 464, 470 (W.D. Va. 2001); *Graco Children's Prods., Inc. v. Regalo Int'l*, 77 F. Supp. 2d 660, 664–65 (E.D. Pa. 1999).

267. 77 F. Supp. 2d at 664.

268. 326 F.3d 1255 (Fed. Cir. 2003); see also *Dana v. E.S. Originals, Inc.*, 342 F.3d 1320 (Fed. Cir. 2003).

269. *RF Del.*, 326 F.3d at 1261–62 (Fed. Cir. 2003) (quoting *Kaspar Wire Works, Inc. v. Leco Eng'g & Mach., Inc.*, 575 F.2d 530, 542 (5th Cir. 1978)) (emphasis in original).

270. *Id.* at 1261 (quoting *Christo v. Padgett*, 223 F.3d 1324, 1339 n.47 (11th Cir. 2002) (citing RESTATEMENT (SECOND) OF JUDGMENTS § 13 (1980))).

271. *Id.*

272. *Id.* (quoting *Christo*, 223 F.3d at 1339).

notice that the court's order could have preclusive effect, and no final order approving the settlement was ever entered.<sup>273</sup>

The Federal Circuit further addressed the preclusive effect of stipulated constructions and settlements in *Pfizer, Inc. v. Teva Pharmaceuticals, USA, Inc.*<sup>274</sup> Because the parties in the prior proceeding had stipulated that the agreed claim interpretation was for purposes of that litigation only, the Federal Circuit held that the agreement could not preclude litigation in a later case.<sup>275</sup> Looking to jurisprudence on the interpretation of consent decrees, the court declared that “ ‘the scope of a consent decree must be discerned within its four corners’ and the conditions upon which a party has consented to waive its right to litigate particular issues ‘must be respected.’ ”<sup>276</sup>

ii) Essential to the Final Judgment

A final requirement for a prior *Markman* ruling to foreclose later interpretation over a claim term is that the earlier construction was essential to the final judgment. When the prior action turns upon resolution of a particular claim term or terms, the court's construction of other claim terms is “merely dictum, and therefore has no issue preclusive effect.”<sup>277</sup> To have a preclusive effect, the earlier court's interpretation of the particular claim had to be the reason for the previous outcome.<sup>278</sup>

A related principle is that issues of claim construction that cannot be appealed cannot be accorded preclusive effect.<sup>279</sup> Thus, courts will not attach preclusive effect where a patentee loses on the issue of claim interpretation but nonetheless prevails on validity and infringement because the patentee lacked a basis for appealing the *Markman* ruling.<sup>280</sup>

e) Reasoned Deference as a Prudent Approach to Issue Preclusion

Where the basis for applying issue preclusion is open to question, many courts have taken the approach of according prior *Markman* rulings

---

273. *Id.* at 1261–62.

274. 429 F.3d 1364 (Fed. Cir. 2005).

275. *Id.* at 1376.

276. *Id.* (quoting *United States v. Armour & Co.*, 402 U.S. 673, 682 (1971) and citing *In re Graham*, 973 F.2d 1089, 1097 (3d Cir. 1992) (noting that the Third Circuit defers to the intent of parties concerning the preclusive effect of agreed facts or claims in consent decrees and stipulations)).

277. *Phonometrics, Inc. v. N. Telecom Inc.*, 133 F.3d 1459, 1464 (Fed. Cir. 1998).

278. *See Jackson Jordan, Inc. v. Plasser Am. Corp.*, 747 F.2d 1567, 1577 (Fed. Cir. 1984).

279. *See Hartley v. Mentor Corp.*, 869 F.2d 1469, 1472 (Fed. Cir. 1989).

280. *See Graco Children's Prods., Inc. v. Regalo Int'l*, 77 F. Supp. 2d 660, 664–65 (E.D. Pa. 1999); *Schering Corp. v. Amgen, Inc.*, 35 F. Supp. 2d 375, 377 n.2 (D. Del. 1999), *aff'd in part*, 222 F.3d 1347 (Fed. Cir. 2000).

“reasoned deference” in assessing the disputed claim terms.<sup>281</sup> Where no new arguments are offered, no new foundation is laid, and there has been no change in the applicable standards for construing claims, courts generally adopt the prior construction unless it is clearly unsound. Where new argument and evidence is adduced, then the review is more probing and independent. Even in cases in which courts have determined that collateral estoppel applies, they have nonetheless made some independent assessment of claim construction. Thus, even the *TM Patents* court, which held that a *Markman* ruling from a earlier case that settled prior to trial precluded relitigation of claim meaning, used the “reasoned deference” approach as a judicial backstop: “Finally, I have to observe that this issue of collateral estoppel . . . is of marginal practical importance, because I agree with just about everything Judge Young did when he construed the claims in the EMC action.”<sup>282</sup>

### 3. *Judicial Estoppel*

The Federal Circuit has recognized the applicability of the equitable doctrine of judicial estoppel in the context of claim construction.<sup>283</sup> As an equitable doctrine, the contours of judicial estoppel are relatively flexible. Although “[t]he circumstances under which judicial estoppel may appropriately be invoked are probably not reducible to any general formulation of principle,”<sup>284</sup> the Supreme Court has emphasized three factors to consider in determining whether the doctrine applies: (1) whether a party’s later position is “clearly inconsistent” with its earlier position; (2) whether the party succeeded in persuading a court to accept that party’s earlier position, so that judicial acceptance of an inconsistent position in a later proceeding would create “the perception that either the first or second court was misled”; and (3) whether the party seeking to assert an inconsistent position

---

281. *See, e.g.*, *Finisar Corp. v. DirecTV Group, Inc.*, 523 F.3d 1323, 1329 (Fed. Cir. 2008) (noting that “in the interest of uniformity and correctness,” the Federal Circuit “consults the claim analysis of different district courts on the identical terms in the context of the same patent”); *Visto Corp. v. Sproqit Techs., Inc.*, 445 F. Supp. 2d 1104, 1108 (N.D. Cal. 2006) (observing that in cases of interjurisdictional uniformity, a prior interpretation is entitled to “‘reasoned deference’ . . . turning on the persuasiveness of the order; ‘in the end, [however, the Court] will render its own independent claim construction’”) (citation omitted and alteration in original).

282. *TM Patents, L.P. v. Int’l Bus. Machs. Corp.*, 72 F. Supp. 2d 370, 379 (S.D.N.Y. 1999).

283. *See* *Biomedical Patent Mgmt. Corp. v. Cal. Dep’t of Health Servs.*, 505 F.3d 1328, 1341 (Fed. Cir. 2007); *Harris Corp. v. Ericsson*, 417 F.3d 1241, 1252 n.2 (Fed. Cir. 2005); *RF Del., Inc. v. Pac. Keystone Techs., Inc.*, 326 F.3d 1255, 1262 (Fed. Cir. 2003).

284. *Allen v. Zurich Ins. Co.*, 667 F.2d 1162, 1166 (4th Cir. 1982).

would derive an unfair advantage or impose an unfair detriment on the opposing party if not estopped.<sup>285</sup>

The requirements for judicial estoppel partially overlap with the standard for issue preclusion (such as the element of identity of issues), but there are substantial differences as well. Unlike issue preclusion, judicial estoppel does not require strict mutuality,<sup>286</sup> or even that the issue was actually litigated in the prior proceeding.<sup>287</sup> On the other hand, judicial estoppel typically requires strong evidence of improper intent to mislead a tribunal.<sup>288</sup>

Judicial estoppel is also closely related to equitable estoppel.<sup>289</sup> Unlike equitable estoppel, a party asserting judicial estoppel does not have to prove detrimental reliance because judicial estoppel is designed to protect the integrity of the courts rather than any interests of the litigants.<sup>290</sup> Therefore, judicial estoppel may apply in a particular case “where neither collateral estoppel nor equitable estoppel . . . would apply.”<sup>291</sup>

As with issue preclusion and other non-patent procedural issues, courts apply the standards for judicial estoppel developed by their regional circuit.<sup>292</sup> Such standards vary across the circuits. For example, although most circuits do not require mutuality of judicial estoppel, some courts limit the doctrine to those who were party to (or in privity with a party to) the prior proceeding.<sup>293</sup> The relative importance of particular factors varies as well. Some circuits consider intent—whether the inconsistency in position was for the purpose of gaining unfair advantage—to be most determinative.<sup>294</sup>

---

285. *See* *New Hampshire v. Maine*, 532 U.S. 742, 750–51 (2001).

286. *Ryan Operations G.P. v. Santiam–Midwest Lumber Co.*, 81 F.3d 355, 360 (3d Cir. 1996) (stating that privity is not required for judicial estoppel).

287. *Lowery v. Stovall*, 92 F.3d 219, 223 n.3 (4th Cir. 1996).

288. *See* *Scarano v. Cent. R. Co.*, 203 F.2d 510, 513 (3d Cir. 1953) (noting that judicial estoppel prevents parties from “playing ‘fast and loose with the courts’” by forbidding “intentional self-contradiction . . . as a means of obtaining unfair advantage” (quoting *Stretch v. Watson*, 69 A.2d 596, 603 (N.J. Super. Ct. Ch. Div. 1949))).

289. *See id.* at 514 n.2.

290. *Teledyne Indus., Inc. v. NLRB*, 911 F.2d 1214, 1220 (6th Cir. 1990).

291. *Allen v. Zurich Ins. Co.*, 667 F.2d 1162, 1166–67 (4th Cir. 1982).

292. *See* *Lampi Corp. v. Am. Power Prods., Inc.*, 228 F.3d 1365, 1377 (Fed. Cir. 2000).

293. *See* *Nichols v. Scott*, 69 F.3d 1255, 1272 n.33 (5th Cir. 1995).

294. *See* *Lowery v. Stovall*, 92 F.3d 219, 224 (4th Cir. 1996). The Federal Circuit holds that judicial estoppel does not normally prevent a party from altering on appeal an unsuccessful position on claim construction that it advocated before the trial court. *See* *RF Del., Inc. v. Pac. Keystone Tech., Inc.*, 326 F.3d 1255, 1262 (Fed. Cir. 2003) (“The doctrine of judicial estoppel is that where a party *successfully* urges a particular position in a legal proceeding, it is estopped from taking a contrary position in a subsequent proceeding where its interests have changed.”) (emphasis in original).

#### 4. *Stare Decisis*

Since claim construction is considered a question of law, lower courts must adhere to prior claim construction determinations by the Federal Circuit, even if the claim construction is applied to a party who was not involved in the prior litigation.<sup>295</sup> The Supreme Court considered this a virtue of categorizing claim construction as a matter of law: “treating interpretive issues as purely legal will promote (though it will not guarantee) intrajurisdictional certainty through the application of stare decisis on those questions not yet subject to interjurisdictional uniformity under the authority of the single appeals court.”<sup>296</sup>

A decision of a district court is not binding precedent on a different judicial district, the same judicial district, or even the same judge in a different case under the doctrine of stare decisis. Rather, stare decisis requires only that the later court encountering the issue give consideration and careful analysis to that sister court’s decision where applicable to a similar fact pattern.<sup>297</sup> Courts sometimes accord prior decisions from within their district somewhat greater consideration than those decided outside the district.<sup>298</sup>

Just as issue preclusion requires an issue to have been actually *litigated* in order for collateral estoppel to attach, stipulations of claim meaning may not be entitled to stare decisis effect “because it is only the judiciary—not the parties—that declares what the law is.”<sup>299</sup> The court in *Amgen, Inc. v. F. Hoffmann–La Roche Ltd.* noted, however, that “[s]uch agreements, of course, may, where appropriate, implicate judicial estoppel and, where a final

---

295. See *Amgen, Inc. v. F. Hoffmann–La Roche Ltd.*, 494 F. Supp. 2d 54, 59–60 (D. Mass. 2007); *Tate Access Floors, Inc. v. Interface Architectural Res., Inc.*, 185 F. Supp. 2d 588, 595 n.4 (D. Md. 2002); *Wang Labs., Inc. v. Oki Elec. Indus. Co.*, 15 F. Supp. 2d 166, 175 (D. Mass. 1998) (holding a prior Federal Circuit claim construction binds a party that was not a party to (or allowed intervention in) prior litigation interpreting the claim term in question).

296. *Markman v. Westview Instruments, Inc.*, 517 U.S. 370, 391 (1996); see also *Visto Corp. v. Sproqit Techs., Inc.*, 445 F. Supp. 2d 1104, 1108 (N.D. Cal. 2006) (observing that “interjurisdictional uniformity” refers to claim constructions reviewed by the Federal Circuit).

297. See *Amgen*, 494 F. Supp. 2d at 60 (D. Mass. 2007) (citing *United States v. Rodriguez–Pacheco*, 475 F.3d 434, 441 (1st Cir. 2007)); *Tex. Instruments, Inc. v. Linear Techs. Corp.*, 182 F. Supp. 2d 580, 589 (E.D. Tex. 2002); cf. *Finisar Corp. v. DirecTV Group, Inc.*, 523 F.3d 1323, 1329 (Fed. Cir. 2008) (noting that “[i]n the interests of uniformity and correctness,” the Federal Circuit “consults the claim analysis of different district courts on the identical terms in the context of the same patent”).

298. See, e.g., *Visto*, 445 F. Supp. 2d at 1107–08 (noting that intra-judicial uniformity warrants an even higher level of deference); *Verizon Cal. Inc. v. Ronald A. Katz Tech. Licensing, L.P.*, 326 F. Supp. 2d 1060, 1069 (C.D. Cal. 2003).

299. *Amgen*, 494 F. Supp. 2d at 70 n.1.

judgment occurs, the doctrine of issue preclusion.<sup>300</sup> Also as with issue preclusion, stare decisis applies only to rulings that were necessary to the decision rendered.<sup>301</sup>

A distinct tension arises when courts look to prior *Markman* rulings under the doctrine of stare decisis in circumstances that do not satisfy the more exacting requirements of issue preclusion. In practice, courts have alleviated this strain by affording a party who did not participate in that earlier action a full and fair opportunity to be heard in the later proceeding. At the same time, the court can be mindful of prior rulings.<sup>302</sup>

### III. CLAIM CONSTRUCTION PROCEDURE

As in most areas of litigation, procedure plays a critical role in the quality and efficiency of claim construction and the ultimate resolution of patent disputes. In the decade plus since *Markman*, courts have experimented with various approaches to the claim construction process. Most notably, the Northern District of California developed Patent Local Rules (hereinafter “PLRs”) for the primary purpose of structuring the disclosure of contentions leading up to *Markman* hearings.<sup>303</sup> Eleven other districts have since adopted PLRs modeled in varying degrees on the Northern District of California’s PLRs.<sup>304</sup> Beyond PLRs, courts have experimented with different approaches to the timing of *Markman* hearings; the use of tutorials, experts, and advisors in claim construction; and integrating *Markman* determinations with resolution of dispositive motions that can turn on claim construction. Drawing upon our survey of court practices, meetings with judges in the most patent-intensive districts,<sup>305</sup> and discussions with patent litigators, this

---

300. *Id.*

301. *See* Miken Composites, LLC v. Wilson Sporting Goods Co., 515 F.3d 1331, 1338 n.\* (Fed. Cir. 2008); Zenith Radio Corp. v. United States, 783 F.2d 184, 187 (Fed. Cir. 1986) (holding that stare decisis applied where resolution of issue was a “necessary predicate” to earlier Federal Circuit ruling).

302. *See Tex. Instruments*, 182 F. Supp. 2d at 590.

303. *See* N.D. CAL. PATENT LOC. R.

304. *See* D. MASS. PATENT LOC. R.; D. N.J. PATENT LOC. R.; E.D. MO. PATENT LOC. R. (Judge Charles Shaw); E.D. N.C. PATENT LOC. R.; E.D. TEX. PATENT LOC. R.; N.D. GA. PATENT LOC. R.; N.D. ILL. PATENT LOC. R. (proposed); S.D. CAL. PATENT LOC. R.; S.D. TEX. PATENT LOC. R.; W.D. PA. PATENT LOC. R.; W.D. WASH. PATENT LOC. R.

305. Between December 2007 and August 2008, the authors met with district judges and magistrate judges in the Northern District of California, Central District of California, District of Delaware, Northern District of Illinois, District of New Jersey, Southern District of New York, Eastern District of Texas, and Eastern District of Virginia, as well as the Federal Circuit, to discuss the range of patent case management practices.

section explores the landscape of case management approaches and describes established and emerging best practices for the process of claim construction.

#### A. PATENT LOCAL RULES

In an effort to provide fair and efficient management of patent cases, some districts have adopted PLRs<sup>306</sup> or have adopted standard practices under the Federal Rules of Civil Procedure and Civil Local Rules that have markedly affected the conduct of patent cases (e.g., Eastern District of Virginia<sup>307</sup>). The impetus for PLRs arose out of a clash between the liberal notice pleading policy underlying the Federal Rules of Civil Procedure and the need for patent litigants to have more specific notice of the issues they were litigating.<sup>308</sup> Under the Federal Rules of Civil Procedure, a patent plaintiff need only plead that a defendant is infringing its patent.<sup>309</sup> The plaintiff has not traditionally been required to specify which claims are infringed. Nor has the plaintiff needed to plead its theory of the meaning of the claim terms and the features of the defendant's products (or even the products themselves) that are alleged to infringe. Because a plaintiff may assert multiple claims in multiple patents, a defendant reading a notice pleading complaint is typically left to guess as to the boundaries of a plaintiff's case and the available defenses.

A patent plaintiff reading a notice pleading answer and counterclaim is equally in the dark about the substance of the defendant's case. The defendant, for example, need not identify the prior art on which its invalidity defense relies. Nor does the defendant have to plead its theories of claim

---

306. See generally James Ware & Brian Davy, *The History, Content, Application and Influence of the Northern District of California's Patent Local Rules*, 25 SANTA CLARA COMPUTER & HIGH TECH. L.J. 965 (2009) (providing a detailed account of the evolution of the Northern District of California's Patent Local Rules and a detailed appendix comparing the principal provisions of patent local rules throughout the districts).

307. See T.S. Ellis III, Presentation at the Proceedings of the 1999 Summit Conference on Intellectual Property: Quicker and Less Expensive Enforcement of Patents: United States Courts (1999), in *Streamlining International Intellectual Property* 11, 12–14 (5 CASRIP Publ'n Series 2000), available at <http://www.law.washington.edu/casrip/symposium/Number5/pub5atcl2.pdf>; see also Robert E. Scott & George G. Triantis, *Anticipating Litigation in Contract Design*, 115 YALE L.J. 814, 829 n.34 (2006) (describing Eastern District of Virginia's reputation for rapid resolution of litigation).

308. See *O2 Micro Int'l Ltd. v. Monolithic Power Sys.*, 467 F.3d 1355, 1365–66 (Fed. Cir. 2006).

309. FED. R. CIV. P. 8; FED. R. CIV. P. Form 16; see also *Phonometrics, Inc. v. Hospitality Franchise Sys., Inc.*, 203 F.3d 790, 794 (Fed. Cir. 2000); *Gammino v. Cellco P'ship*, No. 04-4303, 2005 WL 2397168, at \*1–3 (E.D. Pa. Sept. 27, 2005). *But cf.* *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 556 (2007) (raising the quantum of factual matter that must be pled (in the context of a Sherman Act cause of action) to survive a motion to dismiss).

construction or which combinations of prior art references might invalidate each of the claims. Only the defense of unenforceability due to inequitable conduct in procurement of the patent has to be pled with particularity, because it is viewed as a species of fraud.<sup>310</sup>

Initial disclosures required under Rule 26 do not alleviate this problem. Routine discovery procedures such as service of contention interrogatories or expert discovery ultimately could provide the necessary information. However, contention interrogatories are often not required to be meaningfully answered until the late stages of discovery. Expert discovery provides an opportunity to focus the case, but arises on the verge of trial. The associated delay can be highly prejudicial to litigants.

As a result, absent forced, early substantive disclosure, patent litigants have been known to engage in a “shifting sands” approach to litigation based on “vexatious shuffling of positions.”<sup>311</sup> That is, litigants may offer initial, substantially hedged, theories of infringement or invalidity, only to change those theories later by asserting different patent claims, different prior art, or different claim constructions if their initial positions founder. Resulting extensions of fact and expert discovery can unduly prolong the litigation, sapping the court’s and the parties’ resources.

PLRs were developed to facilitate efficient discovery by requiring patent litigants to promptly disclose the bases underlying their claims. By requiring parties to disclose contentions in an orderly, sequenced manner, PLRs prevent the “shifting sands” tendencies. Neither litigant can engage in a strategic game of saying it will not disclose its contentions until the other side reveals its arguments. In discussing the Northern District of California’s PLRs, the Federal Circuit explained that they are designed to require

both the plaintiff and the defendant in patent cases to provide early notice of their infringement and invalidity contentions, and to proceed with diligence in amending those contentions when new information comes to light in the course of discovery. The rules thus seek to balance the right to develop new information in discovery with the need for certainty as to the legal theories.<sup>312</sup>

---

310. *See* *MedImmune, Inc. v. Centocor, Inc.*, 271 F. Supp. 2d 762, 771–72 (D. Md. 2003); *Environ Prods., Inc. v. Total Containment, Inc.*, 951 F. Supp. 57, 59 (E.D. Pa. 1996).

311. *See* *LG Elecs., Inc. v. Q-Lity Computer, Inc.*, 211 F.R.D. 360, 367 (N.D. Cal. 2002).

312. *O2 Micro*, 467 F.3d at 1365–66; *see also* *Nova Measuring Instruments Ltd. v. Nanometrics, Inc.*, 417 F. Supp. 2d 1121, 1122–23 (N.D. Cal. 2006) (“The [patent local] rules are designed to require parties to crystallize their theories of the case early in the litigation and to adhere to those theories once they have been disclosed.”).

PLRs adopted by a district, or by an individual judge as a standing order or a case-specific order, supplement the Federal Rules of Civil Procedure. Courts may modify the procedures dictated by PLRs as necessary to suit the issues presented in a particular case.<sup>313</sup> All modifications, as well as the rules or standing orders, must be consistent with Federal Circuit case law to the extent an issue “pertains to or is unique to patent law.”<sup>314</sup> For example, Federal Circuit law was applied in cases addressing whether claim charts exchanged by parties pursuant to PLRs could be amended to add new statutory bases for invalidity and infringement.<sup>315</sup> In these situations, the Federal Circuit held that the sufficiency of notice regarding defenses or theories of liability under specific statutory provisions of patent law “clearly implicat[ed] the jurisprudential responsibilities of this court within its exclusive jurisdiction.”<sup>316</sup>

Chart 10 depicts a typical timeline for a patent case utilizing patent-specific initial disclosures, a structured claim construction briefing process including a joint claim construction statement, and a *Markman* hearing. The process depicted here is consistent with the requirements of PLRs in the Northern District of California.

---

313. See, e.g., N.D. CAL. PATENT LOC. R. 1–2.

314. See *O2 Micro*, 467 F.3d at 1364 (citing *Sulzer Textil A.G. v. Picanol N.V.*, 358 F.3d 1356, 1363 (Fed. Cir. 2004)).

315. *Genentech Inc. v. Amgen Inc.*, 289 F.3d 761, 774 (Fed. Cir. 2002); *Advanced Cardiovascular Sys., Inc. v. Medtronic, Inc.*, 265 F.3d 1294, 1303 (Fed. Cir. 2001).

316. See also *Advanced Cardiovascular*, 265 F.3d at 1303; *In re Spalding Sports Worldwide, Inc.*, 203 F.3d 800, 803–04 (Fed. Cir. 2000) (holding that Federal Circuit law applies when determining the applicability of the attorney–client privilege to an invention record because it implicates the substantive patent issue of inequitable conduct).

**Chart 10: Patent Local Rules Timetable, Northern District of California**

(1) Case Management Conference	Set by Court	Patent Local Rule
(2) Disclosure of Asserted Claims and Infringement Contentions	Within 10 days of (1)	3-1 & 3-2
(3) Invalidity Contentions	Within 45 days of (2)	3-3 & 3-4
(4) Identify Claim Terms to be Construed	Within 10 days of (3)	4-1
(5) Preliminary Claim Constructions	Within 20 days of (4)	4-2
(6) Joint Claim Construction Statement	Within 60 days of (3)	4-3
(7) Close of Claim Construction Discovery	Within 30 days of (6)	4-4
(8) Opening Claim Construction Brief	Within 45 days of (6)	4-5(a)
(9) Responsive Claim Construction Brief	Within 14 days of (8)	4-5(b)
(10) Reply Claim Construction Brief	Within 7 days of (9)	4-5(c)
(11) <i>Markman</i> Hearing	Within 14 days of (10)	4-6
(12) Claim Construction Order	TBD by Court	
(13) Produce Advice of Counsel, if any	Within 50 days of (12)	3-7

An accelerated timeline may be appropriate for less complex cases, for example where the technology is quite simple or there is little dispute as to the structure, function, or operation of accused devices. Under a particularly streamlined plan, the parties would not make patent-specific initial disclosures or file joint claim construction statements. The court might also forgo a *Markman* hearing and address claim construction as part of summary judgment.<sup>317</sup> Chart 11 provides an example of such a timeline. The decision to adopt an accelerated timeline can best be made after discussion with the parties of the substantive issues that will drive the case.

---

317. See *infra* Section III.D.

Chart 11: Accelerated Patent Case Management Timeline

(1) Case Management Conference (CMC)	Set by court
(2) Produce Opinion of Counsel, if any	Within 2 months after CMC
(3) Close of Fact Discovery	5 months after CMC
(4) Close of Expert Discovery	2 months after (3)
(5) Opening Briefs on Claim Construction and Summary Judgment	Within 30 days of (4)
(6) Responsive Briefs on Claim Construction and Summary Judgment	Within 14 days of (5)
(7) Reply Briefs on Claim Construction and Summary Judgment	Within 7 days of (6)
(8) Claim Construction and Summary Judgment Hearing	Within 14 days of (7)
(9) Claim Construction and Summary Judgment Order	TBD by court

#### B. TIMING OF *MARKMAN* HEARINGS

Perhaps the most important case management decision relating to the *Markman* process is its timing. More than a decade of practice has taught important lessons on when to hold the *Markman* hearing and has shown the need for flexibility to accommodate the needs of different cases.

Early *Markman* hearings (i.e., within about five months of the case management conference) may be appropriate in some contexts. In cases that appear to present a well-crystallized question of claim construction that may resolve liability without the need for extensive discovery, an early *Markman* hearing may be advantageous. Providing parties with an early ruling on key claim construction issues can promote settlement and avoid the cost and burden of lengthy discovery. However, in practice, these advantages are often outweighed by several disadvantages. Knowing what issues to present at a *Markman* hearing frequently requires extensive discovery into the nature of the accused device and of the prior art. Thus, an early *Markman* ruling often will need revisiting when new issues emerge.

In practice, the dominant and recommended approach is to hold *Markman* hearings mid-way through, or toward the end, of fact discovery, but prior to expert discovery. This affords the advantage of allowing sufficient discovery in advance of claim construction proceedings to more fully identify the issues that need to be resolved. Such mid-phase *Markman* hearings allow a more focused expert discovery process (assuming that the *Markman* ruling is issued in advance). This approach avoids the need for requiring expert witnesses to prepare reports that address the range of potential claim constructions. Such reports can be especially difficult to prepare.

Furthermore, if the expert does not anticipate the ultimate claim construction, expert discovery might have to be redone following a *Markman* decision.

Some courts defer *Markman* hearings until completion of expert discovery and resolve the disputes in conjunction with summary judgment briefing or immediately before trial. Although there may be some advantages to holding a *Markman* hearing at or near the end of a case (such as framing claim construction disputes in the context of dispositive motions), in practice this approach has been found to have too many drawbacks. Furthermore, holding a late-phase *Markman* hearing may deprive litigants of enough time to settle the case before trial. Late-phase *Markman* rulings are likely to upset the experts' positions and may inject new issues into the case, especially where the court arrives at its own construction that does not squarely adopt what either of the parties proposed.<sup>318</sup>

#### C. STREAMLINING THE PRE-MARKMAN PROCESS

In order to promote efficient and effective *Markman* hearings, many courts address the procedures and ground rules for such proceedings at a relatively early stage in case management. PLRs place particular emphasis on timely and orderly identification of disputed claim terms. We begin this section with further discussion of best practices to bring those disputes and the parties' arguments to the surface prior to the *Markman* hearing. Depending on the complexity of the technology at issue, it is often useful to plan for technology tutorials in conjunction with the *Markman* proceeding. We discuss several practical issues relating to the timing, form, and conduct of such tutorials and the use of court-appointed experts to assist in claim construction.

##### 1. *Mandatory Disclosure of Positions*

Two primary goals of the procedures before a *Markman* hearing exist: (1) insurance that the parties' claim construction positions are squarely joined, reducing false and hidden disputes; and (2) resolution of any disputes about how the *Markman* hearing should be conducted so the hearing itself is efficient, helpful to the court, and without procedural disarray. The following practices have proven especially effective in accomplishing these objectives.

---

318. See *Magarl, LLC v. Crane Co.*, No. IP 02-0478-C-T/L, 1:03-CV-01255-JDT-TWL, 2004 U.S. Dist. LEXIS 24283, at \*44 (S.D. Ind. Sept. 29, 2004) (encouraging holding *Markman* hearings in advance of summary judgment briefing, because a "claim construction which precedes summary judgment could avoid unnecessary alternative briefing and evidentiary submissions, including expert witness testimony addressed to or based on rejected claim constructions").

a) Early Disclosure of Infringement and Invalidity Contentions

Requiring disclosure of infringement contentions at the start of the case focuses the disputes at issue for the *Markman* hearing. In jurisdictions that have not adopted PLRs, courts are free to build these disclosure requirements into their scheduling orders. These infringement contentions require the patentee to specify, among other things, each claim of each patent-in-suit that is allegedly infringed; each instrumentality that allegedly infringes each asserted claim; and a claim chart detailing where each element of an asserted claim is found in each accused instrumentality.<sup>319</sup>

With its infringement contentions, the party must produce, among other things, all documents evidencing the conception and reduction to practice of each asserted claim, along with documents sufficient to show the disclosure of the claimed inventions to others prior to filing of the patent application.<sup>320</sup> Similarly, the court can help focus *Markman* issues by requiring the alleged infringer to disclose invalidity contentions after receipt of the infringement contentions. This requires the alleged infringer to specify, among other things, the identity of each item of prior art that allegedly anticipates each asserted claim or renders it obvious, and any grounds for invalidity due to indefiniteness, enablement, or written description.<sup>321</sup> With its invalidity contentions, the accused infringer must produce all prior art not already of record, as well as documents sufficient to show the operation of the accused devices.

These disclosures force parties to crystallize their theories early in the case, and thereby to identify the matters that need to be resolved through the *Markman* hearing. They also help streamline discovery by mandating the disclosures that are core to patent cases, thus reducing the need for interrogatories, document requests, and contention depositions. Early infringement contentions can, however, lead to more discovery because they may occur before parties fully understand their own positions. In practice, this may result in under-production of evidence.

b) Disclosure of Claims to Construe and Proposed Constructions

A widespread problem in patent cases is that the parties' *Markman* briefing may not effectively join the issues to be litigated at the *Markman* hearing, or may not confront claim construction issues that that will ultimately be litigated at trial. To avoid this problem, it is advisable that the

---

319. See N.D. CAL. PATENT LOC. R. 3-1.

320. See, e.g., N.D. CAL. PATENT LOC. R. 3-2.

321. See 35 U.S.C. § 112 (2006).

court set a meet-and-confer schedule in its scheduling order to require parties to identify terms that need construction. These procedures help to ensure that the issues for the *Markman* hearing be specified in advance of the briefing cycle, as opposed to having issues disclosed for the first time in briefing. Ordering a meet-and-confer process also helps to ensure that the parties' briefing is not wastefully directed to false or merely hypothetical disputes. Ordering parties to disclose their claim construction positions also discourages "hidden" disputes that may otherwise arise at trial. This structured meet-and-confer process is part of the PLRs of the Northern District of California and the Eastern District of Texas, and is required within ten days of service of the invalidity contentions.<sup>322</sup>

As part of this process, the court's scheduling order should set a date for the parties to exchange proposed constructions of the identified terms. Setting this date approximately twenty days after exchanging lists of terms is appropriate. As part of this disclosure, some jurisdictions also require that the parties disclose their supporting evidence, including whether they will be relying on expert witnesses.<sup>323</sup>

c) Mechanisms for Limiting the Number of Claim Terms to Construe

Cases commonly involve multiple patents, dozens or even hundreds of claims, and multitudes of claim terms that may need construction. If left unmanaged, the sheer complexity of this tangle of terms can overwhelm the merits of a lawsuit. Courts should exercise their inherent case management authority to limit the number of claims and claim terms at issue, as appropriate.

At the *Markman* phase, courts have wide discretion to limit the number of claim terms at issue. Restricting the scope of the *Markman* hearing may have the benefit of focusing the court's attention on the key issues (which may dispose of the case), and of allowing a more prompt and well-reasoned ruling on the central matters in the case. Courts have experimented widely with various approaches to managing the scope of *Markman* hearings. By contrast, asking the parties to brief all the potential claim construction disputes invites false or inconsequential disputes, particularly because parties reflexively seek to avoid the risk of a waiver finding if they refrain from raising peripheral disputes.

---

322. See N.D. CAL. PATENT LOC. R. 4-1 to 4-3; E.D. TEX. PATENT LOC. R. 4-1 to 4-3.

323. See S.D. CAL. PATENT LOC. R. 4.1(d); E.D. TEX. PATENT LOC. R. 4-2(b).

The Northern District of California has recently adopted local rules requiring parties to identify “the terms whose construction will be most significant to the resolution of the case up to a maximum of 10.”<sup>324</sup> The ten-term limit is a default rule that can be adjusted upwards or downwards depending on the circumstances of the case. The number should vary depending on the number of patents in dispute. Ten can be high for single patent cases, but low for multi-patent cases. The parties are required to meet and confer to identify the ten most significant terms in dispute. In addition to any terms that the parties mutually agree upon as being the most significant, the parties are each allocated half of the remaining terms of the ten, and can identify additional terms they wish to have construed under this allocation. This is not a fixed limit altogether of the number of terms to be construed, and litigants may seek to construe terms at later phases in the case. However, for purposes of the main *Markman* hearing, this channeling of the most significant terms allows courts to deploy their resources most efficiently to resolve the key disputes in the case.

There are many factors that may influence whether to increase the number of terms to be construed. For example, means-plus-function claims generally must be construed in order to identify the corresponding structure in the specification.<sup>325</sup> Also, allowing each party to have a fixed number of claim terms to be construed may not make sense. In many cases, a plaintiff will assert dozens of patent claims, often out of multiple patents, and may not want to construe any of the terms, seeking to leave their interpretation to the jury. Typically, the defendant is the party with a greater interest in having claims construed, and it may be prejudicial to the defendant to limit its ability to only have ten claim terms construed (particularly where the plaintiff has asserted a large number of claims). Thus, a rigid, formulaic approach will not accommodate all cases, and the parties should be allowed, where appropriate, to structure the *Markman* proceedings in a flexible manner to suit the unique aspects of the case.

Other customary mechanisms for managing the scope of *Markman* proceedings include page limits on briefing, and time restrictions at the *Markman* hearing. Parties will naturally allocate limited presentation times (written or oral) to the key disputes, and limits on briefing or oral argument will have some effect at streamlining the *Markman* proceedings. However, parties may feel that they will be faced with a waiver situation if all disputed terms are not addressed at the *Markman* proceedings. In such cases, there will

---

324. N.D. CAL. PATENT LOC. R. 4-3(c).

325. See 35 U.S.C. § 112 para. 6 (2006).

inherently be a tendency to cram additional arguments into the written or oral presentations. Ultimately, this is a less helpful mechanism than limiting the number of terms that the court will address in the main *Markman* proceeding.

Courts risk upsetting trial dates and may invite reversal if they overly constrain or defer the *Markman* process. Ultimately, all material claim construction disputes must be ruled upon by the court for cases that go to trial.<sup>326</sup> It is legal error for the court to allow the parties to argue competing claim construction positions to the jury.<sup>327</sup> The more that outstanding claim construction issues are deferred until the late phases of litigation (or are not resolved until trial) the greater the likelihood of legal error and that trial will be a game of surprise. Resolving the material claim construction disputes well in advance of trial will prevent procedural aberrations from overwhelming the merits of a case and minimize the risk of reversal and the need for retrial.

d) Severance versus Postponement

In cases involving many patents, frequently with diverse technologies, courts have struggled to find ways to reduce the case to a manageable size that the court and a jury can handle in one trial. Often the court is able to persuade the parties to reduce the number of patents to be tried to a manageable number, but if that is unsuccessful, the court does not have the power to order a party not to pursue a patent claim it has lawfully filed. District courts typically have addressed this issue in the context of multi-patent disputes in one of two ways: (1) limiting the total number of disputed terms to be construed, and hoping that those terms will resolve the dispute; or (2) allowing the parties to select a limited subset of patents to be tried in the first instance, and severing the remaining patents for a subsequent trial if needed. The primary risk in the first approach is that the chosen terms will not resolve the dispute, in which case the court will be faced with two unattractive options: (1) either doing claim construction hurriedly at the end of the pretrial schedule, which disrupts expert reports, summary judgment, and other pretrial scheduling, or (2) postponing the trial for another round of claim construction. The Federal Circuit has made clear that the district court may not proceed to trial without resolving any remaining claim construction

---

326. See *O2 Micro Int'l. Ltd. v. Beyond Innovation Tech. Co.*, 521 F.3d 1351, 1360–61 (Fed. Cir. 2008).

327. See *CytoLogix Corp. v. Ventana Med. Sys., Inc.*, 424 F.3d 1168, 1172 (Fed. Cir. 2005) (“[B]y agreement the parties also presented expert witnesses who testified before the jury regarding claim construction, and counsel argued conflicting claim constructions to the jury. This was improper, and the district court should have refused to allow such testimony despite the agreement of the parties.”).

disputes.<sup>328</sup> In general, courts have gravitated toward the severance and stay option, and have found that the subsequent trials are not needed.

e) Recommended Approach: Mandatory Disclosure of Impact of Proposed Constructions

Many infringement and invalidity disputes hinge on legal questions of claim interpretation and can be properly resolved on summary judgment. Requiring parties to state the anticipated impact of their proposed constructions on the merits of the case enables the court to better appreciate the ramifications of its claim construction. Integrating claim construction with consideration of those dispositive motions dictated by claim construction streamlines adjudication.

We recommend that parties state the reasons for seeking construction of any terms that are litigated in the *Markman* process, regardless of whether they are being asserted for summary judgment purposes. This approach not only gives courts the context for making important rulings in the *Markman* process, but also minimizes unnecessary disputes. In practice, parties are often unable to articulate why their definition is materially different from their opponent's, but may nonetheless adhere to it. Left unresolved, these less-than-meaningful discrepancies in wording may result in wasteful briefing and unnecessary consumption of the court's time. Requiring disclosure of *why* these terms need to be construed should reduce false disputes. Where there is not a meaningful dispute underlying a party's request for a construction, courts may be well within their authority to decline construing that term.<sup>329</sup>

Terms that are to be construed for summary judgment purposes should be specifically identified, along with a statement of which party (or both) would be seeking summary judgment on the basis of that term, and why. As an example of the form of disclosure recommended, Table C illustrates a sample claim chart showing a term to be construed ("steering wheel"), along with the defendant's reasons for seeking summary judgment.

---

328. See *O2 Micro*, 521 F.3d at 1360–63.

329. See *Vivid Techs., Inc. v. Am. Sci. & Eng'g, Inc.*, 200 F.3d 795, 803 (Fed. Cir. 1999) ("AS & E is correct that although the claims are construed objectively and without reference to the accused device, only those terms need be construed that are in controversy, and only to the extent necessary to resolve the controversy.").

Table C: Summary Judgment Term: “Steering Wheel”

	Plaintiff	Defendant
Proposed construction	any device for directing a vehicle	a circular device for directing a vehicle
Summary Judgment Context (non-infringement)		Accused device lacks a circular steering device, so summary judgment of no infringement is proper.
Summary Judgment Context (invalidity)		If plaintiff's proposed construction prevails, then ABC reference anticipates the claims as a matter of law.

Many claim terms are not the focus of summary judgment motions, but will be the focus of claims or defenses to be presented at trial. There may also be collateral reasons for parties to seek construction of terms, such as ensuring that a defendant's future products will be safely outside the scope of an asserted patent. Courts should require the parties to disclose why they are seeking constructions of these other terms.

One approach used in some courts to focus the *Markman* inquiry is to conduct a short telephone conference with the parties after they file the list of terms to be construed and the reasons for their submission, prior to the briefing cycle. During this call, the court can state which summary judgment motions it is willing to entertain in connection with the *Markman* proceedings. Moreover, forcing the parties to explain why they need to have terms construed goes a long way towards eliminating unnecessary disputes. Minor disputes over wording choices can also be resolved in this manner.

This process integrates the summary judgment process and the *Markman* hearing. The court may wish to schedule summary judgment briefing in tandem with claim construction briefing, or may wish to stagger summary judgment briefing to take place shortly after the *Markman* hearing.

An open question is whether courts could or should penalize a party for failing to take advantage of opportunities to bring summary judgment in connection with the *Markman* process. We expect that parties would take advantage of a formalized summary judgment process in connection with *Markman*, and parties should be encouraged to do so. However, there are many reasons why parties may legitimately want to defer filing a summary judgment motion until later in the case, even where a claim construction question is at the heart of the dispute. It may be difficult to craft a summary judgment position until the claim construction ruling issues. Also, it is frequently desirable to close out fact discovery before filing summary

judgment motions to preclude unforeseen facts from being “lobbed in” to defeat a summary judgment motion. Courts should address with care any efforts to penalize a party that does not file an early summary judgment motion in connection with the *Markman* process.

2. *Use of Tutorials, Experts, and Advisors in Claim Construction*

Claim terms are interpreted from the perspective of a person having ordinary skill in the art at the time of invention. Thus, the parties will need to educate the court about the science, technology, and perspective of a person having ordinary skill in the art as of the time period of the invention. The most common vehicle for accomplishing this task is the use of technology tutorials typically done in connection with a *Markman* hearing. In addition, courts occasionally go a significant step further and appoint a technical advisor, special master, or expert for the court. Table D summarizes the principal characteristics of these educational aids.

Table D: Educating the Court and Court Appointed Experts

Nature of Expert/ Legal Authority	Process/Role	Procedural Safeguards
<b>1. Tutorial Process</b>	presented by counsel, experts for each side, or agreed expert demonstratives often useful (e.g., PowerPoint presentation, simulation video, CD that can be reviewed later)	typically scheduled within two weeks of <i>Markman</i> hearing usually best to allow each side to make their own presentation, with court actively questioning advance disclosure (at least 48 hours) of demonstratives often useful to video proceedings for later review
<b>2. Technical Advisor</b>  pursuant to inherent powers <i>TechSearch, LLC v. Intel Corp.</i> , 286 F.3d 1360, 1381 (Fed Cir 2002) (approved for use in <i>Markman</i> )	“sounding board” and tutor who aids the court in understanding “jargon and theory” not analogous to law clerk because advisor’s superior technical knowledge can override judge’s prerogative	fair and open procedure for appointment; address allegations of bias, lack of qualifications court must clearly define and limit duties in writing guard against ex parte communications; advisor cannot contribute evidence or conduct independent investigation make explicit (perhaps through a report or record), the nature and content of the advisor’s tutelage concerning technology
<b>3. Special Master</b>  Fed. R. Civ. P. 53	prepares report and recommendations (e.g., proposed claim construction) Court adopts, rejects, or modifies	parties must be given opportunity to object court may receive additional evidence factual and legal issues decided de novo procedural decisions reviewed for abuse of discretion
<b>4. Expert Witness</b>  Fed. R. Evid. 706	instructed by court in writing provides findings to parties and court court or any party may call expert as a witness	court must allow parties to present views may be deposed by any party

## a) Technology Tutorials

Technology tutorials can be especially helpful in educating the court about the underlying technology. Although tutorials will always be shaped by the issues the parties are litigating, the goal of the tutorial should be to give the court neutral, useful background information about the technology.

Cases vary widely on the need for technology tutorials. Some cases need little more than a brief introduction by the lawyers at the *Markman* hearing. Others may benefit from a lengthy, separate presentation with animations and live witnesses. A common practice is to schedule the technology tutorial within two weeks of the *Markman* hearing. It is often best to have the attorneys give the main presentations, with each side's technical expert in attendance for questioning. This approach recognizes that attorneys generally will be the most efficient at tailoring the background technology presentation to the issues the court will confront in *Markman* and throughout the remainder of the case. Having each side's expert in attendance allows the court to ask questions about the science, technical background, and technical terminology. Not all courts share this view, and some discourage attorneys from presenting the tutorial.<sup>330</sup> Several courts have successfully utilized what is referred to as the "hot tub" method, in which experts for each side engage in a dialogue with the court moderating the discussion and probing to determine areas of agreement and disagreement.

The education process involving complex technologies can be improved through the use of video animations, which has the benefit of giving the court a tutorial that can be played at any time, including for newly-arrived law clerks. However, videos are a costly and time-consuming undertaking for the parties and may be less useful than allowing in-court presentations, with the opportunity for live questioning by the court. Some courts videotape in-court tutorials (or use a simple webcam) to achieve the benefits of having a live presentation where the court's questions can be answered, and preserve a copy of the presentation for chambers' use (which captures more than a bare transcript might).

As discussed below, some courts appoint technical experts in patent cases. It is not recommended that the court use a court-appointed expert to deliver the tutorial. Preparing for these tutorials is a lengthy and expensive undertaking, typically with large investments in graphics and multimedia teaching tools. This function cannot be readily delegated to a court-appointed expert under a cost-sharing agreement by the parties, because the parties would never agree on what should be taught, or how the message should be conveyed. Moreover, allowing a court-appointed expert to present the tutorial would inject substantial uncertainty into the proceedings, and would leave the parties to try to present their own views of the technology through cross-examination of the court-appointed expert, which would

---

330. See JUDGE SAUNDRA BROWN ARMSTRONG, JUDGE ARMSTRONG'S PATENT STANDING ORDER EFFECTIVE OCT. 15, 2004, ¶ 6.

detract from the neutral presentation that these tutorials contemplate. It is better to allow each side to present their own view of the technology.

It is important to bear in mind that the Federal Circuit faces challenges comparable to those encountered by district courts in understanding the background technology in patent cases. The appellate court lacks the opportunity to hear from science and technology experts about the background of the technology. Therefore, it will be valuable for the background information to be filed with the court to make it part of the record so that it can be reviewed on appeal. Concise tutorial videos prepared by the parties can be particularly valuable. In addition, transcripts of hearings and PowerPoint slides can assist the Federal Circuit in comprehending the background science and more fully understanding the basis for the district court's claim construction.

b) Court-Appointed Experts

Due to the challenges of understanding the technical issues in particularly complex patent cases, some courts have turned to the appointment of experts. As reflected in Table D, there are three options: (1) technical advisor; (2) special master; and (3) expert witness. These roles vary significantly.

i) Technical Advisor

Given the demands of *Markman* proceedings to construe claims from the perspective of a person of ordinary skill in the art, there can be an appropriate role for technically skilled persons to assist the court, particularly in technologically complex cases.<sup>331</sup> The Federal Circuit expressly approved appointing a technical advisor for *Markman* proceedings in *TechSearch LLC v. Intel Corp.*,<sup>332</sup> although the court emphasized the need to establish “safeguards to prevent the technical advisor from introducing new evidence and to assure that the technical advisor does not influence the district court’s review of the factual disputes.”<sup>333</sup> Applying Ninth Circuit law, the Federal Circuit noted the following guidelines for appointing a technical advisor: (1) “use a ‘fair and open procedure for appointing a neutral technical advisor’ addressing any allegations of bias, partiality or lack of qualifications”; (2) “clearly define and limit the technical advisor’s duties in a writing disclosed to all parties”; (3) “guard against extra-record information”; and (4) “make explicit, perhaps

---

331. See generally John S. Wiley, Jr., *Taming Patent: Six Steps for Surviving Scary Patent Cases*, 50 UCLA L. REV. 1413 (2003) (promoting the use of technical advisors).

332. 286 F.3d 1360 (Fed. Cir. 2002).

333. *Id.* at 1377.

through a report or record, the nature and content of the technical advisor's tutelage concerning the technology."<sup>334</sup> The Federal Circuit cautioned, however, that "district courts should use this inherent authority sparingly and then only in exceptionally technically complicated cases."<sup>335</sup>

The proper role of the advisor is to be a sounding board or tutor who aids the judge's understanding of the technology. This includes explanation of the technical terminology used in the field, the underlying theory or science of the invention, or other technical aspects of the evidence being presented by the parties. The advisor can also assist the judge's analysis by helping think through critical technical problems. In this latter function, case law admonishes that the court must be careful to assure that the decision-making is not delegated to the advisor.<sup>336</sup>

First, one common concern with the appointment of a technical advisor is that the judge's role in applying the legal rules of claim construction may be surrendered to the technical expert, who could then have undue influence over the proceedings. Although in form the relationship between a judge and a technical advisor is much like the interaction between a judge and law clerk, the former relationship differs in that because of a judge's knowledge of law, a clerk cannot usurp the judicial role. Unlike the judge's law clerk, who may have undergraduate and possibly some graduate training in the relevant field and understands his or her role in assisting the judge through legal education and familiarity with the judicial system, a technical advisor will typically be a nationally or internationally known scientist or engineer with limited exposure to legal institutions. They are less likely to appreciate the nature of judicial decision-making and the unique, constitutionally-grounded authority of the court. Perhaps recognizing that parties often do not voluntarily raise these issues to the court, some judges are now including in their standard scheduling order a date for parties to submit agreed-upon names of technical advisors.

Second, a related concern with the use of court-appointed advisors for claim construction is that they distance the judge from some of the most important decisions relating to the case. It is essential for the court to be fully engaged in the interpretation of claim language as these determinations often play a decisive role in the litigation, may require adjustment or further analysis later in the case, and affect the conduct of the trial (e.g., relevance of expert testimony, jury instructions, and what arguments can be made to the

---

334. *Id.* at 1379 (citing *Ass'n of Mexican-Am. Educators v. California*, 231 F.3d 572, 611 (9th Cir. 2000)).

335. *Id.* at 1378.

336. *See id.* at 1377-81.

jury). For this reason, some experienced patent jurists have disavowed use of advisors in claim construction and caution against their use.

Third, there is concern about the transparency of the technical advisor process. The *TechSearch* decision emphasizes the need to guard against extra record information, and makes explicit the nature and content of the technical advisor's tutelage concerning the technology.<sup>337</sup> These principles run counter to using the technical advisor in the same manner as a law clerk, in which the court has informal, off-the-record communication with a member of his or her staff. A technical advisor is not a member of the court's staff. One solution to this concern would be to have all interactions between the judge and the technical advisor in open court with counsel present. Such a procedure, however, could make use of the technical advisor so inconvenient and costly as to render it infeasible. An alternative approach is to have all interactions between the court and the special master transcribed, along with a record made of all correspondence, documents reviewed, and other materials considered by the technical advisor and discussed with the court. A third variation on this alternative, used by one court,<sup>338</sup> is to have transcripts of interaction between the court and the technical advisor sealed and released to the parties only after the trial court proceedings have concluded. This approach has the advantage of enabling the court some flexibility in use of the technical advisor while assuring that the parties will have a full opportunity to review that interaction prior to potential appeal.

#### ii) Special Master

Some courts, pursuant to Federal Rule of Civil Procedure 53, have delegated initial consideration of claim construction to a special master. Such special masters often have general legal training as well as experience with patent law specifically. They might also be familiar with the technical field in question. The special master will typically conduct a claim construction process, with briefing and argument. The special master will then prepare a formal report with recommendations regarding the construction of disputed claim terms. After the parties have had an opportunity to object to that report the court will often conduct a hearing at which the court may receive additional evidence and then adopt, reject, or modify the recommended claim constructions.

---

337. *See id.*

338. This procedure was used by Judge William G. Young in the District of Massachusetts. Interview with William G. Young, Judge, District of Massachusetts, in Boston, Mass. (Mar. 25, 2008).

The use of a special master for the purpose of claim construction alleviates some of the due process concerns inherent in the use of a technical advisor. The special master does not engage in off-the-record communications with the court. On the other hand, the use of a special master runs an even greater risk of distancing the court from the details of claim construction. This limits the court's involvement in some of the most critical aspects of many patent cases and can create problems should claim construction require adjustment later in the case. It may limit the court's ability to gain command over the background science and technology, which could be important later in the case—such as in addressing non-obviousness.

iii) Expert Witness

A third option is the formal appointment of an expert pursuant to Federal Rule of Evidence 706. This procedure is not usually appropriate for the *Markman* process. If there is a role for expert witnesses at the *Markman* hearing, it is likely that the parties will provide their own experts, on their own budgets, and on their own initiative. Because the court will be free to question the experts at the *Markman* hearing, the court should be able to fully explore whatever questions it has on the underlying technology. Of course, courts are free to submit questions to the parties in advance of the hearing to ensure that the experts are fully prepared to respond to the courts' questions. Because the court should be able to resolve its questions through the parties' own witnesses, it is unnecessary to enlist a court-appointed expert to fill this role. These experts can be enormously expensive, and preparing for the all-important confrontation of this expert would drive up costs tremendously. Court-appointed experts have been used in at least one recent jury trial, where there was a serious risk of juror confusion,<sup>339</sup> but the justifications for using a court-appointed testifying expert are lacking in a *Markman* hearing, where the judge should be fully briefed on the issues and is free to question the witness.

D. SUMMARY JUDGMENT AND CLAIM CONSTRUCTION

Effective utilization of the summary judgment process is especially important in patent cases because they present so many complex issues. Summary judgment can play a critical role in narrowing or simplifying the issues in claim construction, thereby promoting settlement or simplifying the trial. On the other hand, the summary judgment process in a patent case can

---

339. *See* *Monolithic Power Sys., Inc. v. O2 Micro Int'l Ltd.*, 558 F.3d 1341, 1348 (Fed. Cir. 2009).

put a significant burden on the court, particularly if the parties file numerous, voluminous motions.

1. *Summary Judgment and Claim Construction*

As with any case, the timing of summary judgment motions can be critical. If the court holds summary judgment proceedings too early in the process, the inability to determine which factual issues are in dispute precludes summary resolution that might be possible at a later stage of the pretrial process. Deferring summary judgment too long in a given case may waste time and resources of the parties and the court on issues that could have been resolved with little discovery.

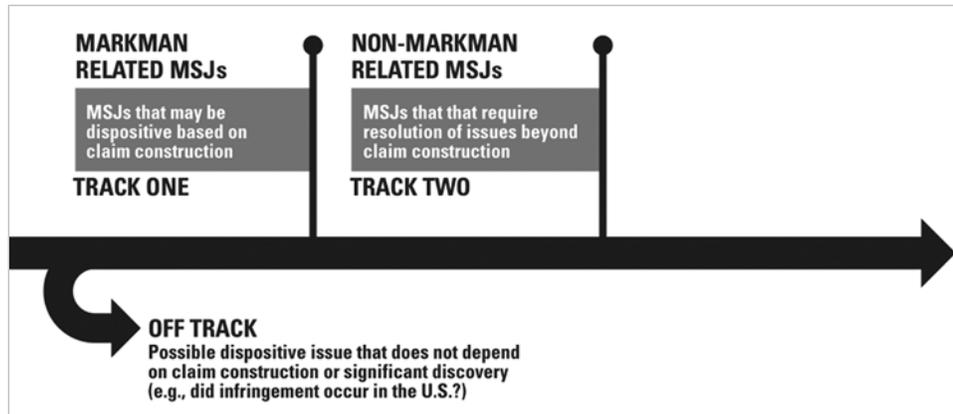
Claim construction plays a central role in scheduling and managing summary judgment motions. Generally, the pretrial issues requiring the largest investment of judicial resources in a patent case are claim construction and summary judgment. Furthermore, most of the weighty issues in a patent case—the technical aspects of infringement and most allegations of invalidity—depend in some way on claim construction. As a result, summary judgment on the main issues in a patent case (infringement and validity) generally cannot be resolved without construing at least some disputed claim terms.

Resolving claim construction issues does not by itself resolve a case unless it fosters settlement. Moreover, not all claim construction disputes are essential to a case—sometimes construing just a single disputed claim term is all that is needed to decide a case-dispositive summary judgment motion. Thus, it can be inefficient to spend the judicial resources needed to resolve all of the claim construction disputes in a case before considering summary judgment motions that could obviate further trial court proceedings.

2. *Recommended Dual-Track Approach to Summary Judgment*

The tension between devoting judicial and party resources to claim construction while at the same time preparing for dispositive motions can be productively resolved by using a dual-track approach to the summary judgment process. On the first (“fast”) track are motions that depend primarily or exclusively on claim construction. On the second track are motions that require resolution of substantial issues beyond claim construction. In rare cases, it may be worthwhile to consider a summary judgment outside either of these tracks—what we refer to as “off-track” summary judgment motions. Figure 1 illustrates the tracks along a time line.

Figure 1: Multi-Track Motion for Summary Judgment (MSJ) Process for Patent Cases



a) “First-Track” Summary Judgment Motions

“First-track” motions are typically non-infringement motions. For example, in *Planet Bingo v. Gametech International*, the claims at issue required “establishing a predetermined winning combination.”<sup>340</sup> The accused bingo machines determined winning combinations after the bingo game began. The parties disputed whether this could be encompassed by the claim term “predetermined.” The district court construed “predetermined” to mean a determination made before the game began. This precluded literal infringement. Based on this construction, and a finding that making a determination *after* the bingo game began could not be equivalent to making the determination *before* the game began, the district court granted summary judgment of non-infringement.<sup>341</sup> The Federal Circuit affirmed.<sup>342</sup> In this case, all that needed to be resolved was the construction of “predetermined” and the issue of what could be “equivalent” to “predetermined”—all other disputes, claim construction or otherwise, were mooted.<sup>343</sup>

In most cases, first-track motions should be resolved as a part of, or in temporal proximity to, the claim construction process. Waiting to address such motions a significant time after claim construction eliminates the potential efficiency of resolving the case based on the construction of a single term or a small set of terms. If the court does not have first-track

340. 472 F.3d 1338, 1340 (Fed. Cir. 2006).

341. *See id.* at 1341.

342. *Id.* at 1345.

343. *See, e.g.,* Schoenhaus v. Genesco, Inc., 440 F.3d 1354, 1356, 1360 (Fed. Cir. 2006) (affirming issuance of “carefully-crafted summary judgment opinion” that “construed two limitations of claim 1 of the patent” in lieu of a claim construction order).

summary judgment issues properly before it during the claim construction process, the court may find itself addressing most or all of the claim construction disputes presented by the parties, only to later find that only one of those disputes actually mattered to the resolution of the case. Thus, while claim construction is often complex in and of itself, hearing a first-track summary judgment motion concurrently with claim construction has the potential to significantly reduce the expenditure of judicial and parties' resources by eliminating the need to consider all the claim construction issues.

Another possibility is to hear first-track motions before claim construction. This is generally not recommended, though it may make sense in some cases if the court is able to determine early in the case that there is a first-track motion with a strong chance of success. The reason this approach is generally not recommended is that it can disrupt and delay the case if the summary judgment motion is denied. Many districts have established local rules for patent cases that set up a structured series of disclosures leading up to claim construction briefing and a hearing.<sup>344</sup> Such procedures are recommended even if they are not required by the district's local rules. It generally does not make sense to postpone or interfere with this process just because one party argues that it has a strong first-track motion.

Hearing first-track summary judgment motions with claim construction strikes a good balance. The case will remain on track even if the motion is denied, or taken under submission at the hearing. An alternative benefit occurs because the summary judgment hearing has been held early enough that the court can avoid unnecessary effort. If the court decides to grant the motion after the hearing, it need only issue an opinion on the claim terms whose construction is necessary to resolve the summary judgment motion. If, on the other hand, the court decides not to grant the motion, then the case can proceed like any other case with the issuance of a claim construction order.

Further, hearing first-track summary judgment motions with claim construction informs the court of important context for understanding the parties' claim construction disputes. Technically, the accused product is not a factor in claim construction.<sup>345</sup> Nonetheless, the Federal Circuit has expressly directed district judges to construe claims with an understanding of the ultimate issues and disputes in a case.<sup>346</sup> Indeed, it is "highly undesirable" to

---

344. *See supra* Section III.A.

345. *Scripps Clinic & Research Found. v. Genentech, Inc.*, 927 F.2d 1565, 1580 (Fed. Cir. 1991) ("[T]he words of the claims are construed independent of the accused product.").

346. *Id.* ("Of course the particular accused product (or process) is kept in mind, for it is

consider claim construction issues “without knowledge of the accused devices,”<sup>347</sup> because these provide the “proper context for an accurate claim construction.”<sup>348</sup> Summary judgment briefing can be an effective vehicle for revealing the motivations underlying claim construction disputes. Of course, information about the issues in the case need not be provided to the court by summary judgment motions. For example, the court can obtain this information through a tutorial, at a case-management conference, or through the claim construction briefing or hearing.

b) “Second-Track” Summary Judgment Motions

“Second-track” summary judgment motions involve substantial issues beyond claim construction—such as jurisdiction, standing, and patentable subject matter—and, therefore, should not normally be considered as part of the claim construction process. Claim construction issues are often interrelated and involve a common set of legal principles and evidence. It makes sense to consider them together. Second-track summary judgment motions involve different sets of legal principles and evidence in addition to underlying claim construction issues. Moreover, most courts have found that it is best to resolve claim construction issues midway through the pretrial process, both to facilitate settlement and so that the parties can prepare for trial knowing the proper claim construction. Unless the second-track motion is straightforward and unaffected by claim construction (for example, a challenge to standing), making the effort to consider a second-track summary judgment motion before issuing a claim construction order diverts judicial resources.

c) Implementing a Dual-Track Approach to Summary Judgment

This dual-track approach to summary judgment in patent cases depends on the ability to distinguish between first-track and second-track motions and to enforce the distinction. It also requires the court to manage the case so that any first-track summary judgment motions are briefed prior to or simultaneous with the claim construction process, and so that Federal Rule of Civil Procedure 56(f) issues do not derail the court’s ability to grant a meritorious first-track motion and dispose of the case early on.

---

efficient to focus on the construction of only the disputed elements or limitations of the claims.”).

347. *Mass. Inst. of Tech. v. Abacus Software*, 462 F.3d 1344, 1350–51 (Fed. Cir. 2006).

348. *Lava Trading, Inc. v. Sonic Trading Mgmt., LLC*, 445 F.3d 1348, 1350 (Fed. Cir. 2006).

The most essential component of this approach is to provide early notice to the parties of the procedure the court intends to follow. The court should explain the first-track motion concept to the parties in a standing order for patent cases, at the initial case-management conference, or both.

There should be a deadline in the case schedule for a summary judgment motion believed to be a first-track motion. To avoid unfairness or problems with Rule 56(f), there should also be a deadline for providing notice to the other party of the basis for any planned first track motion, including, at least, the identity of any witnesses who will submit evidence in support of the motion. These deadlines could be the same, provided that the deadline is far enough in advance of the claim construction hearing to allow the opposing party time to perform reasonably necessary discovery, such as deposing the witnesses who submit declarations in support of the first track motion.

Courts also need to set expectations to avoid having the parties submit multiple first-track summary judgment motions. One option is to limit each party to a single motion. Once the briefing is complete, the court could review it and decide whether to consider it along with claim construction. Another option is to require a party to obtain leave of court before filing a first-track motion. For example, the court could require that a party wishing to file a first-track motion submit a two- or three-page letter brief with the court within two weeks of submitting the Joint Claim Construction Statement required under some courts' PLRs. The letter brief would describe the proposed "first-track" motion and why it should be heard with claim construction. The court could then evaluate how to proceed. This would also afford the opposing party notice of the basis of the motion, to avoid Rule 56(f) problems.

#### d) Recognizing First-Track Summary Judgment Motions

Non-infringement motions based on a small set of claim terms are the most likely to be first track motions. This is because judgment of non-infringement is appropriate if any single claim limitation is not met. Often, the same or similar claim limitations appear in each of the independent claims. If those claim limitations are not met, literal infringement (and quite possibly non-literal infringement) cannot be established and the case, or at least some aspects of it, is resolved. Dependent claims need not be considered because they cannot be infringed if the independent claims are not infringed.

While non-infringement motions are the most common, first-track motions can also include certain invalidity motions, particularly motions for indefiniteness or lack of written description under § 112, or motions

asserting the claims are not patentable subject matter under § 101. Even enablement motions under § 112 can be amenable to early resolution. Whether a claim is patentable subject matter under § 101 is a question of law.<sup>349</sup> Enablement and indefiniteness are also both ultimately legal conclusions for the court, albeit based on underlying facts.<sup>350</sup> While the issue of written description is a question of fact, a patent can nonetheless be held invalid “on its face” for lack of adequate written description.<sup>351</sup> Importantly, enablement, indefiniteness, and written description are issues that often turn on the meaning of a single claim limitation that appears throughout the claims in dispute. For example, modifying the *Planet Bingo* facts slightly, the defendant could have argued that if “predetermined combination” was construed to include winning combinations generated after the bingo game began, the claim was not supported by the patent’s written description.<sup>352</sup> If the patent only described determining winning combinations before the game started, and emphasized the benefits of determining the combinations before the game started, the written description motion could be meritorious and would dispose of the case.

Similarly, motions that argue that claims are not patentable under 35 U.S.C. § 101 are often resolvable without claim construction.<sup>353</sup> Even if some claim construction is required, it may still make sense to consider a § 101 motion as a first-track motion. For example, one court granted summary judgment of invalidity under § 101 using the constructions proposed by the plaintiff, the non-moving party.<sup>354</sup> It is possible, albeit unlikely, for virtually any infringement or validity motion to fall into this category. The key questions are how many disputes the court needs to resolve, and of what type.

Normally, a motion based on anticipation or obviousness will not be a first-track motion because to prove either the moving party must show that every limitation in every claim is present in the prior art. This typically gives rise to a host of disputes, at least some of which are not governed primarily

---

349. *In re Bilski*, 545 F.3d 943, 951 (Fed. Cir. 2008) (addressing the § 101 standard).

350. *Warner-Lambert Co. v. Teva Pharms. USA, Inc.*, 418 F.3d 1326, 1336–37 (Fed. Cir. 2006) (addressing the enablement standard); *Datamize, LLC v. Plumtree Software, Inc.*, 417 F.3d 1342, 1347–48 (Fed. Cir. 2005) (addressing the indefiniteness standard).

351. *Univ. of Rochester v. G.D. Searle & Co.*, 358 F.3d 916, 927 (Fed. Cir. 2004) (describing written description standard and listing cases where a patent was held invalid “on its face” under this standard).

352. *See supra* notes 344–46 and accompanying text.

353. *See, e.g., Fort Props., Inc. v. Am. Master Lease, LLC*, 609 F. Supp. 2d 1052 (C.D. Cal. 2009) (invalidating claims under § 101 without discussion of claim construction).

354. *See CyberSource Corp. v. Retail Decisions, Inc.*, 620 F. Supp. 2d 1068 (N.D. Cal. 2009).

by claim construction issues. Thus, these motions are normally not first-track motions. However, it is possible for a question of anticipation or obviousness to turn on a small number of issues that are manageable early on in the case. For example, if it is beyond reasonable dispute that the patented invention is a specific improvement on a specific prior art device, the validity of the patent may turn on whether the specific improvement is obvious. Now that the Supreme Court has emphasized that obviousness is a legal conclusion for the courts, it is much more likely that fact patterns will arise where even under the patentee's version of the facts it is clear that the claimed inventions are obvious.<sup>355</sup>

### 3. *Summary Judgment Independent from Claim Construction (Off-Track)*

The above discussion focuses on motions that depend on claim construction. In a patent case, this includes most case dispositive issues. However, there are issues that typically do not require the claims to be construed before the motion is decided. For example, a territoriality issue—did the alleged infringement occur “in the United States”?<sup>356</sup>—often will not involve claim construction.

For such motions, the tracked approach does not apply as directly. Still, it remains true that making the effort to consider a summary judgment motion before issuing a claim construction order diverts the resources of both the court and the parties from the goal of focusing and resolving the claim construction issues by the mid-point in a case. Thus, in general, considering an off-track summary judgment motion before claim construction may make sense if the issue is potentially dispositive of the case as a whole or of a significant issue or issues.

In any event, it is important that courts recognize the disconnect that may occur between *Markman* disputes and summary judgment positions. As noted above, *Markman* hearings tend to funnel down to the meaning of isolated terms or phrases in a claim. By contrast, infringement and validity positions often tend to focus on the overall structure or flow of a claim. Resolving a *Markman* dispute as to a particular term may not be sufficient for a party to bring a summary judgment motion relating to the same term. Parties may rightly seek to have a term defined through the *Markman* proceedings, and then wait for trial to press their claims or defenses on the merits to the jury. Absent exceptional circumstances, courts should not penalize parties for deciding not to bring summary judgment motions

---

355. See *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 425 (2007); *PharmaStem Therapeutics, Inc. v. ViaCell, Inc.*, 491 F.3d 1342, 1362 (Fed. Cir. 2007).

356. See 35 U.S.C. § 271(a) (2006).

relating to the terms that are construed in the *Markman* process. Likewise, there may be good reason for parties to forego a *Markman* dispute where the meaning of the words in the claim is not in dispute, but rather the overall claim structure is the focus of a non-infringement or invalidity position.

#### E. CONDUCT OF THE *MARKMAN* HEARING

As courts have experimented with *Markman* hearings, they have had to determine how such proceedings should be characterized and what rules apply.

##### 1. “Evidentiary” Nature of *Markman* Hearings

The “evidentiary” nature of *Markman* hearings is a concept in flux. *Markman* hearings are referred to as “evidentiary hearings.”<sup>357</sup> Nonetheless, the Federal Circuit has ruled that claim construction is strictly a matter of law.<sup>358</sup> This view, however, has increasingly been questioned.<sup>359</sup> A widely-held understanding has been that consideration of fact-intensive “extrinsic” evidence was generally taboo.<sup>360</sup> That line of authority (especially as articulated in *Vitronics Corp. v. Conceptoronic, Inc.*<sup>361</sup>) has been repeatedly discredited and overruled by the Federal Circuit.<sup>362</sup> In recent years the Federal Circuit has allowed consideration of extrinsic evidence, and *Phillips* should put to rest any doubt that extrinsic evidence is proper for consideration.<sup>363</sup> Indeed, several members of the Federal Circuit believe that the time is ripe to reconsider *Cybor*’s rule of de novo review for claim construction.<sup>364</sup> Relying on extrinsic evidence (especially by considering the parties’ expert submissions and making credibility determinations as to their

---

357. See, e.g., *EMI Group N. Am., Inc. v. Intel Corp.*, 157 F.3d 887, 891–92 (Fed. Cir. 1998).

358. *Cybor Corp. v. FAS Techs., Inc.*, 138 F.3d 1448, 1464 (1998) (en banc).

359. See *infra* Section II.A.2.b.

360. “Intrinsic” evidence refers to the patent and its file history, including any reexaminations and reissues. Intrinsic evidence also includes related patents and their prosecution histories. In addition, the Federal Circuit generally treats as intrinsic evidence the prior art that is cited or incorporated by reference in the patent-in-suit and prosecution history. “Extrinsic evidence” refers to all other types of evidence, including inventor testimony, expert testimony, and documentary evidence of how the patentee and alleged infringer have used the claim terms. Dictionaries are considered to be “extrinsic” evidence. *Phillips v. AWH Corp.*, 415 F.3d 1303, 1318 (Fed. Cir. 2005) (en banc); see also *supra* Section II.A.2.

361. 90 F.3d 1576, 1583–84 (Fed. Cir. 1996).

362. See, e.g., *AFG Indus. v. Cardinal IG Co.*, 239 F.3d 1239, 1248 (Fed. Cir. 2001).

363. See *Phillips*, 415 F.3d at 1303, 1324.

364. See *Amgen Inc. v. Hoechst Marion Roussel, Inc.*, 469 F.3d 1039, 1041 (Fed. Cir. 2006) (denying petition for rehearing en banc).

respective merit) may be a way of bolstering the “factual” nature of *Markman* rulings and improving chances of deferential review on appeal.<sup>365</sup> Nonetheless, intrinsic evidence should ordinarily be the primary focus of claim construction determinations.<sup>366</sup>

A frequent and related question is whether, and to what extent, courts should apply the Federal Rules of Evidence in *Markman* proceedings. The dominant and recommended approach is to apply evidentiary rules loosely, in part because *Markman* hearings are not heard by a jury. Furthermore, requiring available witnesses to appear live at a *Markman* hearing and discovery to overcome hearsay and other objections would significantly increase the cost and burden of conducting the hearing. Thus, absent particular concerns about the unreliability of certain forms of proffered evidence, we recommend taking a liberal approach to applying the Federal Rules of Evidence in *Markman* proceedings, such as allowing use of depositions instead of live testimony, declarations (as long as there has been an opportunity for cross-examination), and freer use of documents without a foundational witness as long as there is not a dispute about the authenticity of the document.

## 2. *Safeguards on Extrinsic Evidence*

The court should provide safeguards to ensure that extrinsic evidence is reliable. Allowing depositions of experts prior to a *Markman* hearing reduces this risk and may eliminate the need to call witnesses at the *Markman* hearing. If expert testimony occurs, parties should be permitted to cross-examine any witnesses and allow examination into any sources of documentary evidence that may be proffered. Courts need to scrutinize expert submissions and should actively question the opinions of experts. Typically, experts are highly paid consultants and there is an inherent risk that their opinions will be biased and unreliable. Thus, while it may be extremely probative to hear from persons who are truly experts in the particular field of technology at issue, courts must actively guard against the risk of bias. Cross-examination will usually be a sufficient mechanism to expose bias and unreliability, and conversely, confirm that an expert’s opinions are sound. Courts may choose to apply a *Daubert* standard<sup>367</sup> for qualifying expert witnesses to present

---

365. See *Ortho–McNeil Pharm., Inc. v. Caraco Pharm. Labs., Ltd.*, 476 F.3d 1321, 1328 (Fed. Cir. 2007) (affirming construction based in part on expert testimony that claim term “about 1:5” means “approximately 1:5, encompassing a range of ratios no greater than 1:3.6 to 1:7.1”).

366. See *Phillips*, 415 F.3d at 1319; *supra* Section II.A.2.a.

367. See *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 592–95 (1993) (describing a multi-factor test for determining the admissibility of expert testimony).

expert opinions in a *Markman* hearing. Because *Markman* hearings are not heard by a jury, the need for applying *Daubert* is not as compelling as for a jury trial; however, it would be within the trial court's discretionary powers to exclude any testimony of a witness whose proffered opinions lack the hallmarks of reliability and relevance mandated by *Daubert*.

### 3. *Evidence of the Accused Device*

Another common question is whether, and to what extent, the court should consider the accused device during the *Markman* hearing. In theory, the accused device should have no role in the *Markman* process because the claims should be construed based on the patent language and relevant supporting documentation. Older en banc authority from the Federal Circuit holds that the accused device should not be considered during claim construction.<sup>368</sup> More recently, the Federal Circuit expressly approved consideration of the accused device during claim construction.<sup>369</sup> As stressed by this more recent authority, it is often useful for trial courts to understand the context of the infringement dispute to know what it is that they are deciding when ruling on claim construction. Moreover, knowing the context of the infringement (or validity) dispute gives courts a better sense of whether they even need to construe a term, or if they can simply let the "plain meaning" of a term speak for itself. But the accused device has no relevance to how a person having ordinary skill in the art would interpret claim terms.

### 4. *Evidence of the Prior Art*

Relatedly, courts are free to consider the prior art when ruling on claim construction. Prior art may be directly relevant to claim construction, especially where the patent applicant's dialogue with the Patent Office concerning the prior art may have given rise to a disclaimer.<sup>370</sup> Also, statements in the patent specification about the prior art may be important evidence for construing claim terms.<sup>371</sup> Even apart from prior art recited in

---

368. *See SRI Int'l v. Matsushita Elec. Corp. of Am.*, 775 F.2d 1107, 1118 (Fed. Cir. 1985) ("It is only *after* the claims have been *construed without reference to the accused device* that the claims, as so construed, are applied to the accused device to determine infringement." (emphasis in original)).

369. *Wilson Sporting Goods Co. v. Hillerich & Bradsby Co.*, 442 F.3d 1322, 1326–27 (Fed. Cir. 2006); *Pall Corp. v. Hemasure Inc.*, 181 F.3d 1305, 1308 (Fed. Cir. 1999) ("Although the construction of the claim is independent of the device charged with infringement, it is convenient for the court to concentrate on those aspects of the claim whose relation to the accused device is in dispute.").

370. *See supra* Section II.B.2.e.

371. *See supra* Section II.B.2.d.iii.

the patent and the prosecution history, it is important for trial courts to have the context of other prior art that will form the basis of a validity defense. Those prior art references may play as large a role in shaping the claim construction dispute as does the accused device.

5. *The Need to Focus Markman Proceedings on Claim Construction*

There are limits on the extent to which the court should consider the accused device and prior art during *Markman* proceedings. The *Markman* case seeks to establish distinct roles for the court and for the jury.<sup>372</sup> It is the court's job to perform the legal task of interpreting the scope of the claim terms to the extent possible based upon the patent document from the perspective of a person having ordinary skill in the art. It is the role of the factfinder (typically the jury) to apply these construed terms to the accused device (to determine infringement) and to the prior art (to determine validity). If the court prejudices infringement or validity in its *Markman* ruling, then the court is subject to reversal for having usurped the role of the jury.<sup>373</sup> As we see below, these roles can become blurred in the context of non-technical claim terms and terms of degree.<sup>374</sup> Following the *Markman* ruling, the court is free to entertain summary judgment motions that turn on claim

---

372. See *MacNeill Eng'g Co. v. Trisport, Ltd.*, 126 F. Supp. 2d 51, 54 n.1 (D. Mass. 2001). The court stated:

To open up *Markman* hearings to detailed comparisons between the patented and allegedly infringing device creates the unacceptable risk of conflating claim construction (law teaching) with infringement (fact finding).

Let's face it, when *Markman* hearings become miniature or full blown infringement trials, the actual *language* of the claim diminishes in importance relative to the *context* of the particular dispute, despite the Supreme Court's admonition that it was the judiciary's particular facility for construing *language* that warranted denoting claim construction as a legal, and hence judicial, function.

*Id.* (emphasis in original).

373. See *PPG Indus. v. Guardian Indus. Corp.*, 156 F.3d 1351, 1355 (Fed. Cir. 1998). The court stated:

Claims are often drafted using terminology that is not as precise or specific as it might be . . . . That does not mean, however, that a court, under the rubric of claim construction, may give a claim whatever additional precision or specificity is necessary to facilitate a comparison between the claim and the accused product. Rather, after the court has defined the claim with whatever specificity and precision is warranted by the language of the claim and the evidence bearing on the proper construction, the task of determining whether the construed claim reads on the accused product is for the finder of fact.

*Id.*

374. See *supra* Section II.B.1.e.

construction. We recommend that courts schedule summary judgment motions that can be resolved on the basis of claim construction simultaneously with claim construction hearings. Nonetheless, it will be important for the court to avoid trenching upon the jury's role.

#### 6. *Sequence of Argument*

Courts have broad discretion as to how they conduct *Markman* hearings. Some allocate multiple days to the hearing, while others determine claim construction on the papers.

When there is an oral hearing, it may be appropriate to hear from the lawyers on a term-by-term basis. Particularly when there are many terms at issue, hearing each side's positions for each term can help crystallize the dispute for each term. In other cases, it makes sense for each side to give its complete presentation. Allowing each party to do so may be a better way for appreciating the overall themes of a case. Hybrid approaches may work, as well, with the court hearing from each side on groups of terms.

It is highly recommended that courts allow the parties to make a visual presentation. Multimedia presentations, animations, and other visual aids can be highly instructive tools for teaching the technological concepts and claim construction principles that shape a dispute. They are also especially helpful in illustrating the particular issues in dispute. To the extent possible, the court should endeavor to preserve this record for appellate review.

### F. THE *MARKMAN* RULING

#### 1. *Interrelationship to Jury Instructions*

The *Markman* ruling becomes the basis for the court's jury instructions.<sup>375</sup> Courts should draft their *Markman* rulings with an eye towards making the claim terms understandable to the jury when the time comes for instructions. In this regard, it is highly recommended that courts include a conclusion section at the end of their *Markman* orders setting forth the exact construction that will be used in the jury instructions. Any lack of clarity in this regard invites further disputes in the midst of trial during the drafting of jury instructions.

#### 2. *Basis for Appellate Review*

Comparably important, the court should provide a detailed explanation for the basis for its ruling. Although the Federal Circuit currently reviews

---

375. *IPPV Enters., LLC v. Echostar Commc'ns Corp.*, 106 F. Supp. 2d 595, 601 (D. Del. 2000).

claim construction rulings de novo, it is more likely to defer to the trial court's interpretation when the ruling is detailed and is accompanied by a detailed record. Furthermore, even if the Federal Circuit reaches a different interpretation, a fuller record might provide the basis for an alternative disposition short of remand and a second trial.

The district court should also scrutinize factual stipulations that underlie summary judgment motions following or in combination with claim construction. The parties may enter into such stipulations so as to obtain finality of the district court proceedings and secure appellate review (such as the patentee stipulating to non-infringement after receiving a narrow claim construction). If the stipulation is devoid of context, or overly vague and ambiguous, the Federal Circuit may lack the context it needs to properly resolve the appeal, including making decisions on whether to remand the case. Accordingly, the district court should be vigilant to ensure that any such stipulations provide the necessary facts to justify the finality of the judgment below.

### 3. *The Court May Adopt Its Own Construction*

The court is free to devise its own construction of claim terms rather than adopt a construction proposed by either of the parties. However, the consequence of issuing the court's own construction is that it may upset the foundations of the parties' expert reports and any pending motions before the court. This problem may be particularly acute in late-phase *Markman* hearings where the parties' expert reports may have already been rendered based on the particular wording of the parties' proposed constructions. In such circumstances, departing from the parties' proposed constructions may throw a case off track by requiring new expert reports and re-drafting of case dispositive motions.

### 4. *Tentative Rulings Prior to the Markman Hearing*

Many courts report success with issuing tentative rulings prior to the *Markman* hearing. The ability to follow this approach is naturally constrained by the resources of chambers to issue a tentative ruling in advance of the *Markman* hearing. It may also be infeasible where the invention involves complex science and technology. The court may understandably wish to hear from experts and see demonstrative exhibits before opining, even if only tentatively.

When the court is able to issue a tentative pre-hearing ruling, it has the benefit of informing the parties what issues are most important to the court, in order to most effectively channel the in-court presentations at the *Markman* hearing. This approach allows the court to confirm its

understanding of the record and the governing authorities in a direct dialogue with the attorneys. Issuing a tentative ruling prior to the hearing is a good way for the court to clear up any misperceptions that might otherwise result in reversible error. But given the lack of familiarity that the court may have with the science and technology at issue and the blurred fact and law aspects of claim construction, the court should view its tentative position with less conviction than might otherwise be the case in other areas of the law.

5. *Integrating the Markman Ruling into Trial*

a) Amendments to Infringement and Invalidity Contentions

The court's *Markman* ruling may alter the landscape for a party's infringement or invalidity contentions. Accordingly, for those courts that employ PLRs, or provide for similar provisions in their scheduling orders, it is appropriate to allow limited amendments to a party's infringement or invalidity contentions to account for the *Markman* ruling or other events that may arise during discovery (such as newly discovered prior art or non-public information about the accused devices).<sup>376</sup> Such amendments, however, should only be allowed on a showing of good cause. Freely allowing such amendments would invite litigants to change the playing field late in the case and disrupt the orderly framework that the PLRs are designed to establish.

b) Integrating the *Markman* Ruling into Jury Instructions

The central role of the *Markman* ruling at trial is to provide the basis for the jury instructions. The *Markman* ruling establishes the claim limitations that must be met for the patent to be infringed and for the prior art to invalidate the patent. The *Markman* ruling also establishes the scope of the claims that must be enabled in order for the patent to be valid, and it defines the scope of art that must have been disclosed to the Patent Office during prosecution. Thus, the *Markman* ruling is critical to most of the substantive matters of patent law in the jury instructions. Having a clear, concise *Markman* ruling, which spells out the final constructions for disputed claim terms, is essential to avoiding disputes at trial over the jury instructions. It is useful to place these constructions in a summary conclusion at the end of an opinion so that these constructions can be readily adapted into jury instructions. It is essential that the instructions on claim construction come from the court and that the attorneys not be permitted to re-argue claim

---

376. See, e.g., N.D. CAL. PATENT LOC. R. 3-6.

construction positions inconsistent with the court's instructions, at the risk of a new trial being ordered or of reversal.<sup>377</sup>

Aside from the actual constructions adopted by the court, which are incorporated into jury instructions, the *Markman* opinion should usually not be shown to the jury. The *Markman* ruling will ordinarily include language rejecting the claim construction positions of one of the parties. Conveying that information to the jury would be prejudicial to the party whose position was rejected. Giving the *Markman* ruling to the jury might also interfere in the jury's analysis of the infringement and invalidity arguments, particularly when (as is common) the *Markman* ruling contains a discussion of the accused device and the prior art.

There may be situations in which it is appropriate for portions of the *Markman* ruling to be shown at trial. For example, where the opinion of an expert witness is inconsistent with the claim construction standards ordered by the court, it may be appropriate in some cases to cross-examine the expert on his or her alleged misapplication of the claim construction ruling. In such circumstances, the court should be vigilant in restricting the portions of the ruling that may be shown at trial.

c) Interlocutory Appeal of *Markman* Rulings

Due to Federal Circuit practice, it has become widely accepted that *Markman* rulings cannot be appealed until there has been a final judgment of all claims and counterclaims. In the mid 1990s, various parties attempted to appeal *Markman* rulings prior to obtaining a final judgment on all claims and counterclaims at the district court level. Arguments in favor of such early appeals note that claim construction is a matter of law and that obtaining a definitive claim construction from the Federal Circuit could avoid the costs to all parties of trial on a multitude of issues that hinge on claim construction. Moreover, given the relatively high rate of reversal of claim construction rulings, trial rulings frequently need to be vacated when the claim construction is changed on appeal, even in part. Thus, parties frequently argue that early appeals of claim construction rulings should be allowed to avoid the expense of time and money (including the trial court's own resources) for resolving issues that may likely be disposed of when claim construction is determined on appeal.

---

377. See *CytoLogix Corp. v. Ventana Med. Sys., Inc.*, 424 F.3d 1168, 1172 (Fed. Cir. 2005) (“[B]y agreement the parties also presented expert witnesses who testified before the jury regarding claim construction, and counsel argued conflicting claim constructions to the jury. This was improper, and the district court should have refused to allow such testimony despite the agreement of the parties.”).

Nonetheless, the Federal Circuit disfavors interlocutory review of claim construction rulings. One basis for the Federal Circuit's reluctance to accept early appeals of *Markman* rulings is that claim construction is frequently not finished until trial is complete. It is routine for additional *Markman* issues to arise during trial—either based on new claim construction issues, or the all-too-frequent exercise of “construing the construction,” when the initial claim construction of a court does not squarely resolve the issues presented for trial. Furthermore, because claim construction is tied to so many issues in the case, the Federal Circuit is leery of giving an early ruling on claim construction while unaware of the other issues tied to it. And seeking Federal Circuit review of an interim ruling is disruptive of the underlying litigation because such appeals would be handled on the Federal Circuit's regular appeal schedule, without expedited relief. Another concern is that granting such appeals could discourage settlements<sup>378</sup> and lead to a deluge of appeals that would adversely affect appellate resources.<sup>379</sup>

In 2007, the Federal Circuit granted interlocutory appeal of a *Markman* ruling pursuant to 28 U.S.C. § 1292(b),<sup>380</sup> although the circumstances were somewhat unusual—an earlier case involving the same claims was before the Federal Circuit in an appeal from a denial of a preliminary injunction.<sup>381</sup> The

---

378. See Kimberly A. Moore, *Are District Court Judges Equipped to Resolve Patent Cases?*, 15 HARV. J.L. & TECH. 1, 34–35 (2001). Moore stated:

Although patent appeals only represent about 20% of the Federal Circuit's docket in terms of the number of cases, they are the most complex and time consuming of the cases the court hears. There are approximately 2200 patent cases resolved each year in the district courts. The Federal Circuit judges may fear that if claim construction were appealable on an interlocutory basis, many parties who settle rather than endure expensive and time-consuming litigation would appeal claim construction prior to settlement because a Federal Circuit appeal is relatively inexpensive compared to a district court trial.

*Id.* (footnotes omitted).

379. Chief Judge Michel of the Federal Circuit has expressed this concern to Congress: “I would expect an interlocutory appeal in virtually every patent infringement case as soon as a claim construction order issues.” Letter from Paul R. Michel, Chief Judge, U.S. Court of Appeals, Federal Circuit, to Senators Patrick Leahy and Arlen Specter, U.S. Senators, 2 (June 13, 2007), *cited in* William C. Rooklidge & Mansi H. Shah, *Creation of the Right to Interlocutory Appeal of Patent Claim Construction Rulings and Mandatory Stay Pending Appeal* 3 n.11, *available at* [http://www.patentmatter.com/issue/pdfs/Interlocutory\\_Review\\_Paper.pdf](http://www.patentmatter.com/issue/pdfs/Interlocutory_Review_Paper.pdf).

380. This provision authorizes a district court to certify and order for interlocutory reviews that turn on a controlling question of law as to which there is substantial ground for difference of opinion and an immediate appeal from the order may materially advance the ultimate termination of the litigation.

381. See *Regents of Univ. of Cal. v. Dako N. Am., Inc.*, 477 F.3d 1335, 1336 (Fed. Cir. 2007).

court noted its general disfavor of such interlocutory appeals, but explained that there was already a co-pending appeal of a denial of a preliminary injunction and thus that it made sense to hear the interlocutory appeal in connection with the co-pending appeal.<sup>382</sup>

Partially in response to proposed legislation liberalizing the standard for interlocutory review of *Markman* determinations, Chief Judge Michel has publicly invited litigators to seek interlocutory appeals on claim construction.<sup>383</sup> While this does not appear to be signaling an invitation to review every (or even many) *Markman* rulings on an interlocutory basis, this case-management option may be appropriate in limited circumstances. Procedurally, litigants have had the most success obtaining early appellate review when the *Markman* ruling renders the claims non-infringed. The parties may at that point stipulate to non-infringement, and ask the trial court to enter final judgment as to non-infringement under Federal Rule of Civil Procedure 54(b). On occasion, the Federal Circuit has granted review of partial judgments entered under Rule 54(b).<sup>384</sup> However, because the issues of invalidity and unenforceability generally remain pending below, the Federal Circuit commonly will deny such review.<sup>385</sup> At least one judge has remarked that allowing such piecemeal review of issues “portends chaos in process.”<sup>386</sup> Litigants seeking to invoke such review may maximize their chances by fully describing the basis for non-infringement so as to provide meaningful review of that ruling on appeal.<sup>387</sup> Furthermore, parties arranging for dismissal of the remaining claims would also facilitate review (although such dismissal may be with prejudice).<sup>388</sup>

#### IV. CONCLUSIONS

More than any other facet of patent cases, claim construction distinguishes patent litigation from other forms of civil actions. The substantive law of claim construction can be analogized to interpretation of other texts, but various nuanced features—the perspective of the person of

---

382. *Id.* at 1336–37.

383. Tony Dutra, *Chief Judge Issues Call to Action to Bring Cases for En Banc Federal Circuit Review*, 76 PAT., TRADEMARK, & COPYRIGHT J. (BNA) 755 (2008) (“Litigators . . . should be seeking interlocutory appeals on claim construction. For 15 years, litigators stopped the practice, but he noted that ‘we got one this year and granted it.’”).

384. *See Lava Trading, Inc. v. Sonic Trading Mgmt., LLC*, 445 F.3d 1348, 1350 (Fed. Cir. 2006).

385. *See Linear Tech. Corp. v. Impala Linear Corp.*, 31 F. App’x. 700 (Fed. Cir. 2002).

386. *Lava Trading*, 445 F.3d at 1355 (Mayer, J., dissenting).

387. *See id.* at 1350.

388. *See Nystrom v. TREX Co.*, 339 F.3d 1347, 1351 (Fed. Cir. 2003).

ordinary skills, the technical nature of the subject matter, distinctions between lay and technical terms, the importance of prosecution history, the interplay of multiple claims, and the need to safeguard the jury's role in determining infringement—distinguish interpretation of patent claims from contractual and statutory interpretation. While the Federal Circuit's en banc *Phillips* decision clarifies many of the principles underlying claim construction, neither that decision nor other sources provide a cohesive step-by-step process or overarching framework to guide lower courts in rendering decisions. Through synthesis of the vast jurisprudence and working with a broad range of jurists and practitioners, this Article provides a pragmatic approach to applying the substantive principles.

In the decade and a half since the Supreme Court's *Markman* decision, district courts have come a long way in developing effective strategies for managing claim construction and patent case management. Most significantly, the Northern District of California pioneered the development of specialized PLRs to promote orderly resolution of claim construction. Such rules, which have now been adopted in more than ten districts (including many of the most patent-intensive jurisdictions), provide for joint, sequenced, staged, and timely disclosure of claim construction contentions. Beyond PLRs, a growing number of district courts have developed effective means for limiting the number of claim terms that must be construed, integrating summary judgment with claim construction, coming up to speed on the science and technology necessary to interpret claims, conducting claim construction hearings, and integrating claim construction rulings into patent trials.

**APPENDIX:  
NARROWING OR BROADENING "ORDINARY MEANING"**

Doctrine	Citation
<b>I. Narrowing Construction</b>	
<b>A. Description of Invention</b>	
Characterization of "the present invention."	Netcraft Corp. v. eBay, Inc., 549 F.3d 1394, 1398 (Fed. Cir. 2008) ("We agree with Netcraft that use of the phrase 'the present invention' does not 'automatically' limit the meaning of claim terms in all circumstances, and that such language must be read in the context of the entire specification and prosecution history. For the reasons below, however, we agree with the district court that the common specification's repeated use of the phrase 'the present invention' describes the invention as a whole, and, as will be discussed further below, that the prosecution history does not warrant a contrary result."); Verizon Servs. Corp. v. Vonage Holding Corp., 503 F.3d 1295, 1308 (Fed. Cir. 2007) ("In the course of describing the 'present invention,' the specification then

	states that “[t]he gateway compresses and decompresses voice frequency communication signals and sends and receives the compressed signals in packet form via the network.’ When a patent thus describes the features of the ‘present invention’ as a whole, this description limits the scope of the invention.”); <i>Honeywell Int’l, Inc. v. ITT Indus., Inc.</i> , 452 F.3d 1312, 1318 (Fed. Cir. 2006) (holding that the “fuel injection system component” was limited to a fuel filter because all the disclosed embodiments disclosed only fuel filters and the specification repeatedly described the fuel filter as “this invention” and “the present invention”).
Distinctions over prior art.	<i>SafeTCare Mfg., Inc. v. Tele-Made, Inc.</i> , 497 F.3d 1262, 1270 (Fed. Cir. 2007) (holding that patentee’s statements throughout specification revealed an intentional disclaimer or disavowal of coverage (“In this case, the written description repeatedly emphasizes that the motor of the patented invention applies a pushing force, not a pulling force, against the lift dog. The inventor makes clear that this attribute of the invention is important in distinguishing the invention over the prior art. Thus we are persuaded by the language used by the patentee that the invention disclaims motors that use pulling forces against the lift dogs.”)); <i>SciMed Life Sys., Inc. v. Advanced Cardiovascular Sys., Inc.</i> , 242 F.3d 1337, 1343–44 (Fed. Cir. 2001) (limiting claim to unitary lumen where the specification distinguished the prior art in part on the ground of the use of dual lumen configurations).
Consistent usage of claim terms in patent and prosecution history.	<i>Irdeto Access, Inc. v. EchoStar Satellite Corp.</i> , 383 F.3d 1295, 1303 (Fed. Cir. 2004) (“[W]hile the specification does not contain any statements of explicit disavowal or words of manifest exclusion, it repeatedly, consistently, and exclusively uses ‘group’ to denote fewer than all subscribers, manifesting the patentee’s clear intent to so limit the term.”); <i>Int’l Rectifier Corp. v. IXYS Corp.</i> , 361 F.3d 1363, 1371–72 (Fed. Cir. 2004) (“The correct construction of the term ‘polygonal,’ consistent with the written description, is simply ‘a closed plane figure bounded by straight lines.’ The patentee, being fully aware of the effects of the doping process, could have claimed the regions more broadly but chose to use the word “polygonal” without modification or qualification. The district court was not free to attribute new meaning to the term or to excuse the patentee from the consequences of its own word choice.”); <i>Bell Atl. Network Servs., Inc. v. Covad Commc’ns Group</i> , 262 F.3d 1258, 1273 (Fed. Cir. 2001) (“[T]he patentees defined the term ‘mode’ by implication, through the term’s consistent use throughout the ’786 patent specification. Given this definition, the three modes described in the Detailed Description of the Preferred Embodiments describe the three possible modes of the invention, and the claims are not entitled to any broader scope.”).
<b>B. Prosecution Disclaimer</b>	
Surrendering claim scope during prosecution	<i>MBO Labs., Inc. v. Becton, Dickinson &amp; Co.</i> , 474 F.3d 1323, 1330 (Fed. Cir. 2007) (“Prosecution arguments like this one which draw distinctions between the patented invention and the prior art are useful for determining whether the patentee intended to surrender

narrows claim construction.	territory, since they indicate in the inventor’s own words what the invention is not.”(citation omitted)); <i>Bass Pro Trademarks, LLC v. Cabela’s, Inc.</i> , 485 F.3d 1364, 1369 (Fed. Cir. 2007) (holding that where a patentee procures a patent based upon the unique combination of elements stressed in the prosecution history, such combination is a “material limitation to the claim”); <i>Hakim v. Cannon Avent Group, PLC</i> , 479 F.3d 1313, 1318 (Fed. Cir. 2007) (holding that a patentee may not recapture through a continuation application what was surrendered during prosecution of the parent application); <i>Atofina v. Great Lakes Chem. Corp.</i> , 441 F.3d 991, 998 (Fed. Cir. 2006) (holding that while “it frequently happens that patentees surrender more through amendment than may have been absolutely necessary to avoid particular prior art,” the patentee is still limited “to the scope of what they ultimately claim,” and cannot “assert that claims should be interpreted as if they had surrendered only what they had to”).
“Clear and unmistakable disavowal” required for prosecution disclaimer.	<i>Abbott Labs. v. Sandoz, Inc.</i> , 566 F.3d 1282, 1290 (Fed. Cir. 2009) (“[T]he prosecution history of the ’507 patent shows a clear and intentional disavowal of claim scope beyond Crystal A.”); <i>Gillespie v. Dywidag Sys. Int’l, USA</i> , 501 F.3d 1285, 1291 (Fed. Cir. 2007). (“Although [plaintiff] argues that this distinction was not material to the grant of his patent . . . he nonetheless argued this distinction from the [prior art] . . . . The patentee is held to what he declares during the prosecution of his patent.”); <i>Verizon Servs. Corp. v. Vonage Holding Corp.</i> , 503 F.3d 1295, 1306 (Fed. Cir. 2007). (“We have held that a statement made by the patentee during [the] prosecution history of a patent in the same family as the patent-in-suit can operate as a disclaimer. To operate as a disclaimer, the statement in the prosecution history must be clear and unambiguous, and constitute a clear disavowal of scope.” (internal citations omitted)).
<b>C. Special Cases</b>	
Inventors may expressly define terms differently than ordinary meaning.	<i>Chamberlain Group, Inc. v. Lear Corp.</i> , 516 F.3d 1331, 1337 (Fed. Cir. 2008) (“[T]he ’544 patent specification gives particular limiting meanings to the language in the claims . . . . While the district court’s construction may represent an ordinary or customary reading of “binary code,” the ’544 patent restricts “binary code” to a narrower meaning.”); <i>Sinorgchem Co., Shandong v. Int’l Trade Comm’n</i> , 511 F.3d 1132, 1136 (Fed. Cir. 2007) (“The specification states, ‘A ‘controlled amount’ of protic material is an amount up to that which inhibits the reaction of aniline with nitrobenzene . . . .’ The term ‘controlled amount’ is set off by quotation marks—often a strong indication that what follows is a definition. Moreover, the word ‘is,’ again a term used here in the specification, may ‘signify that a patentee is serving as its own lexicographer.’ ” (citation omitted)); <i>Honeywell Int’l, Inc. v. Universal Avionics Sys. Corp.</i> , 493 F.3d 1358, 1361 (Fed. Cir. 2007) (“The specification and prosecution history make clear, however, that the patentees used the term ‘heading’ in a manner different from its ordinary meaning. When a patentee defines a claim term, the patentee’s definition governs, even if it is contrary to the conventional meaning of the

	term.”); <i>Abraxis Bioscience, Inc. v. Mayne Pharma (USA) Inc.</i> , 467 F.3d 1370, 1376 (Fed. Cir. 2006) (holding that the patentee acted as their own lexicographers by defining “edentate” in the specification).
Specification may disclaim coverage to embodiments.	<i>SciMed Life Sys., Inc. v. Advanced Cardiovascular Sys., Inc.</i> , 242 F.3d 1337, 1343–44 (Fed. Cir. 2001) (limiting claim to unitary lumen where the specification stated that the unitary lumen configuration is the “basic . . . structure for all embodiments of the present invention contemplated and disclosed herein”).
Ambiguity in claim term may permit limiting scope to preferred embodiment.	<i>E-Pass Techs., Inc. v. 3COM Corp.</i> , 343 F.3d 1364, 1370 n.4 (Fed. Cir. 2003) (“Where claim language is ambiguous, the purpose of the invention described in the specification may, of course, sometimes be useful in resolving the ambiguity.”); <i>Comark Commc’ns, Inc. v. Harris Corp.</i> , 156 F.3d 1182, 1187 (Fed. Cir. 1998) (noting that interpreting claim language in light of the specification is proper when a term is “so amorphous that one of skill in the art can only reconcile the claim language with the inventor’s disclosure by recourse to the specification”).
Means-plus-function terms are limited to structures in specification and equivalents.	<i>Welker Bearing Co. v. PHD, Inc.</i> , 550 F.3d 1090, 1095–97 (Fed. Cir. 2008) (holding that the term “mechanism for moving said finger” was a limitation subject to means-plus-function treatment); <i>TriMed, Inc. v. Stryker Corp.</i> , 514 F.3d 1256, 1259 (Fed. Cir. 2008) (noting that a patentee’s use of the word “means” in a claim limitation creates a presumption that 35 U.S.C. § 112 para. 6 applies); <i>Applied Med. Res. Corp. v. U.S. Surgical Corp.</i> , 448 F.3d 1324, 1333 (Fed. Cir. 2006) (noting that literal infringement of a claim limitation in means-plus-function format “requires that the relevant structure in the accused device perform the identical function recited in the claim and be identical or equivalent to the corresponding structure in the specification”); <i>Mass. Inst. of Tech. v. Abacus Software</i> , 462 F.3d 1344, 1354 (Fed. Cir. 2006) (noting that “[t]he generic terms ‘mechanism,’ ‘means,’ ‘element,’ and ‘device,’ typically do not connote sufficiently definite structure” to avoid means-plus-function treatment).
<b>II. Broadening Construction</b>	
<b>A. Claim Differentiation</b>	
“Pure” claim differentiation creates a presumption that independent claims are broader than dependent claims.	<i>Praxair, Inc. v. ATMI, Inc.</i> , 543 F.3d 1306, 1326 (Fed. Cir. 2008) (“While no mention of uniformity appears in independent claim 1, the uniformity criterion defined in the specification—‘variation in diameter of different capillary passages does not exceed 15%’—is set forth in dependent claim 4. It therefore appears that the uniformity requirement, as set forth in the specification, was intended to be added by dependent claim 4, and was not already present in independent claim 1 or the invention overall.”); <i>Voda v. Cordis Corp.</i> , 536 F.3d 1311, 1320 (Fed. Cir. 2008) (“Cordis concedes that ‘claim 1 does not expressly recite a ‘straight portion.’” By contrast, claims 4 and 5 of the ’213 patent specifically require that the contact portion of the catheter be a ‘substantially straight leg’ in its rest state. Therefore, the fact that claim 1—and dependent claims 2 and 3—does not expressly recite a ‘straight’ or

	<p>‘substantially straight’ portion strongly implies that claims 1 through 3 do not require the contact portion of the catheter to be straight in its rest state.”); <i>Phillips v. AWH Corp.</i>, 415 F.3d 1303, 1309–10 (Fed. Cir. 2005) (holding that “baffles” included metal supports oriented at ninety degrees to the wall because a dependent claim in the patent recited baffles “projecting inwardly from the outer shell at angles tending to deflect projectiles that penetrate the outer shell”); <i>Liebel–Flarsheim Co. v. Medrad, Inc.</i>, 358 F.3d 898, 910 (Fed. Cir. 2004) (“Although that presumption [of claim differentiation] can be overcome if the circumstances suggest a different explanation, or if the evidence favoring a different claim construction is strong, the presumption is un rebutted in this case, as Medrad has offered no alternative explanation for why the ‘pressure jacket’ limitation is found in the dependent claims but not in the corresponding independent claims. In such a setting, where the limitation that is sought to be ‘read into’ an independent claim already appears in a dependent claim, the doctrine of claim differentiation is at its strongest.”).</p>
<p>Presumption may be rebutted based on specification or prosecution history, or where § 112 para. 6 involved.</p>	<p><i>Regents of the Univ. of Cal. v. Dakocytomation Cal., Inc.</i>, 517 F.3d 1364, 1375 (Fed. Cir. 2008) (“[T]he prosecution history overcomes the presumption [of claim differentiation]; the correct construction of ‘heterogeneous mixture’ is one that excludes repetitive sequences, notwithstanding the presence of certain dependent claims that do not exclude them.”); <i>Sinorgchem Co., Shandong v. Int’l Trade Comm’n</i>, 511 F.3d 1132, 1140 (Fed. Cir. 2007) (“Because claim 41 refers merely to a subset of the solvent systems described in claim 30, and is significantly narrower in scope, the claims are not rendered identical and present no claim differentiation problem.”); <i>SRAM Corp. v. AD-II Eng’g, Inc.</i>, 465 F.3d 1351, 1357–58 (Fed. Cir. 2006) (restricting independent claim to use of “precision index downshifting” even though this term was present in dependent claim, when additional differences existed between the independent and dependent claim); <i>Seachange Int’l, Inc. v. C-COR Inc.</i>, 413 F.3d 1361, 1369 (Fed. Cir. 2005) (noting that the presumption created by the doctrine of claim differentiation is “not a hard and fast rule and will be overcome by a contrary construction dictated by the written description or prosecution history”).</p>
<b>B. Preferred Embodiment Generally Not Limiting</b>	
<p>Preferred embodiment generally not limiting absent a clear intention to limit scope.</p>	<p><i>Epistar Corp. v. U.S. Int’l Trade Comm’n</i>, 566 F.3d 1321, 1337 (Fed. Cir. 2009) (refusing to limit “substrate” to a preferred embodiment that describes the thicker layer as a “substrate” since the specification explains that the thickness identified for the substrate is merely “exemplary”); <i>Linear Tech. Corp. v. Int’l Trade Comm’n</i>, 566 F.3d 1049, 1057–58 (Fed. Cir. 2009) (refusing to limit claim to cover the only disclosed embodiments or examples in the specification even when only one embodiment is disclosed); <i>Howmedica Osteonics Corp. v. Wright Med. Tech., Inc.</i>, 540 F.3d 1337, 1345–46 (Fed. Cir. 2008) (refusing to limit otherwise broad claim language to a single disclosed embodiment where there was nothing in the specification to indicate the inventor meant to limit</p>

	<p>the claim language); <i>Decisioning.com, Inc. v. Federated Dep't Stores, Inc.</i>, 527 F.3d 1300, 1314 (Fed. Cir. 2008) (“[The] description of a preferred embodiment, in the absence of a clear intention to limit claim scope, is an insufficient basis on which to narrow the claims.”); <i>Acumed LLC v. Stryker Corp.</i>, 483 F.3d 800, 805 (Fed. Cir. 2007) (“[The defendant’s] argument is essentially an assertion that since the patent says broaching is desirable, the term ‘curved’ must be construed to cover only embodiments whose curvature allows them to be inserted into a broached hole, excluding ‘angled bends or small radius curves.’ That assertion is flawed: it is an attempt to import a feature from a preferred embodiment into the claims.”).</p>
--	--



# COMMUNICATIONS PRIVACY IN THE MILITARY

*Justin Holbrook*<sup>†</sup>

## TABLE OF CONTENTS

I.	<b>INTRODUCTION</b> .....	832
II.	<b>FIRST PRINCIPLES OF PRIVACY IN THE MILITARY WORKPLACE</b> .....	835
	A. THE TOUCHSTONE OF FOURTH AMENDMENT ANALYSIS .....	835
	B. THE MILITARY AND CONSTITUTIONAL PROTECTIONS .....	838
	C. THE FOURTH AMENDMENT AS APPLIED TO THE MILITARY .....	841
	D. PRIVACY IN THE PUBLIC WORKPLACE .....	844
	E. PRIVACY IN THE MILITARY WORKPLACE .....	847
	F. FIRST PRINCIPLES OF MILITARY WORKPLACE PRIVACY .....	850
III.	<b>THE FOURTH AMENDMENT AND MILITARY E-MAIL</b> .....	851
	A. THE FOURTH AMENDMENT AND THE FEDERAL STATUTORY SCHEME FOR ELECTRONIC COMMUNICATIONS PRIVACY .....	852
	1. <i>Katz v. United States</i> .....	852
	2. <i>The Wiretap Act</i> .....	853
	3. <i>The Electronic Communications Privacy Act and Stored Communications Act</i> .....	855
	4. <i>Fourth Amendment Challenges to the SCA</i> .....	857
	B. E-MAIL PRIVACY IN THE PUBLIC WORKPLACE .....	860
	C. E-MAIL PRIVACY IN THE MILITARY WORKPLACE .....	864
	1. <i>United States v. Maxwell—The Fourth Amendment and Commercial E-Mail</i> .....	867
	2. <i>United States v. Monroe—The Fourth Amendment and Government E-Mail Monitoring</i> .....	870
	3. <i>United States v. Long—The Fourth Amendment and Government E-Mail Searches</i> .....	874
	4. <i>United States v. Larson—The Fourth Amendment and Government Computer Searches</i> .....	878
IV.	<b>THE LANDSCAPE AFTER <i>LONG</i> AND <i>LARSON</i></b> .....	885

---

© 2010 Justin Holbrook.

<sup>†</sup> Associate Professor of Law and Director of Veterans Law Clinic, Widener University School of Law. J.D. Harvard Law School (2001); B.A.L.S Georgetown University (1998). The author wishes to thank Grant Wahlquist and Drew Brown for their invaluable assistance with this article.

A.	THE DOD’S NEW LOGON BANNER AND CONSENT AGREEMENT .....	885
B.	GENERAL WARRANTS AND FOURTH AMENDMENT PROTECTIONS.....	893
	1. <i>General Warrants and the Particularity Requirement</i> .....	893
	2. <i>Neutral and Detached Interposition</i> .....	898
	3. <i>Voluntariness of Consent</i> .....	900
C.	THE NEED FOR A NORMATIVE APPROACH .....	903
V.	<b>CONCLUSION</b> .....	907

## I. INTRODUCTION

The warrant is to seize all the plaintiff’s books and papers without exception, and carry them before Lord Halifax; what? Has a Secretary of State a right to see all a man’s private letters of correspondence, family concerns, trade and business? This would be monstrous indeed; and if it were lawful, no man could endure to live in this country.<sup>1</sup>

It is axiomatic that subjective expectations of privacy must be objectively reasonable to warrant constitutional protections under the Fourth Amendment.<sup>2</sup> When it comes to computers in the workplace, proponents of robust privacy expectations face difficult battles on both the subjective and objective fronts. As a question of fact, at least three (and probably many more) factors diminish subjective privacy expectations in workplace electronic communications<sup>3</sup>: the public nature of work environments, the technical oversight of network administrators, and the ubiquity of employee notices in handbooks, user agreements, and logon banners. As a question of law, the operational realities of the workplace and the need for employer oversight further encumber objectively reasonable privacy expectations.<sup>4</sup>

In the context of the federal workplace, such subjectively held and objectively reasonable privacy expectations in electronic communications

1. *Entick v. Carrington*, 95 Eng. Rep. 807 (K.B. 1765) (recounting plaintiff’s response to defendants’ arguments).

2. *See Minnesota v. Olson*, 495 U.S. 91, 95 (1990); *Katz v. United States*, 389 U.S. 347, 361 (1967) (Harlan, J., concurring).

3. For purposes of this Article, the term “communications” specifically refers to electronic communications, including e-mail, chat, and instant messaging.

4. *See United States v. Maxwell*, 45 M.J. 406, 417 (C.A.A.F. 1996) (quoting *Tri-State Steel Constr., Inc. v. Occupational Safety & Health Review*, 26 F.3d 173, 176 (D.C. Cir. 1994)) (observing subjective expectations of privacy are questions of fact reviewed under clearly erroneous standard while objective expectations of privacy are questions of law subject to de novo review).

appear even more incredible, especially if that federal workplace happens to be within the U.S. military. That constitutional protections are colored by the nature of military service is well-settled law.<sup>5</sup> What service member, therefore, could honestly and reasonably believe that a government computer issued for official use by the military provides even a sliver of privacy expectation?

As it turns out, the issue is one of deceptive simplicity. Like courts elsewhere, military courts have struggled to find an articulable standard with which to analyze Fourth Amendment electronic privacy issues. This difficulty is partly due to the factual differences between electronic privacy cases which, when painted with the broadest of brushes, appear factually similar, but whose technological distinctions justify different subjective and objective privacy expectations. It is partly due to the ferment engendered by courts applying Fourth Amendment jurisprudence to evolving technologies. And it is also partly due to the unique legal standards applied to the military by virtue of its history, customs, and statutory authority to both supervise and self-police its members.

Weaving these concerns into a coherent analytical model is no easy task. At issue is a straightforward question cloaked in complexity: Do service members have a legitimate expectation of privacy in their workplace electronic communications? Following from this question are others. If privacy expectations are justified, how much privacy? Would the acknowledgment of some measure of privacy degrade the military's operational effectiveness? What about its ability to self-police remotely deployed members who commit criminal misconduct using government information systems?

In exploring these questions, I draw on a line of cases issued by the Court of Appeals for the Armed Forces (CAAF)<sup>6</sup> grappling with service members' communications privacy expectations, including *United States v.*

---

5. For a discussion of the applicability of the Bill of Rights to military service members, see *infra* Section II.B.

6. The United States Court of Appeals for the Armed Forces (CAAF), an Article I court created by Congress, functions as "the supreme court of the military judicial system." *United States v. Rorie*, 58 M.J. 399, 403 (C.A.A.F. 2003) (quoting *McPhail v. United States*, 1 M.J. 457, 462 (C.M.A. 1976)). The court is composed of five civilian judges appointed for 15-year terms by the President with the Senate's advice and consent. *Id.* Prior to its current designation, the CAAF was known as the Court of Military Appeals (1950–1968) and the United States Court of Military Appeals (1968–1994). THE UNITED STATES COURT OF APPEALS FOR THE ARMED FORCES (2006), available at <http://www.armfor.uscourts.gov/CAAFBooklet2006.pdf>; see also Karen A. Ruzic, Note and Comment, *Military Justice and the Supreme Court's Outdated Standard of Deference: Weiss v. United States*, 70 CHI.-KENT L. REV. 265, 276–77 (1994) (discussing congressional motivation behind changing the CAAF's name).

*Long*<sup>7</sup> and *United States v. Larson*,<sup>8</sup> two cases that have caused widespread concern throughout the Department of Defense (DoD).<sup>9</sup> I also reflect on the possibility that the nature of military service itself warrants both subjective and objective expectations of privacy in workplace communications. On the one hand, the unique realities of military service justify diminished privacy expectations in everything from garrison barracks to government e-mail accounts.<sup>10</sup> On the other hand, the dislocation of military members from friends and family by virtue of their military service arguably creates an inescapable user reliance on military resources to communicate with the outside world. This reliance forces the military to acknowledge that government information systems may, in fact, be used for more than official use, which could create both a subjective and objective expectation of privacy in any personal use.<sup>11</sup>

This Article's argument proceeds in three steps. First, the Article considers Fourth Amendment jurisprudence and how it applies to military members in the workplace. Several principles are distilled from this jurisprudence to mark the contours of communications privacy in the

---

7. 64 M.J. 57 (C.A.A.F. 2006).

8. 66 M.J. 212 (C.A.A.F. 2008).

9. See, e.g., Lawrence A. Edell, *A Reasonable Expectation of Privacy: Is a Government E-mail Account the Equivalent of a Wall Locker in a Barracks Room?*, 2008 ARMY LAW. 1, 2 (2008) (observing that the CAAF's decision in *Long* created DoD concern regarding "the ability to monitor its computer networks"); Jamie L. Mendelson, *Government Computers and Email—Get the Search Authorization!*, 3 JAJG PERSP. 9 (2008) (discussing post-*Larson* search authorization for service members' workplace e-mails and computers); U.S. DEP'T OF AIR FORCE, INSTRUCTION 51-201, ADMINISTRATION OF MILITARY JUSTICE § 13.13.2.2 (2010) [hereinafter AFI 51-201] (noting the unsettled nature of electronic privacy law after *Long* and *Larson* and encouraging officials to "consider obtaining search authorization for cases involving alleged criminal activity").

10. For a discussion of the applicability of the service members' privacy expectations under the Fourth Amendment, see *infra* Section II.E.

11. Because of the nature of military service, and in accordance with the Joint Ethics Regulation (JER), military service regulations allow limited personal use of official information systems. See U.S. DEP'T OF DEFENSE, DIR. 5500.7R, JOINT ETHICS REGULATION § 2-301 (Nov. 29, 2007); U.S. DEP'T OF AIR FORCE, INSTRUCTION 33-129, WEB MANAGEMENT AND INTERNET USE § 2-1 (2005) ("Appropriate officials may authorize personal uses . . ."); U.S. DEP'T OF ARMY, ARMY REG. 25-1, ARMY KNOWLEDGE MANAGEMENT AND INFORMATION TECHNOLOGY para. 6-1.d(5) (2008) ("Official use includes health, morale, and welfare (HMW) communications by military members and DoD employees who are deployed in remote or isolated locations for extended periods of time on official DoD business."); Joshua E. Kastenbergh, *Changing the Paradigm of Internet Access from Government Information Systems: A Solution to the Need for the DoD to Take Time-Sensitive Action on the NIPRNET*, 64 A.F. L. REV. 175, 184–88 (2009) (discussing JER and noting it authorizes commanders "to permit DoD personnel at a normal workplace to conduct brief internet searches beyond matters involving official business or family communications").

military. Second, the Article considers Fourth Amendment protections in the context of electronic communications, working through the seminal Supreme Court decisions, the congressional enactment of the Stored Communications Act, and the CAAF's communications privacy cases. Finally, the Article reviews the DoD's response to the CAAF's communications privacy cases and explore how a normative inquiry could assist the CAAF in situating service members' communications privacy rights in the context of traditional Fourth Amendment jurisprudence. The Article concludes by suggesting that military courts adopt an analytical framework that explicitly distinguishes between work-related and law enforcement searches in determining the degree of communications privacy properly afforded service members.

## II. FIRST PRINCIPLES OF PRIVACY IN THE MILITARY WORKPLACE

This Part reviews Fourth Amendment jurisprudence and discusses how that jurisprudence has been applied in the public workplace generally and the military workplace specifically. This Part addresses the relevant portions of the Military Rules of Evidence (MRE), which distinguish between searches for work-related and law enforcement purposes, and analyzes federal workplace case law, which similarly distinguishes between work-related and law enforcement searches. Distilled from this discussion are several principles of military workplace privacy which, when considered in light of electronic communications, provide critical guidance in determining what level of electronic communications privacy service members should receive.

### A. THE TOUCHSTONE OF FOURTH AMENDMENT ANALYSIS

The Fourth Amendment protects the "right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures . . . ."<sup>12</sup> Warrants must issue upon "probable cause, supported by Oath or affirmation, and particularly describing the place to be searched and the persons or things to be seized."<sup>13</sup> For purposes of the Fourth Amendment, a "search" is conducted when the government, acting on its own or through an authorized agent, intrudes into a person's "constitutionally protected reasonable expectation of privacy."<sup>14</sup> The

---

12. U.S. CONST. amend. IV.

13. *Id.*

14. *California v. Ciraolo*, 476 U.S. 207, 211 (1986) (quoting *Katz v. United States*, 389 U.S. 347, 360 (1967) (Harlan, J., concurring)); *see also Soldal v. Cook County*, 506 U.S. 56, 63 (1992) (quoting *United States v. Jacobsen*, 466 U.S. 109, 113 (1984)) ("A 'search' occurs when an expectation of privacy that society is prepared to consider reasonable is infringed."); *United States v. Daniels*, 60 M.J. 69, 71 (C.A.A.F. 2004) ("The Supreme Court defines a

Supreme Court highlighted this emphasis on privacy rights, as opposed to property rights, in Fourth Amendment analysis as early as 1967 when it observed in *Warden v. Hayden* that “[w]e have recognized that the principal object of the Fourth Amendment is the protection of privacy rather than property . . . . This shift in emphasis from property to privacy has come about through a subtle interplay of substantive and procedural reform.”<sup>15</sup> An individual’s reasonable expectation of privacy now serves as the “touchstone” of Fourth Amendment inquiry.<sup>16</sup>

To determine whether an individual’s privacy expectations are constitutionally protected, courts generally apply a two-part test, first articulated by Justice Harlan in his concurring opinion in *Katz v. United States*, a case involving the expectation of privacy in telephone calls.<sup>17</sup> Observing that “the Fourth Amendment protects people, not places,” Justice Harlan conceived a sliding scale of protections, in which the appropriate level of protection is determined by inquiring (1) whether the person involved has a subjective (actual) expectation of privacy, and (2) whether that expectation is objectively reasonable.<sup>18</sup> A privacy expectation is objectively reasonable if it is “one that society is prepared to recognize as ‘reasonable.’”<sup>19</sup> The first query is a question of fact, while the second is a matter of law.<sup>20</sup>

Importantly, situations exist in which the two-part *Katz* inquiry may be inadequate. Writing for the majority in *Smith v. Maryland*, Justice Blackmun commented:

[f]or example, if the Government were suddenly to announce on nationwide television that all homes henceforth would be subject to warrantless entry, individuals thereafter might not in fact entertain any actual expectation of privacy regarding their homes, papers, and effects. . . . In such circumstances, where an

---

Fourth Amendment ‘search’ as a government intrusion into an individual’s reasonable expectation of privacy.”).

15. 387 U.S. 294, 304 (1967) (internal citations omitted); *cf. Soldal*, 506 U.S. at 64 (observing that, although protection of privacy is “principal” object of contemporary Fourth Amendment jurisprudence, protection of property remains integral).

16. *Ciraolo*, 476 U.S. at 211.

17. 389 U.S. at 361.

18. *Id.*; *see also* *Smith v. Maryland*, 442 U.S. 735, 740 (1979) (discussing *Katz* and noting that a subjective expectation exists when an individual “seeks to preserve [something] as private” and an objective expectation exists when, “viewed objectively, [it] is ‘justifiable’ under the circumstances” (internal quotation marks omitted) (citations omitted)).

19. *Katz*, 389 U.S. at 361 (Harlan, J., concurring); *see also* *Minnesota v. Olson*, 495 U.S. 91, 95–96 (1990).

20. *United States v. Maxwell*, 45 M.J. 406, 417 (C.A.A.F. 1996) (quoting *Tri-State Steel Construction Inc. v. Occupational Safety & Health Review Comm’n*, 26 F.3d 173, 176 (D.C. Cir. 1994)).

individual's subjective expectations had been "conditioned" by influences alien to well-recognized Fourth Amendment freedoms, those subjective expectations obviously could play no meaningful role in ascertaining what the scope of Fourth Amendment protection was. In determining whether a "legitimate expectation of privacy" existed in such cases, a normative inquiry would be proper.<sup>21</sup>

Justice Blackmun's comment underscores the importance of basing Fourth Amendment protections on normative valuations of what should be protected rather than simply what is protected, especially when the government may have sought to perform an end-run around Fourth Amendment privacy expectations by issuing blanket pronouncements of prescribed intrusions.<sup>22</sup> Even in *Katz*, in which the facts supported the Court's two-pronged inquiry, the core holding rested on a normative judgment that an individual placing a telephone call "is surely entitled to assume that the words he utters into the mouthpiece will not be broadcast to the world."<sup>23</sup> Such normative judgments, as discussed in Section IV.C below, continue to play a critical role in evaluating society's expanding reliance on communications technologies.<sup>24</sup>

---

21. 442 U.S. at 741 n.5. In *Smith*, the Supreme Court held that using a "pen register" to record a dialed phone number did not violate the Fourth Amendment because individuals "realize that they must 'convey' phone numbers to the telephone company, since it is through telephone company switching equipment that their calls are completed." *Id.* at 742. For an excellent discussion of how *Smith* applies to electronic communications, see Matthew J. Tokson, *The Content/Envelope Distinction in Internet Law*, 50 WM. & MARY L. REV. 2105, 2110 (2009) ("Perhaps the most practically significant of these unresolved questions is whether novel categories of Internet communications data, such as e-mail subject lines, website Uniform Resource Locators (URLs), and website IP addresses should be protected as the contents of electronic communications, or whether they should be treated as noncontent 'envelope' information."). In *Larson*, it was precisely this "novel" internet information which law enforcement agents seized and which the trial court admitted. *See infra* Section III.C.4.

22. Professor Susan Freiwald has observed that one of the Framers' motivations for the Fourth Amendment was their concern about the government issuing blanket pronouncements in the form of general warrants. Susan Freiwald, *First Principles of Communications Privacy*, 2007 STAN. TECH. L. REV. 3, 59 (2007); *see also* *Berger v. New York*, 388 U.S. 41, 64 (1967) (striking down New York's wiretapping statute as violating the Fourth Amendment's particularity requirement similar to a general warrant); *infra* Section IV.B.1 (discussing how the DoD's mandatory logon banner functions like a general warrant).

23. *Katz*, 389 U.S. at 352.

24. *See infra* Section IV.C; *see also* Freiwald, *supra* note 22, at 29 (discussing *Katz* court's normative acknowledgment of "vital role that the public telephone has come to play in private communications" and comparing it with vital role of e-mail in contemporary society).

## B. THE MILITARY AND CONSTITUTIONAL PROTECTIONS

Constitutional protections “take on a different complexion” when applied to members of the military.<sup>25</sup> As the Supreme Court observed in *Parker v. Levy*, “the different character of the military community and of the military mission requires a different application of those protections. The fundamental necessity for obedience, and the consequent necessity for imposition of discipline, may render permissible within the military that which would be constitutionally impermissible outside of it.”<sup>26</sup> Courts have observed this alternate constitutional complexion in the context of the First Amendment,<sup>27</sup> Fifth Amendment,<sup>28</sup> Sixth Amendment,<sup>29</sup> Seventh Amendment,<sup>30</sup> and, as discussed in this Article, the Fourth Amendment.<sup>31</sup>

---

25. See *United States v. Allen*, No. ACM 32727, 1999 CCA LEXIS 116, at \*11 (A.F. Ct. Crim. App. Apr. 22, 1999). Jurists have long recognized the distinctive nature of military service. See, e.g., *In re Grimley*, 137 U.S. 147, 153 (1890) (“An army is not a deliberative body. It is the executive arm. Its law is that of obedience. No question can be left open as to the right to command in the officer, or the duty of obedience in the soldier.”). This distinction implicates the color and scope of constitutional protections. See *Solorio v. United States*, 483 U.S. 435, 447–48 (1987) (discussing constitutional protections for military members); *Parker v. Levy*, 417 U.S. 733, 758 (1974) (same).

26. 417 U.S. at 758.

27. See, e.g., *id.* at 761 (holding speech of commissioned officer urging enlisted personnel to refuse to obey orders not protected under First Amendment); *United States v. Forney*, 67 M.J. 271, 274–75 (C.A.A.F. 2009) (holding virtual child pornography not protected under military law even if protected under First Amendment in civilian society); see also John A. Carr, *Free Speech in the Military Community: Striking a Balance Between Personal Rights and Military Necessity*, 45 A.F. L. REV. 303 (1998); Katherine C. Den Bleyker, *The First Amendment versus Operational Security: Where Should the Milblogging Balance Lie?*, 17 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 401 (2007); David E. Fitzkee & Linell A. Letendre, *Religion in the Military: Navigating the Channel Between the Religion Clauses*, 59 A.F. L. REV. 1 (2007).

28. See, e.g., *Weiss v. United States*, 510 U.S. 163, 177 (1994) (holding due process under Fifth Amendment differs in military); *Ex parte Quirin*, 317 U.S. 1, 40 (1942) (noting Fifth Amendment explicitly excepts “cases arising in the land or naval forces”); *United States v. Culp*, 14 C.M.A. 199, 205 (C.M.A. 1963) (“[N]either has the Supreme Court held that all of the rights covered by the Fifth and Sixth Amendments apply to those in the military.”).

29. See, e.g., *Quirin*, 317 U.S. at 40 (noting Sixth Amendment exempts by implication “cases arising in the land or naval forces”); *United States v. Witham*, 47 M.J. 297, 301 (C.A.A.F. 1997) (noting “a military accused has no Sixth Amendment right to trial by jury”).

30. See *Culp*, 14 U.S.C.M.A. at 208 (observing Seventh Amendment right to trial by jury is limited to suits at common law, which does not include “suits in equity or proceedings in admiralty”).

31. See, e.g., *Comm. for GI Rights v. Callaway*, 518 F.2d 466, 476–77 (D.C. Cir. 1975) (holding warrantless inspections for drug abuse not violative of Fourth Amendment); *United States v. McCarthy*, 38 M.J. 398, 401 (C.M.A. 1993) (“Fourth Amendment requirement that a warrant be supported by oath or affirmation does not apply to searches and seizures by military authorities” (citation omitted)); *United States v. Stuckey*, 10 M.J. 347, 360–61 (C.M.A. 1981) (noting military commander, though not magistrate for Fourth Amendment

These provisional limitations flow from judicial recognition that Congress retains a plenary role “over rights, duties, and responsibilities in the framework of the Military Establishment, including regulations, procedures, and remedies related to military discipline[.]”<sup>32</sup> It is for this reason that the Supreme Court acknowledged that “[j]udicial deference . . . is at its apogee when reviewing congressional decisionmaking in this area.”<sup>33</sup> In modern jurisprudence, it is settled law that the military is a “separate community” subject to congressional regulation with limited oversight by the federal judiciary.<sup>34</sup>

While military law limits certain rights, others are expanded beyond constitutional minimums by federal statutes, executive orders, federal case law, executive orders, or service regulations.<sup>35</sup> Two common examples include the right to counsel before special and general courts martial, regardless of the indigent circumstances of the accused,<sup>36</sup> and the requirement to administer warnings against self-incrimination, regardless of

purposes, may authorize searches).

32. *Chappell v. Wallace*, 462 U.S. 296, 301 (1983). As set forth in Art. I, § 8, the legislature has broad responsibility for both maintaining and directing the military:

Congress shall have Power . . . To make Rules for the Government and Regulation of the land and naval Forces [and] to provide for organizing, arming, and disciplining, the Militia, and for governing such Part of them as may be employed in the Service of the United States . . . .

U.S. CONST. art. I, § 8; *see also* *Solorio v. United States*, 483 U.S. 435, 447–48 (1987) (“Congress has primary responsibility for the delicate task of balancing the rights of servicemen against the needs of the military. . . . [W]e have adhered to this principle of deference in a variety of contexts where, as here, the constitutional rights of servicemen were implicated.”).

33. *Weiss*, 510 U.S. at 177 (citation omitted) (internal quotation marks omitted).

34. *See* *Parker v. Levy*, 417 U.S. 733, 743 (1974). The Court noted

[t]his Court has long recognized that the military is, by necessity, a specialized society separate from civilian society. . . . The differences between the military and civilian communities result from the fact that “it is the primary business of armies and navies to fight or be ready to fight wars should the occasion arise.”

*Id.* (quoting *United States ex rel. Toth v. Quarles*, 350 U.S. 11, 17 (1955)). For an insightful discussion of the “separate community” doctrine, see James M. Hirschhorn, *The Separate Community: Military Uniqueness and Servicemen’s Constitutional Rights*, 62 N.C. L. REV. 177 (1984).

35. *See generally* Francis A. Gilligan, *The Bill of Rights and Service Members*, 1987 ARMY LAW. 3 (1987) (discussing Bill of Rights as applied to service members and arguing service members sometimes enjoy rights broader than those required by Constitution).

36. 10 U.S.C. § 838(b)(1) (2006) (providing that an “accused has the right to be represented in his defense before a general or special court-martial or at an investigation under [Article 32]” while implementing procedures for detailing defense counsel). Military defense counsel are provided without expense to the accused. *See, e.g.*, *United States v. Culp*, 14 U.S.C.M.A. 199, 202 (C.M.A. 1963) (“[N]o service man appears before a court-martial alone and there are no ‘indigents’ before courts-martial. Inevitably, by law, the accused before a court-martial appears with appointed counsel and no funds are required.”).

the custodial nature of the interrogation.<sup>37</sup> The basis for this rights expansion is often that, although both visible and understandable within the canopy of contemporary legal society, the roots of American military jurisprudence predate the Constitution, the Articles of Confederation, and the American Revolution itself, finding their origin in Roman and British military tradition.<sup>38</sup> Military jurisprudence thus exists distinct and separate from the law governing Article III courts.<sup>39</sup>

---

37. Article 31(b) of the Uniform Code of Military Justice (UCMJ) provides that: [n]o person subject to this chapter may interrogate, or request any statement from, an accused or a person suspected of an offense without first informing him of the nature of the accusation and advising him that he does not have to make any statement regarding the offense of which he is accused or suspected and that any statement made by him may be used as evidence against him in a trial by court-martial.

10 U.S.C. § 831(b) (2006); *see also Culp*, 14 U.S.C.M.A. at 206 (discussing protections afforded service members under Bill of Rights and noting “[h]e cannot be compelled to be a witness against himself and that provision of the Fifth Amendment is preserved and expanded” (citations omitted)). Service members also have broader rights under the Eighth Amendment. *See United States v. Matthews*, 16 M.J. 354, 368 (C.M.A. 1983) (“[W]e have held that, in enacting Article 55, Congress ‘intended to grant protection covering even wider limits’ than ‘that afforded by the Eighth Amendment.’” (quoting *United States v. Wappler*, 2 U.S.C.M.A. 393, 396 (C.M.A. 1953))).

38. *See Culp*, 14 U.S.C.M.A. at 205. For an introduction to the historical roots of the U.S. military legal system, see Ruzic, *supra* note 6, at 266–70.

39. *See generally* Note, *Military Justice and Article III*, 103 HARV. L. REV. 1909 (1990) (discussing military courts and their relationship to Article III courts). In *Burns v. Wilson*, 346 U.S. 137, 140 (1953), the Supreme Court rejected two applications for writs of habeas corpus from petitioners who had been convicted at court-martial of murder and rape and sentenced to death. In reviewing a civil court’s power to review a military court-martial’s judgment, the Supreme Court commented on the distinctive nature of military law:

[m]ilitary law, like state law, is a jurisprudence which exists separate and apart from the law which governs in our federal judicial establishment. This Court has played no role in its development; we have exerted no supervisory power over the courts which enforce it; the rights of men in the armed forces must perforce be conditioned to meet certain overriding demands of discipline and duty, and the civil courts are not the agencies which must determine the precise balance to be struck in this adjustment. The Framers expressly entrusted that task to Congress.

*Id.* at 140 (internal citations omitted). The Supreme Court’s reasoning echoed sentiments expressed nearly a century earlier in *Dynes v. Hoover*, 61 U.S. 65 (1857):

[t]hese provisions show that Congress has the power to provide for the trial and punishment of military and naval offences in the manner then and now practiced by civilized nations; and that the power to do so is given without any connection between it and the 3d article of the Constitution defining the judicial power of the United States; indeed, that the two powers are entirely independent of each other.

*Id.* at 79; *cf. In re Grimley*, 137 U.S. 147, 150 (1890) (“[I]t is equally clear that by habeas corpus the civil courts exercise no supervisory or correcting power over the proceedings of a

Despite being a “separate community,” service members are not outside the umbrella of constitutional protections. Quoting Blackstone, the Court of Military Appeals explained in *United States v. Culp*, “he puts not off the citizen when he enters the camp; but it is because he is a citizen, and would wish to continue so, that he makes himself for a while a soldier.”<sup>40</sup> Courts thus apply the Bill of Rights to service members provision-by-provision, similar to the manner in which the Bill of Rights applies to the states.<sup>41</sup>

### C. THE FOURTH AMENDMENT AS APPLIED TO THE MILITARY

Within the context of the Fourth Amendment, “[t]he ‘expectation of privacy’ is different in the military than it is in civilian life.”<sup>42</sup> The foundation for this jurisprudential distinction lies in both necessity and custom. The responsibilities of a military commander extend to the welfare and combat readiness of his troops, investigation of their alleged misconduct, and the safety of the installation on which they live.<sup>43</sup> As a result of these responsibilities, which often are carried out in foreign, remote locations, military commanders necessarily require the power to inspect, search, and seize both persons and property under their command.<sup>44</sup> The breadth of such powers is supported by military custom, which “has long granted military commanders broad powers of search and seizure,”<sup>45</sup> and is now expressly provided for in the MRE.<sup>46</sup>

---

court[-]martial; and that no mere errors in their proceedings are open to consideration. The single inquiry, the test, is jurisdiction.”).

40. 14 C.M.A. at 206 (quoting WILLIAM BLACKSTONE, COMMENTARIES \*408 (Wendell ed.)) (internal quotation marks omitted).

41. *See* *United States v. Stuckey*, 10 M.J. 347, 349 (C.M.A. 1981). The court noted [t]he time is long past when scholars disputed the applicability of the Bill of Rights to service personnel. Instead, our premise must be that the Bill of Rights applies with full force to men and women in the military service unless any given protection is, expressly or by necessary implication, inapplicable.

*Id.* (internal quotation marks and citation omitted).

42. *Comm. for GI Rights v. Callaway*, 518 F.2d 466, 477 (D.C. Cir. 1975) (internal citations omitted).

43. *Stuckey*, 10 M.J. at 359.

44. *Id.*

45. *Id.* at 360 (citing *United States v. Middleton*, 10 M.J. 123, 126 (C.M.A. 1981)).

46. *See* MIL. R. EVID. 312–317. Although memorialized in Military Rule of Evidence (MRE) 315, which addresses a commander’s power to authorize searches based on probable cause, the basis for search authority is rooted in the nature of command itself. As the *Stuckey* court noted, “the commander’s long-recognized power to authorize searches within the area of his command is generally viewed as derived from and correlative with his position and responsibilities in the military community—which, of course, is ‘a specialized society separate from civilian society.’” 10 M.J. at 360 (quoting *Parker v. Levy*, 417 U.S. 733, 743 (1974)). Of course, a commander’s power to authorize probable cause searches is subject to

The Military Rules of Evidence (MRE) recognize several intrusions into potential zones of privacy, only some of which require a prior probable cause determination. MRE 312, for example, outlines the procedures for consensual and non-consensual visual examinations of the body, body cavity intrusions, extraction of body fluids (i.e., blood and urine), and intrusions for medical necessity. All of these, with the exception of intrusions for medical necessity, must be based on consent or must be conducted as either a lawful inspection or as a search and seizure pursuant to probable cause.<sup>47</sup>

MRE 313 provides for the admissibility of evidence obtained from inspections and inventories.<sup>48</sup> An “inspection” is an examination of a unit, installation, vessel, aircraft, or vehicle conducted by a command “to determine and to ensure the security, military fitness, or good order and discipline.”<sup>49</sup> The primary purpose of the inspection cannot be to obtain evidence of criminal wrongdoing.<sup>50</sup> Similarly, an “inventory,” which must be conducted for a legitimate administrative purpose, cannot be used as a ruse to obtain evidence of criminal wrongdoing.<sup>51</sup>

MRE 314, drawing on well-established case law, identifies ten searches not requiring probable cause: (1) border searches; (2) searches upon entry to or exit from a U.S. installation, aircraft, or vessel abroad; (3) searches of government property; (4) consent searches; (5) searches incident to a lawful stop; (6) searches incident to a lawful apprehension; (7) searches within jails and confinement facilities; (8) emergency searches to save life; (9) searches of open fields or woodlands; and (10) other searches not requiring probable cause “under the Constitution of the United States as applied to members of the armed forces.”<sup>52</sup> Even within these ten areas, however, searches may be unlawful if the primary purpose of the search is to obtain evidence of

---

scrutiny. *See* 10 U.S.C. § 898 (2006) (addressing noncompliance with procedural rules); 10 U.S.C. § 938 (2006) (addressing complaints of wrongs).

47. MIL. R. EVID. 312.

48. MIL. R. EVID. 313.

49. MIL. R. EVID. 313(b).

50. *Id.* One commentator has suggested that the military’s authority to monitor its information systems stems from its authority to conduct inspections pursuant to MRE 313. *See* Edell, *supra* note 9, at 23. Because MRE 313 expressly prohibits inspections for law enforcement purposes, it cannot serve as a basis for the military to search workplace communications systems for “any” purpose. *See infra* text accompanying notes 307–10 (discussing DoD’s new logon banner, which permits inspections for any purpose). As a result, the military must look elsewhere for its authority to conduct warrantless searches into protected zones of privacy. *See* Edell, *supra* note 9, at 23 (arguing Army policy allowing computer monitoring for law enforcement purposes violates Supreme Court precedent).

51. MIL. R. EVID. 313(c).

52. MIL. R. EVID. 314.

criminal wrongdoing or if the search is conducted contrary to an individual's reasonable expectation of privacy.<sup>53</sup> If a reasonable expectation of privacy exists, an impartial commander or military judge must give search authorization based on probable cause.<sup>54</sup>

With respect to searches not requiring probable cause, military courts have adopted the Supreme Court's two-part *Katz* inquiry to analyze the reasonableness of service members' privacy expectations.<sup>55</sup> In *United States v. Daniels*, the CAAF considered whether a service member appellant's Fourth Amendment rights were violated when his barracks roommate seized a vial of cocaine from his bedside nightstand at the direction of a ranking military member. After finding that the appellant had a reasonable expectation of privacy in his nightstand and that the roommate was acting as an agent of the government, the court held the warrantless search of the nightstand unlawful.<sup>56</sup> *Daniels* holds that service members, like civilians, receive Fourth Amendment protections when they have a subjective expectation of privacy that is objectively reasonable.<sup>57</sup> Also, *Daniels* holds that in the military, as elsewhere, private actors may become agents of the government when their actions are "encouraged, endorsed, and participated in" by government officials acting in a law enforcement capacity.<sup>58</sup>

Perhaps the most fundamental difference in the Fourth Amendment as applied to military members lies not in the prohibition against unreasonable searches, but in the requirement to obtain a search warrant supported by an oath.<sup>59</sup> MRE 315 distinguishes between a search warrant, which is issued by a

---

53. For example, the primary purpose of searches conducted upon entrance to or exit from an installation cannot be to "obtain[] evidence for use in a trial by court-martial or other disciplinary proceeding . . ." MIL. R. EVID. 314(c). Also, searches of government property are not permissible if "the person to whom the property is issued or assigned has a reasonable expectation of privacy" depending on "the facts and circumstances at the time of the search." MIL. R. EVID. 314(d).

54. MIL. R. EVID. 315(d). That the privacy expectation of service members may be intruded upon in certain circumstances without issuance of a warrant is another example of the provision-by-provision application of the Bill of Rights to the armed forces. For a discussion of the distinction between search warrants and search authorization, see *infra* text accompanying notes 59–62. For a discussion of the provisional application of the Bill of Rights to service members, see *supra* Section II.B.

55. See, e.g., *United States v. Daniels*, 60 M.J. 69, 71 (C.A.A.F. 2004); *United States v. Portt*, 21 M.J. 333, 334–35 (C.M.A. 1986).

56. *Daniels*, 60 M.J. at 71–72.

57. *Id.* at 71.

58. *Id.*; see also *Skinner v. Ry. Labor Executives Ass'n*, 489 U.S. 602, 614 (1989) ("Whether a private party should be deemed an agent or instrument of the Government for Fourth Amendment purposes necessarily turns on the degree of the Government's participation in the private party's activities.").

59. See Gilligan, *supra* note 35, at 4 ("In at least one area the courts have applied

competent civilian authority pursuant to the Fourth Amendment's "warrant" clause, and search authorization, which is issued by a competent military authority pursuant to military law.<sup>60</sup> This distinction was challenged in *United States v. Chapman*, in which a service member who was the subject of a civilian prosecution urged the Seventh Circuit Court of Appeals to suppress evidence obtained from the service member's barracks room pursuant to a military commander's search authorization.<sup>61</sup> Holding that the military search authorization complied with the Fourth Amendment's prohibition against unreasonable searches, the court stated:

It is too late in the day to suggest that the Fourth Amendment's basic protection against unreasonable searches and seizures does not apply to members of the armed forces. Nevertheless, the military implementation of that guarantee is different from that employed in civilian matters. In civilian cases, the warrant requirement has been abrogated by judicial decision only in certain carefully described situations. In the military situation, Congress "has primary responsibility for the delicate task of balancing the rights of servicemen against the needs of the military."<sup>62</sup>

The holding in *Chapman* captures the essence of Fourth Amendment privacy protections for service members. While the procedural "complexion" of military privacy rights may differ from their civilian counterparts, the underlying constitutional protection against "unreasonable searches and seizures" remains. As with civilians, the challenge is to determine what privacy protections are reasonable.

#### D. PRIVACY IN THE PUBLIC WORKPLACE

Although the military workplace differs from the civilian workplace, the judicially recognized privacy interests of public employees are instructive in determining the boundaries of service members' workplace privacy interests. With respect to public employees, courts have held that "[i]ndividuals do not lose Fourth Amendment rights merely because they work for the

---

separate standards. That is in the area of the oath. The information given to the judge or the commander need not be under oath, although an oath is preferred.").

60. MIL. R. EVID. 315(b)(2), (d). In the U.S. Air Force, the competent military authority is usually an installation "magistrate" appointed by an installation commander to conduct probable cause search inquiries for on-base searches. AFI 51-201, *supra* note 9, § 3.1. The U.S. Army also has a magistrate program for probable cause search authorizations, although it differs in some respects. U.S. DEP'T OF ARMY, ARMY REG. 27-10, MILITARY JUSTICE (2005).

61. 954 F.2d 1352 (7th Cir. 1992).

62. *Id.* at 1367 (quoting *Solorio v. United States*, 483 U.S. 435, 447 (1987)).

government instead of a private employer.”<sup>63</sup> Rather, public employees enjoy a limited expectation of privacy in their workplace, like those in the private sector, to the extent that the expectation is reasonable under all the circumstances.<sup>64</sup> In reviewing reasonableness, the context of the employment relationship is determinative. Intrusion by a supervisor is more reasonable than intrusion by law enforcement.<sup>65</sup> Office practices and procedures may reduce employees’ expectations of privacy in their offices, desks, file cabinets, and computers.<sup>66</sup> Even then, however, some constitutional protections may remain.<sup>67</sup>

In *O’Connor v. Ortega*, the Supreme Court considered the scope of a public employee’s privacy expectations in his workplace office.<sup>68</sup> Noting that public employers have wide latitude in entering employees’ private workspaces for “the efficient and proper operation of the workplace,” the Court drew clear lines between employer intrusion for work-related and law enforcement purposes.<sup>69</sup> For example, an employer may invade an employee’s private workspace to retrieve a file. An employer may even invade an employee’s workspace in furtherance of a work-related investigation. When the employer becomes an agent of law enforcement, however, and acts within the shadow of a criminal investigation, the employer’s legitimate, work-related purposes end, and an intrusion of an altogether different kind begins.<sup>70</sup>

---

63. *O’Connor v. Ortega*, 480 U.S. 709, 717 (1987) (plurality opinion).

64. *Id.*

65. *Id.*; see also *Mancusi v. DeForte*, 392 U.S. 364 (1968). In the context of military searches, the contrast is even more stark, because the military is both employer and police. See *infra* text accompanying notes 80–81.

66. *O’Connor*, 480 U.S. at 723.

67. In his concurring opinion in *O’Connor*, Justice Scalia criticizes the plurality’s methodology for pinning the expectation of privacy a public employee has in his office to the frequency of office visitors:

in my view, one’s personal office is constitutionally protected against warrantless intrusions by the police, even though employer and co-workers are not excluded. . . . Constitutional protection against *unreasonable* searches by the government does not disappear merely because the government has the right to make reasonable intrusions in its capacity as employer.

*Id.* at 730–31 (Scalia, J., concurring).

68. *Id.* at 709 (majority opinion).

69. *Id.* at 721, 723 (“While police, and even administrative enforcement personnel, conduct searches for the primary purpose of obtaining evidence for use in criminal or other enforcement proceedings, employers most frequently need to enter the offices and desks of their employees for legitimate work-related reasons wholly unrelated to illegal conduct.”).

70. *Id.* at 722 (“In contrast to other circumstances in which we have required warrants, supervisors in offices such as at the Hospital are hardly in the business of investigating the violation of criminal laws.”). In this sense, the civilian workplace is distinguishable from the military workplace, where the roles of supervisor, criminal investigator, and, at times,

The Supreme Court established a standard of reasonableness in the context of legitimate, work-related intrusions. As the Supreme Court held in *O'Connor*, “public employer intrusions on the constitutionally protected privacy interests of government employees for noninvestigatory, work-related purposes, as well as for investigations of work-related misconduct, should be judged by the standard of reasonableness under all the circumstances.”<sup>71</sup> In practice, the test is simply whether the intrusion was “justified at its inception” and, as actually conducted, whether it was “reasonably related in scope to the circumstances which justified the interference in the first place.”<sup>72</sup>

When a public employer’s search is carried out for criminal, rather than work-related, purposes, the reasonableness standard gives way to the probable cause standard, which generally requires a search warrant. Surveying applicable case law in *O'Connor*, the Supreme Court observed, “[t]he only cases to imply that a warrant should be required involve searches that are not work related, or searches for evidence of criminal misconduct . . . .”<sup>73</sup>

The distinction between work-related and law enforcement searches is important. In *United States v. Kaban*, which the Supreme Court cited in *O'Connor*, the defendant, an employee of what then was the Immigration and Naturalization Service (INS), moved to suppress documentary evidence of criminal misconduct taken by INS agents from a desk-side wastebasket used exclusively by the defendant.<sup>74</sup> INS agents had been routinely searching the defendant’s wastebasket as part of a criminal investigation into his official duties. In discussing whether searching the wastebasket constituted a “search” for purposes of the Fourth Amendment, the district court noted (1) the sole purpose of the intrusion was to obtain criminal evidence; (2) the case was distinguishable from those in which a supervisor or co-worker chanced upon criminal evidence while looking through an employee’s desk or wastebasket; and (3) the case did not involve a supervisor engaged in a work-related inspection.<sup>75</sup> The court acknowledged that the government, like private employers, “should be able to manage its . . . offices effectively and without undue restrictions on the supervision of its employees.”<sup>76</sup> The court

---

enforcer of criminal law are merged. *See infra* text accompanying notes 80–81.

71. 480 U.S. at 725–26.

72. *Id.* at 726 (quoting *Terry v. Ohio*, 392 U.S. 1, 20 (1968)).

73. *Id.* at 721.

74. 350 F. Supp. 784, 789–90 (S.D.N.Y. 1972).

75. *Id.* at 791.

76. *Id.*

placed special emphasis, however, on the distinction between private and public employers when employee criminal misconduct is suspected:

If a private employer suspects misconduct on the part of an employee, he will not ordinarily conduct an investigation to substantiate criminal charges against him. Rather, he will simply fire that employee. . . . In contrast, when a government supervisor begins an investigation of suspected criminal activities of an employee in the course of his work, the supervisor's role is no longer that of a manager of an office, but that of a criminal investigator for the government. The purpose of the supervisor's surveillance is no longer simply to preserve efficiency in the office. It is specifically designed to prepare a criminal prosecution against the employee. In that case, searches and seizures by the supervisor or by other government agents are governed by the Fourth Amendment admonition that a warrant be obtained in the absence of exigent circumstances.<sup>77</sup>

Finding the search constituted a Fourth Amendment search, the district court then turned to whether the defendant had a reasonable expectation of privacy in the wastebasket contents.<sup>78</sup> The court concluded (1) government employees, like their private sector counterparts, have an expectation of privacy in their workspace from government criminal investigators; (2) government supervisors may not provide third-party consent to the search of office areas "reserved for the exclusive use of a particular employee"; and (3) by throwing his documents into his wastebasket, the defendant had not abandoned them.<sup>79</sup> The contested evidence, having been obtained in violation of the Fourth Amendment, was suppressed.<sup>80</sup>

#### E. PRIVACY IN THE MILITARY WORKPLACE

The standard for workplace searches is similar in both the military and non-military context. However, in a military workplace the lines between work-related and law enforcement searches are less defined because the roles

---

77. *Id.*

78. *Id.* at 791–92.

79. *See id.* at 795–96. With respect to "abandonment," the district court noted the analysis does not consider whether defendant relinquished possessory interest, but whether he relinquished his privacy interest. The district court adopted

the defense counsel's analogy, offered in oral argument, that a wastebasket should be likened to a paper shredder in that whatever a person throws in, he expects to be destroyed. Once destroyed, the contents and appearance of the papers disposed of are solely within the knowledge of that person and, hence, absolutely private.

*Id.* at 796 n.8. The parallels to electronic documents and e-mails which are deleted by the user but nevertheless retained on the hard drive are obvious.

80. *Id.* at 797.

of supervisor, criminal investigator, and law enforcement agent merge in the form of the commander. As one military court noted,

[w]e must note that the military workplace is not the usual workplace envisioned by the Supreme Court in *O'Connor*. . . . Military commanders have authority and powers not possessed by civilian employers. Military commanders, for example, can authorize searches of their personnel, order them confined, and bring criminal charges against them.<sup>81</sup>

Additionally, actions that may constitute simple workplace misconduct for civilians may constitute criminal offenses for service members.<sup>82</sup> If anything, this would seem to heighten the need for probable cause search authorization prior to conducting searches for wrongdoing, a fact not lost on military courts.

In *United States v. Muniz*, decided just before *O'Connor*, the United States Court of Military Appeals considered whether a service member had a legitimate expectation of privacy in the government-issued credenza inside his government office.<sup>83</sup> While Captain Muniz was on leave for a secret tryst with a female service member, members of his command entered his office, opened his locked credenza, and copied an address from an envelope they thought might identify his location.<sup>84</sup> Their motivation in doing so was to notify Captain Muniz of a medical emergency involving his daughter; only later did they discover he had, among other things, falsified his leave address to disguise his clandestine affair.<sup>85</sup> In canvassing the constitutional landscape of workplace searches, the military court noted Captain Muniz maintained a separate office, the office could be locked, staff members occasionally entered each other's work areas even when the primary occupant was absent, the locked credenza (though government property) was used only by Captain Muniz, and Captain Muniz had the credenza's only key.<sup>86</sup> Comparing military searches to civilian workplaces searches, the military court then observed, "[t]he only seemingly complicating factor in the military is that sometimes business-supervisor and law-enforcement authority merge in the person of

---

81. *Long*, 64 M.J. at 62 (citations omitted).

82. For example, failing to properly perform assigned duties, feigning illness in order to miss work, and being late for work can all result in criminal charges under the UCMJ. *See* 10 U.S.C. §§ 886, 892, 915 (2006). For additional discussion of the merger of workplace and criminal misconduct in the military, see *infra* text accompanying note 365.

83. 23 M.J. 201, 204 (C.M.A. 1987).

84. *Id.* at 203–04.

85. *Id.* at 202–04.

86. *Id.* at 204–05.

the commander.”<sup>87</sup> Without fully resolving this “complicating factor,” the court concluded it was the reasonableness of the appellant’s privacy expectation claim, rather than “which hat the commander happens to be wearing that day,” which controlled, and determined Captain Muniz did not have a legitimate expectation of privacy in his credenza from a search to discover his whereabouts, a non-criminal purpose.<sup>88</sup>

In a case the following year, *United States v. Breseman*, the United States Court of Military Appeals briefly addressed, but did not have occasion to resolve, the issue of a government search authorized by a commander acting in his law enforcement (as opposed to supervisory) capacity.<sup>89</sup> In *Breseman*, a military commander provided probable cause search authorization to a Coast Guard Intelligence agent who wanted to search the defendant’s government-owned desk. After interviewing the agent and reviewing the supporting affidavit and evidence, the commander authorized the search pursuant to MRE 315.<sup>90</sup> Without deciding whether the defendant had a reasonable expectation of privacy in his desk, the court held the commander adequately safeguarded the defendant’s rights by complying with MRE 315.<sup>91</sup> Under *Breseman*, even if there is a reasonable expectation of privacy, if the search is related to law enforcement, obtaining military search authorization is sufficient to protect a service member’s Fourth Amendment rights.

Neither *Muniz* nor *Breseman* definitively answer the question of whether a service member’s reasonable workplace privacy interest can be searched for criminal evidence without probable cause search authorization.<sup>92</sup> Applying

---

87. *Id.* at 205.

88. *Id.* at 205–06. Though dated, for a thorough discussion of military case law regarding military members’ right to privacy in their government desks, see David P. Arcuri, *Muniz, Breseman, Craig, and the Right to Privacy in a Government-Owned Desk*, 1992 ARMY LAW. 26, 28 (1992) (discussing the Military Court of Appeals’ express rejection of “a bright-line rule that a service member never may have a reasonable expectation of privacy in a government-owned desk”). The court’s holding in *Breseman* also underscores the protections afforded service members under MRE 314, which requires probable cause if a reasonable privacy interest exists and the primary purpose of the search is to obtain evidence of criminal wrongdoing. See MIL. R. EVID. 314(d).

89. 26 M.J. 398 (C.M.A. 1988). The Court of Military Appeals first applied *O’Connor* to the military workplace in *United States v. Battles*, in which the court noted “the scope of the expectation of privacy depends in part on the demands of the workplace and its openness to employees and the public.” *United States v. Craig*, 32 M.J. 614, 615 (A.C.M.R. 1991) (citing *United States v. Battles*, 5 M.J. 58, 60 (C.M.A. 1987)).

90. *United States v. Breseman*, 26 M.J. 398, 399 (C.M.A. 1988).

91. *Id.* at 400. In his concurring opinion, Chief Judge Everett phrased the holding in a slightly different manner: “I conclude that the commanding officer had probable cause to authorize a search of appellant’s desk . . . and that he authorized a search in compliance with the Military Rules of Evidence and the Fourth Amendment.” *Id.* (Everett, J., concurring).

92. In *United States v. Craig*, decided in 1991, the U.S. Army Court of Military Review

the reasoning in *Kaban*, however, one would expect military courts to conclude that government searches conducted for law enforcement purposes “are governed by the Fourth Amendment admonition that a warrant be obtained in the absence of exigent circumstances.”<sup>93</sup> This was precisely the issue confronted and decided by the CAAF in *United States v. Long*,<sup>94</sup> discussed in Section III.C.3 below.<sup>95</sup> There, the CAAF held that service members may have a reasonable expectation of privacy from law enforcement searches in their workplace e-mail,<sup>96</sup> a reflection of Justice Warren’s comment that “our citizens in uniform may not be stripped of basic rights simply because they have doffed their civilian clothes.”<sup>97</sup>

#### F. FIRST PRINCIPLES OF MILITARY WORKPLACE PRIVACY

From the preceding discussion, several principles of privacy in the military generally, and the military workplace specifically, emerge. First, the Bill of Rights applies to service members. This includes the Fourth Amendment’s protections against unreasonable searches and seizures. Although guaranteed by procedural safeguards that sometimes differ from those in the civilian sector, service members’ reasonable expectations of privacy are constitutionally protected. Second, the Fourth Amendment’s privacy protections extend into the federal workplace. This includes the military workplace. If, based on all of the circumstances, service members can demonstrate a subjectively held and objectively reasonable expectation of privacy, warrantless searches and seizures are unreasonable absent exigent circumstances. Third, although necessity, custom, and law merge the roles of

---

found that a defendant had no expectation of privacy against his military supervisor’s search of his government desk. 32 M.J. 614, 615 (A.C.M.R. 1991). In that case, the supervisor’s search occurred only after the defendant’s wife told the supervisor about defendant’s criminal misconduct and the supervisor consulted with Army investigators about searching the defendant’s desk. *Id.* Unfortunately, the court did not explore whether the supervisor’s search was for a law-enforcement or work-related purpose and how that determination would impact its Fourth Amendment analysis.

93. *United States v. Kahan*, 350 F. Supp. 784, 791 (S.D.N.Y. 1972).

94. 64 M.J. 57 (C.A.A.F. 2006).

95. For a discussion of *United States v. Long*, 64 M.J. 57 (C.A.A.F. 2006), see *infra* Section III.C.3. For a discussion of privacy expectations in workplace e-mail, see *infra* Section II.B.

96. *Long*, 64 M.J. at 65.

97. Arcuri, *supra* note 88, at 30 n.45 (quoting Earl Warren, *The Bill of Rights and the Military*, 37 N.Y.U. L. REV. 181, 188 (1962)). In his article, Captain Arcuri argues that military commanders searching for criminal evidence are acting as law-enforcement agents rather than supervisors. Turning the language of *Muniz* on its head, he posits that in such cases “the service member, not the commander, should be in ‘no worse position than his civilian counterpart.’” *Id.* at 30.

supervisor and law enforcement agent in the person of the commander, military officials may not use their roles as supervisors to cloak their actions as law enforcement agents. If an inspection, inventory, or routine workplace intrusion into a constitutionally protected area is related to a criminal investigation, both Fourth Amendment protections and the MRE first require search authorization based on probable cause. As discussed in Part III below, an appreciation for these principles is vital in casting both military communications privacy cases and the DoD's response to those cases in their proper light.

### III. THE FOURTH AMENDMENT AND MILITARY E-MAIL

The Supreme Court has long recognized the need for Fourth Amendment protections to apply to electronic intrusions. In his concurring opinion in *Katz v. United States*, Justice Harlan noted that prior case law upholding the constitutionality of electronic surveillance without physical intrusion was “bad physics as well as bad law.”<sup>98</sup> Whether accompanied by physical intrusion or not, “[e]lectronic surveillance is a search for and seizure of words” subject to Fourth Amendment safeguards.<sup>99</sup>

In 1968, the year after the Supreme Court outlined the requisite procedural safeguards for law enforcement agents seeking to conduct electronic surveillance, Congress incorporated the Supreme Court's heightened standards into the Wiretap Act.<sup>100</sup> These provisions were brought into the modern age of electronic communications by the Electronic Communications Privacy Act of 1986 (ECPA), which included the Stored Communications Act (SCA).<sup>101</sup> Unfortunately, as numerous commentators have noted, they have not been meaningfully updated since then to reflect the technological and social evolution of electronic communication.<sup>102</sup> As a

---

98. 389 U.S. 347, 362 (1967) (Harlan, J., concurring) (discussing *Goldman v. United States*, 316 U.S. 129 (1942)).

99. *United States v. Stuckey*, 10 M.J. 347, 349 (C.M.A. 1981) (citing *Katz v. United States*, 389 U.S. 347 (1967)).

100. See Freiwald, *supra* note 22, at 50–53.

101. Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848 (codified as amended in scattered sections of 18 U.S.C. (2000)). The Stored Communications Act (SCA) safeguards electronic communications by prohibiting unauthorized access, restricting disclosure, and permitting a “governmental entity” to compel disclosure of certain electronic communications. *Warshak v. United States (Warshak II)*, 532 F.3d 521, 523 (6th Cir. 2008). For a discussion of the SCA, see *infra* Section III.A.3.

102. See, e.g., Freiwald, *supra* note 22, at 15–17 (noting the SCA was passed when “networked computing was in its infancy”); Orin S. Kerr, *Lifting the “Fog” of Internet Surveillance: How a Suppression Remedy Would Change Computer Crime Law*, 54 HASTINGS L.J. 805, 814 (2003) (discussing ECPA and noting that although amended repeatedly, “all subsequent changes have merely nibbled around the edges of the law that Congress passed with

result, stored electronic communications are protected, at most, by a warrant based on probable cause, rather than the heightened protections of in-transit communication contemplated by *Katz* and the original Wiretap Act.<sup>103</sup>

This Part, following discussion of the first principles of military privacy in Part II, examines precisely how courts—especially military courts—have applied the Fourth Amendment to the “search for and seizure of words” in electronic communications.<sup>104</sup> To lay the necessary foundation, this Part begins with the procedural safeguards outlined in *Katz*. It then moves to brief considerations of the Wiretap Act and the SCA. Finally, this Part reviews the four guidepost communications privacy cases considered thus far by the CAAF and explores whether and under what circumstances service members enjoy the limited workplace privacy expectations outlined by the Supreme Court in *O'Connor*. Particular emphasis is placed on *United States v. Larson*, the most recent computer privacy case decided by the CAAF. This Part argues that it is both factually and analytically distinguishable from prior communications privacy case law.

A. THE FOURTH AMENDMENT AND THE FEDERAL STATUTORY SCHEME FOR ELECTRONIC COMMUNICATIONS PRIVACY

1. *Katz v. United States*

In *Katz v. United States*, the Supreme Court considered whether the Fourth Amendment protects a person’s privacy interest in telephone calls made from a public telephone booth.<sup>105</sup> The petitioner, Mr. Katz, was convicted at trial of communicating wagering information over a telephone in violation of a federal statute. Part of the evidence introduced against him

---

tremendous foresight back in 1986”); Achal Oza, Note, *Amend the ECPA: Fourth Amendment Protection Erodes as E-Mails Get Dusty*, 88 B.U.L. REV. 1043, 1045–46 (2008) (arguing the 180-day distinction in the SCA, which requires probable cause for e-mails stored on third-party servers less than 180 days but not for e-mails stored longer than 180 days, reflects twenty-year-old technology in which e-mails were downloaded to personal computers rather than maintained online).

103. Oza, *supra* note 102, at 1045–46. The argument that stored electronic communications should receive the same heightened safeguards as intercepted telephone calls is based on the normative valuation that electronic communications are as “vital” to private communications today as the telephone was in 1967 when the Supreme Court decided *Katz*, 389 U.S. at 348. Unlike telephone calls, however, electronic communications are generally stored during transmission. Given the normative parity of telephonic and electronic communications, it is difficult to understand why technological distinctions should warrant different statutory protections.

104. *United States v. Stuckey*, 10 M.J. 347, 349 (C.M.A. 1981) (citing *Katz v. United States*, 389 U.S. 347 (1967)).

105. 389 U.S. at 348.

included telephone calls recorded by FBI agents with a device surreptitiously placed on the outside of the telephone booth Mr. Katz used. At trial, Mr. Katz unsuccessfully moved to suppress these telephone call recordings, arguing they constituted a violation of his Fourth Amendment privacy expectations.<sup>106</sup> On appeal, the Supreme Court discussed the “vital role that the public telephone has come to play in private communication” and found that Mr. Katz’ expectation of privacy in his telephone calls was reasonable.<sup>107</sup> The Court held that the government, which had not sought a search warrant, failed to adhere to appropriate procedural safeguards prior to monitoring Mr. Katz’ call:

They were not required, before commencing the search, to present their estimate of probable cause for detached scrutiny by a neutral magistrate. They were not compelled, during the conduct of the search itself, to observe precise limits established in advance by a specific court order. Nor were they directed, after the search had been completed, to notify the authorizing magistrate in detail of all that had been seized.<sup>108</sup>

Importantly, the safeguards outlined by the Court in *Katz* included not only an antecedent probable cause review by a detached magistrate, which would be required for any Fourth Amendment search, but also an articulation of the “precise limits” of the search in a “specific court order,” as well as a detailed, post-search notification to the magistrate of “all that had been seized.”<sup>109</sup> These heightened safeguards having been ignored, the Court overturned Mr. Katz’ conviction.<sup>110</sup>

## 2. *The Wiretap Act*

Congress responded to *Katz* the following year by passing the Wiretap Act as part of the Omnibus Crime Control and Safe Streets Act of 1968.<sup>111</sup> As Congress noted, a primary purpose of the Wiretap Act was the protection of individual privacy.<sup>112</sup> To achieve that end, Congress determined that that non-consensual interception of wire or oral communications “should be allowed only when authorized by a court of competent jurisdiction and

---

106. *Id.*

107. *Id.* at 352.

108. *Id.* at 356.

109. *Id.* The foundation for these standards had been laid in a case issued earlier during the 1967 term, *Berger v. New York*, 388 U.S. 41 (1967). For a discussion of *Berger* and its implications for military communications privacy, see *infra* text accompanying notes 309–14.

110. *Katz*, 389 U.S. at 359.

111. Omnibus Crime Control and Safe Streets Act of 1968, Pub. L. No. 90-351, 82 Stat. 197 (codified as amended at 18 U.S.C. §§ 2510–2522).

112. *Id.* § 801(d), 82 Stat. at 211.

should remain under the control and supervision of the authorizing court.”<sup>113</sup> Congress specifically recognized that communications carriers would occasionally need to intercept communications for work-related purposes, but prohibited common carriers from “service observing or random monitoring except for mechanical or service quality control checks.”<sup>114</sup> The importance of this prohibition is that it acknowledges an individual privacy interest even when that interest is occasionally invaded for routine, work-related purposes. Put another way, in the Wiretap Act, Congress determined, just as the Supreme Court had in *Katz*, that an individual making a telephone call retains a reasonable privacy expectation in that call despite knowing it may be intercepted by the communications carrier for non-law enforcement purposes, which is a foundational principle of judicial decision-making in later communications privacy cases.<sup>115</sup>

To protect an individual’s privacy interest, Congress incorporated into the Wiretap Act the prescriptive guidelines articulated by the Supreme Court in *Katz*, including: (1) an antecedent court order based on probable cause and supported by a detailed oath or affirmation; (2) a description in the order of the person to be surveilled, the communications technology involved, the probable offenses, the identity of the agency and agent, and the period of time (no longer than thirty days) authorized for surveillance; and (3) based on judicial discretion, subsequent reports to the issuing judge regarding the progress and fruits of the surveillance.<sup>116</sup> In exigent circumstances, Congress declared immediate surveillance permissible, provided law enforcement agents seek and obtain a warrant within forty-eight hours.<sup>117</sup> Congress also provided a statutory exclusionary remedy for communications obtained in violation of the Wiretap Act.<sup>118</sup>

---

113. *Id.*

114. 18 U.S.C. § 2511(2)(a) (2006).

115. Recent case law also recognizes that a third-party provider’s ability to access private content does not per se destroy a reasonable expectation of privacy. *See* *Quon v. Arch Wireless Operating Co.*, 529 F.3d 892, 905–06 (9th Cir. 2008) (holding that plaintiff had Fourth Amendment expectation of privacy in text messages even though service provider “may have been able to access the contents of the messages for its own purposes”). Still, some degree of uncertainty regarding a user’s ability to maintain privacy expectations in electronic information disclosed to third-party service providers remains. *See* Orin S. Kerr, *A User’s Guide to the Stored Communications Act, and a Legislator’s Guide to Amending It*, 72 GEO. WASH. L. REV. 1208, 1211 (2004) (discussing judicial application of “disclosure” doctrine to information maintained for users by Internet Service Providers (ISPs)); *see also infra* text accompanying notes 135–36.

116. 18 U.S.C. §§ 2511–2522 (2006).

117. 18 U.S.C. § 2518 (2006).

118. 18 U.S.C. §§ 2511–2522.

3. *The Electronic Communications Privacy Act and Stored Communications Act*

In 1986, Congress passed the Electronic Communications Privacy Act (ECPA) to extend federal statutory protections to “electronic communications.”<sup>119</sup> Among other things, the ECPA added the term “electronic communications” to the Wiretap Act, imposing the same heightened warrant requirements for surveillance of real-time, in-transit electronic communications as for traditional telephone calls.<sup>120</sup> As part of the ECPA, Congress also enacted the SCA, which establishes a two-tiered system for government access to electronic communications stored by third parties based on the duration of storage.<sup>121</sup> When an electronic communication has been stored 180 days or less, the government must seek a warrant based on probable cause.<sup>122</sup> When an electronic communication has been stored more than 180 days, the government may (a) provide notice to the user and then obtain the communication pursuant to a subpoena or court order, or (b) elect not to provide notice to the user and obtain the communications pursuant to a warrant based on probable cause.<sup>123</sup> Unlike the original Wiretap Act, Congress did not include a statutory exclusionary remedy for in-transit or stored electronic communications obtained in violation of the ECPA.<sup>124</sup>

---

119. Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848 (1986) (codified as amended in scattered sections of U.S.C.). For in-depth treatment of the ECPA and Fourth Amendment privacy expectations, see Deirdre K. Mulligan, *Reasonable Expectations in Electronic Communications: A Critical Perspective on the Electronic Communications Privacy Act*, 72 GEO. WASH. L. REV. 1557 (2004). The statutory scheme the ECPA amended is composed of three statutes: the Wiretap Act, 18 U.S.C. §§ 2511–2522, the Pen Register Statute, 18 U.S.C. §§ 3121–3127, and the Stored Communications Act, 18 U.S.C. §§ 2701–2711. Kerr, *supra* note 102, at 815.

120. 18 U.S.C. §§ 2511–2522 (2006); see Freiwald, *supra* note 22, at 13 (“With a few exceptions, the ECPA merely added the word ‘electronic communication’ to every instance of ‘wire communication’ in the statute.”). Unfortunately, these heightened protections did not include stored communications. See *infra* note 125 and accompanying text.

121. 18 U.S.C. §§ 2701–2711 (2006).

122. 18 U.S.C. § 2703(a).

123. 18 U.S.C. § 2703(b).

124. Compare 18 U.S.C. § 2518(10) (2006) (“Any aggrieved person . . . may move to suppress the contents of any wire or oral communication intercepted pursuant to this chapter.”), with 18 U.S.C. § 2708 (2006) (“The remedies and sanctions described in this chapter are the only judicial remedies and sanctions for nonconstitutional violations of this chapter.”). Although violators are subject to civil and criminal penalties, the failure to include a statutory exclusionary remedy means that evidence obtained in violation of the SCA may still be used against a defendant at trial absent a constitutional violation. See *United States v. Kennedy*, 81 F. Supp. 2d 1103, 1110 (D. Kan. 2000). The court noted that:

[t]he statute specifically allows for civil damages and criminal punishment for violations of the ECPA, but speaks nothing about the suppression of information in a court proceeding. Instead, Congress clearly intended for

Because of the ECPA's distinction between in-transit and stored communications, electronic communications occurring in real time receive (with the exception of the suppression remedy) the heightened statutory protections provided in the original Wiretap Act, while those which are stored for more than 180 days do not.<sup>125</sup> As applied to e-mail, this distinction can have the perverse effect of making it more difficult for law enforcement authorities to surveil and seize a single in-transit e-mail than thousands of stored e-mails, particularly those maintained in electronic storage more than 180 days, for which a probable cause warrant is not required.<sup>126</sup> As a result of this deficiency, some commentators have characterized the SCA's protections as "anemic."<sup>127</sup> They argue that the SCA, written more than twenty years ago, fails to recognize the reliance contemporary computer users place on networked access to e-mails, documents, and photographs stored on third-party servers.<sup>128</sup> When the SCA was enacted, most computer users downloaded e-mails and any attached files directly to their personal

---

suppression not to be an option for a defendant whose electronic communications have been intercepted in violation of the ECPA.

*Id.*; see also Kerr, *supra* note 115, at 824–25 (discussing *Kennedy* and arguing for a statutory exclusionary remedy for Internet surveillance).

125. See *United States v. Monroe*, 52 M.J. 326, 330–31 (C.A.A.F. 2000) ("Congress has differentiated between the treatment of the intercept of electronic communications, which is governed by the stringent requirements of 18 USC §§ 2511–2521 (1999), and the access to stored electronic communications, which is governed by the less restrictive provisions of 18 USC §§ 2701–2707 (1999).") (citing David B. Walker, *Privacy in the Digital Age: Encryption Policy—A Call for Congressional Action*, 1999 STAN. TECH. L. REV. 3, 3–12 (1999)). Kerr observes that "as a communication travels across the Internet, different laws apply to it at different times. For example, an e-mail message will be protected by the Wiretap Act when in transit, but by the SCA when it is stored." Kerr, *supra* note 115, at 1231 ("While the SCA protects the privacy of stored Internet communications, the Wiretap Act and Pen Register statute protect the privacy of Internet communications in transit.").

126. Unlike traditional telephone calls, which occur synchronously and generally are not recorded between private parties, e-mails are exchanged asynchronously and, by design, are stored until purposefully or routinely deleted. Even then, copies of an e-mail may be maintained on the computer of the sender or recipient, on the network of a third-party intermediary, or as a hidden file on the computer of the person who deleted the e-mail. See Freiwald, *supra* note 22, at 14 (2007); see also Kerr, *supra* note 115, at 1232 ("Because the Wiretap Act requires the government to obtain a 'super' search warrant rather than the usual warrant required by the SCA, law enforcement agents have an incentive to try to do prospective surveillance normally undertaken under the Wiretap Act using the retrospective authority of the SCA." (citation omitted)).

127. James X. Dempsey, *Communications Privacy in the Digital Age: Revitalizing the Federal Wiretap Laws to Enhance Privacy*, 8 ALB. L.J. SCI. & TECH. 65, 85–89 (1997); Freiwald, *supra* note 22, at 16; see also Mulligan, *supra* note 119, at 1571–76; Oza, *supra* note 102, at 1045–46.

128. See, e.g., David A. Couillard, Note, *Defogging the Cloud: Applying Fourth Amendment Principles to Evolving Privacy Expectations in Cloud Computing*, 93 MINN. L. REV. 2205 (2009) (discussing ECPA's failure to protect communications exchanged via cloud computing).

computers, leaving little, if any, stored communications on third-party servers.<sup>129</sup> Today, however, users frequently rely on third-party services such as Gmail, Yahoo!, Hotmail, and MobileMe to store and retrieve their electronic communications. In such a world, it makes little sense for in-transit e-mails to receive heightened statutory protections, downloaded e-mails stored on a home computer to receive basic Fourth Amendment protections, and web-based e-mails stored on a third-party server to receive less than probable cause statutory protections, especially when the method of communication (e-mail) invokes similar expectations of individual privacy.<sup>130</sup> The Supreme Court's normative conclusion that reasonable privacy expectations are those that "society is prepared to recognize as 'reasonable'"<sup>131</sup> calls for Congress to amend the ECPA and for courts to consider seriously Fourth Amendment privacy rights in stored e-mail.<sup>132</sup>

#### 4. *Fourth Amendment Challenges to the SCA*

Arguably, the competing standards for seizure of substantively similar electronic communications and the absence of a statutory exclusionary remedy should have encouraged, rather than discouraged, both statutory and constitutional challenges. Given that more than twenty years have passed since enactment of the ECPA, one would be justified in assuming the courts have developed a set of analytical tools for reviewing both the federal statutory scheme and its Fourth Amendment implications. As both courts and commentators have repeatedly noted, however, the tool chest remains

---

129. Oza, *supra* note 102, at 1045, 1052–54 (discussing the varying protections afforded e-mail based on the manner in which it is received, including Post Office Protocol (POP), Internet Message Access Protocol (IMAP), and web-based e-mail).

130. Under the federal statutory scheme, telephone calls and real-time electronic communications retain the heightened protections of the original Wiretap Act, while e-mails stored on third-party servers receive probable cause protection or less depending on the duration of storage. E-mails downloaded to a user's personal computer and deleted from the third-party service provider's server are, like other items of personal possession, subject to routine Fourth Amendment analysis. See Kerr, *supra* note 115, at 1232; Mulligan, *supra* note 119, at 1571–72.

131. *Katz v. United States*, 389 U.S. 347, 361 (1967) (Harlan, J., concurring).

132. See Freiwald, *supra* note 22, at 16 (2007) (arguing for abandonment of *Katz* inquiry as analytical model in e-mail privacy cases in favor of Fourth Amendment values espoused in silent video surveillance cases); Kerr, *supra* note 115, at 1233–42 (proposing numerous changes to SCA); Oza, *supra* note 102, at 1045, 1069–70 (proposing removal of the 180-day distinction in the ECPA); Katherine A. Oyama, Note, *E-Mail Privacy After United States v. Councilman: Legislative Options for Amending ECPA*, 21 BERKELEY TECH. L.J. 499, 516–25 (2006) (recommending that Congress reconsider the relationship between ECPA's two-tier framework and modern expectations of e-mail privacy); Scott A. Sundstrom, Note, *You've Got Mail! (And the Government Knows It): Applying the Fourth Amendment to Workplace E-Mail Monitoring*, 73 N.Y.U. L. REV. 2064, 2102 (1998) (urging the courts to apply the Fourth Amendment to e-mail communications).

relatively bare.<sup>133</sup> As a result, communications surveillance case law is famously unsettled, both regarding the constitutionality of the SCA and the privacy interests computer users have in e-mail stored on third-party servers.

A recent Sixth Circuit case, *Warshak v. United States*, illustrates this phenomenon.<sup>134</sup> In *Warshak*, the plaintiff sought declaratory and injunctive relief from the government after he discovered the government had seized personal e-mails from his Yahoo! and local ISP accounts pursuant to the SCA, alleging a violation of his Fourth Amendment rights.<sup>135</sup> On appeal from a district court opinion which held the SCA unconstitutional because it provided for disclosure on less than probable cause, the Sixth Circuit initially ruled that Mr. Warshak retained a reasonable expectation of privacy in his e-mails, even in light of the fact that Mr. Warshak's ISP could, if it chose to do so, review the content of his e-mails.<sup>136</sup>

On rehearing en banc, the Sixth Circuit vacated its earlier decision on grounds of ripeness and standing, holding that Mr. Warshak's request required an inappropriate facial determination of the SCA's constitutionality.<sup>137</sup> The court left open the door to a later constitutional challenge, however, noting that "Mr. Warshak still retains the right to challenge the district court's resolution of his motion to suppress through an appeal of his criminal conviction."<sup>138</sup> While declining to resolve the constitutional issue, the court did foreshadow the complexity of the SCA's constitutional landscape by discussing the difficulty in determining which e-

---

133. See, e.g., *Quon v. Arch Wireless Operating Co.*, 529 F.3d 892, 904 (9th Cir. 2008) ("The extent to which the Fourth Amendment provides protection for the contents of electronic communications in the Internet age is an open question. The recently minted standard of electronic communication via e-mails, text messages, and other means opens a new frontier in Fourth Amendment jurisprudence that has been little explored."); *Freiwald*, *supra* note 22, at 15 (arguing dearth of case law applying Fourth Amendment to ECPA arises from difficulty in applying *Katz*'s two-part inquiry to electronic communication); *Kerr*, *supra* note 102, at 824 (arguing lack of case law interpreting federal surveillance scheme stems from Congress' failure to provide suppression remedy, which in turn discourages criminal defendants from raising challenges).

134. 532 F.3d 521 (6th Cir. 2008).

135. *Id.* at 524.

136. *Warshak v. United States (Warshak I)*, 490 F.3d 455, 473 (6th Cir. 2007) ("[I]ndividuals maintain a reasonable expectation of privacy in e-mails that are stored with, or sent or received through, a commercial ISP. The content of e-mail is something that the user 'seeks to preserve as private,' and therefore 'may be constitutionally protected.' " (quoting *Katz v. United States*, 389 U.S. 347, 351 (1967))). The court's decision in *Warshak I* parallels Congress' determination in the original Wiretap Act that telephone users retain an expectation of privacy despite monitoring for "mechanical or service quality checks." 18 U.S.C. § 2511(2)(a)(i) (1968); see *supra* text accompanying note 123.

137. *Warshak*, 532 F.3d at 526.

138. *Id.* at 534.

mails fell within the SCA's 180-day time period, the varying privacy expectations computer users may have based on their ISP service agreements, and the as-of-yet-unanswered question of how much privacy society is prepared to recognize in computer users' e-mail.<sup>139</sup>

Other courts have applied Fourth Amendment analysis to e-mails stored on third-party servers without specifically confronting the constitutionality of the SCA. In *United States v. Hart*, the Western District of Kentucky considered the defendant's constitutional argument to suppress e-mails obtained by the government from Yahoo! without full compliance with the SCA.<sup>140</sup> Applying the two-part *Katz* inquiry, the district court found that the defendant had agreed to the Yahoo! Terms of Service, which authorized disclosure of "content" as required by law, and therefore "did not show that he sought to preserve . . . his 'Content' as private."<sup>141</sup> Citing *Warshak*, the district court equated "content" with e-mail "messages" and declined to find a constitutional violation. However, the district court limited its holding to the facts of the case, stating "[w]hether or not society is prepared to recognize as reasonable an expectation of privacy in all e-mail communications, the evidence in the record does not show that the defendant sought to preserve as private that which the plaintiff now seeks to introduce into evidence."<sup>142</sup>

Courts elsewhere have issued a range of opinions about e-mail privacy, declining in some cases to find a reasonable expectation of privacy in Internet subscriber information, IP addresses, and to/from address information,<sup>143</sup> while finding in other cases a reasonable expectation of

---

139. *Id.*

140. *United States v. Hart*, No. 08-109-C, 2009 U.S. Dist. LEXIS 72473 (W.D. Ky. Aug. 17, 2009). The government failed to provide proper notice to the defendant as required by the SCA under certain circumstances. *See supra* text accompanying note 121.

141. *Hart*, 2009 U.S. Dist. LEXIS 72473, at \*7. Other courts have relied on contract law to find support for a user's reasonable expectations of privacy in stored e-mail. *See United States v. Maxwell*, 45 M.J. 406, 417 (C.A.A.F. 2006) (finding "[American Online's] contractual obligations with appellant insured him privacy"); *see also* Stephen R. Stewart, *Katy Bar the Door—2006 New Developments in Fourth Amendment Search and Seizure Law*, 2007 ARMY LAW. 1, 13 (2007).

142. *Hart*, 2009 U.S. Dist. LEXIS 72473, at \*7–\*8.

143. *See United States v. Forrester*, 512 F.3d 500, 510 (9th Cir. 2007) (finding "users have no expectation of privacy in to/from addresses of their messages or the IP addresses of the websites they visit because they should know that this information is provided to and used by Internet service providers for the specific purpose of directing the routing of information"); *United States v. Kennedy*, 81 F. Supp. 2d 1103, 1110 (D. Kan. 2000) (finding no reasonable expectation of privacy in Internet "subscriber" information provided by ISP to law enforcement agents pursuant to SCA); *United States v. Ohnesorge*, 60 M.J. 946, 949 (N-M. Ct. Crim. App. 2005) (finding no reasonable privacy expectation in internet "subscriber" information on government computer).

privacy in pager text messages and e-mails.<sup>144</sup> None have yet held the SCA itself unconstitutional.<sup>145</sup>

#### B. E-MAIL PRIVACY IN THE PUBLIC WORKPLACE

Like e-mail privacy generally, e-mail privacy in the public workplace remains an unsettled area of the law, with courts disagreeing about the extent to which employees can retain a reasonable expectation of privacy in e-mails sent over communications systems provided by the government for conducting official business.<sup>146</sup> Two approaches are immediately obvious. First, one might argue that the government provides communications systems for official use only. Because all communications sent over the network are, at least notionally, for official use, users lose any reasonable expectation of privacy.<sup>147</sup> Second, one might argue that the provision of a communications system is no different than the provision of a government desk. It is the actual practice of the government, rather than its general policy of official use, which supports or weakens a limited expectation of privacy. In fact, this latter argument would accord most closely with the Supreme Court's holding in *O'Connor*, which stands for the proposition that employees may possess limited Fourth Amendment privacy rights in the workplace based "on all the circumstances."<sup>148</sup>

Certainly legitimate reasons exist for the government, like other employers, to monitor the online activity of its employees.<sup>149</sup> Public

144. See *Quon v. Arch Wireless Operating Co.*, 529 F.3d 892, 903, 909 (9th Cir. 2008) (finding civil violation of both SCA and Fourth Amendment as pertaining to text messages stored by third-party service provider and obtained by plaintiff's public employer); *United States v. Maxwell*, 45 M.J. 406, 419 (C.A.A.F. 1996) (finding expectation of privacy in e-mails stored on AOL server).

145. See *Warshak v. United States*, 532 F.3d 521, 531 ("The Stored Communications Act has been in existence since 1986 and to our knowledge has not been the subject of any successful Fourth Amendment challenges . . .").

146. For a recent and thorough discussion of online privacy in the workplace generally, see Robert Sprague, *Orwell Was an Optimist: The Evolution of Privacy in the United States and Its De-Evolution for American Employees*, 42 J. MARSHALL L. REV. 83 (2008); see also Micah Echols, *Striking a Balance Between Employer Business Interests and Employee Privacy: Using Respondeat Superior to Justify the Monitoring of Web-Based, Personal Electronic Mail Accounts of Employees in the Workplace*, 7 COMP. L. REV. & TECH. J. 273 (2003); Justin Conforti, Comment, *Somebody's Watching Me: Workplace Privacy Interests, Technology Surveillance, and the Ninth Circuit's Misapplication of the Ortega Test in Quon v. Arch Wireless*, 5 SETON HALL CIR. REV. 461 (2009); Sundstrom, *supra* note 132, at 2102; DEP'T OF JUSTICE, SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS (3d ed.).

147. Even then the slope is slippery, since, in the military at least, "official" and "authorized" use includes limited personal use. See *supra* text accompanying note 11.

148. *O'Connor v. Ortega*, 480 U.S. 709, 726 (1987) (plurality opinion).

149. In the private sector, seventy-seven percent of major U.S. employers reportedly

employers have an interest in ensuring employee productivity, data security, patent and trademark protection, and quality customer relations.<sup>150</sup> They also have an interest in preserving evidence for potential litigation, minimizing behavior that may create a hostile work environment, and ensuring their communications systems are secure, stable, and free of harmful viruses and programs.<sup>151</sup> Finally, with respect to certain computer systems, government agencies must take precautions to ensure their networks are protected from internal and external attacks that would endanger local, state, or national security.<sup>152</sup>

Rather than justifying a blanket exception to Fourth Amendment workplace privacy rights, however, such reasons appear to be yet another variety of the “efficient and proper operation of the workplace” needs the Supreme Court held in *O’Connor* to justify warrantless work-related searches.<sup>153</sup> If so, employees do not have a reasonable expectation of privacy against monitoring for work-related, non-investigatory purposes. Neither do they have a reasonable expectation of privacy against searches related to the investigation of work-related misconduct, provided the search is both “justified at its inception” and “reasonably related in scope.”<sup>154</sup> Whether they

---

monitor employee communications. Echols, *supra* note 146, at 278; *see also* Press Release, Am. Mgmt. Inst., 2007 Electronic Monitoring & Surveillance Survey: Over Half of All Employers Combined Fire Workers for E-Mail & Internet Abuse (Feb. 28, 2008), <http://press.amanet.org/press-releases/177/2007-electronic-monitoring-surveillance-survey/> (last visited Sept. 1, 2009) (finding 66% of employers monitor internet connections, 65% use internet blocking software, and 43% monitor e-mail).

150. *See* Edell, *supra* note 9, at 2 (discussing justifications for employer monitoring); Sprague, *supra* note 146, at 111 (same).

151. These concerns are no different than those expressed by private employers. *See* Echols, *supra* note 146, at 278 (listing legal liability, company security, and productivity as three top reasons employers monitor employee communications).

152. *See* LeEllen Coacher, *Permitting Systems Monitoring: When the Government Can Look and What It Can See*, 46 A.F. L. REV. 155, 155–56 (1999) (discussing monitoring for operational security, law enforcement, and systems protection); Edell, *supra* note 9, at 2–3 (discussing protection of national security assets as a justification for monitoring); Kastenberg, *supra* note 11, at 181 (discussing actions by state and non-state actors to exfiltrate data from government information systems); Joshua E. Kastenberg, *Non-Intervention and Neutrality in Cyberspace: An Emerging Principle in the National Practice of International Law*, 64 A.F. L. REV. 43, 45–50 (2009) (discussing emergence of cyber warfare); *see* Erik Holmes, *Even Routine Work Poses Cyber Threat*, AIR FORCE TIMES, May 11, 2009, at 18 (discussing attacks against DoD computer systems and stolen data regarding F-35 Joint Strike Fighter); William H. McMichael & Bruce Rolfsen, *Despite Network Virus—Avoid Thumb Drives*, AIR FORCE TIMES, Dec. 8, 2008, at 13 (reporting on Air Force directives ordering computer users from using thumb drives due to virus concerns).

153. *O’Connor v. Ortega*, 480 U.S. 709, 723 (1987) (plurality opinion).

154. *Id.* at 726.

have an expectation of privacy for investigations related to criminal misconduct, however, remains largely undecided.<sup>155</sup>

In *United States v. Simons*, the Fourth Circuit considered whether an employee of the Foreign Bureau of Information Services (FBIS), a division of the Central Intelligence Agency (CIA), held a reasonable expectation of privacy in his workplace online activity and workplace computer.<sup>156</sup> During routine systems monitoring for inappropriate uses, it was discovered that Mr. Simons had downloaded child pornography to his workplace computer, which was located in his private office.<sup>157</sup> FBIS remotely examined and copied the contents of Mr. Simons' computer. After the files were viewed by representatives from the CIA Office of Inspector General (OIG), "one of whom was a criminal investigator," FBIS personnel entered Mr. Simons office, removed his hard drive, and replaced it with a copy.<sup>158</sup> Evidence from the hard drive then was used to obtain a search warrant, which eventually led to two searches of Mr. Simons' office and his subsequent arrest.<sup>159</sup>

Following his conviction, Mr. Simons argued on appeal the FBIS had violated his Fourth Amendment privacy rights.<sup>160</sup> The circuit court first concluded that "the remote searches of Simons' computer did not violate his *Fourth Amendment* rights because, in light of the [FBIS] Internet policy, Simons lacked a legitimate expectation of privacy in the files downloaded from the Internet."<sup>161</sup> The FBIS internet policy specifically notified FBIS employees that "FBIS would 'audit, inspect, and/or monitor' employees' use of the Internet, including all file transfers, all websites visited, and all e-mail messages."<sup>162</sup> Thus, even if Mr. Simons held a subjective belief in the privacy of his online activity, his belief was not reasonable. The circuit court then turned to the initial warrantless entry into Mr. Simons' office. Finding Mr. Simons did have a reasonable expectation of privacy in his office, the court concluded the search fell within the *O'Connor* "work-related" exception.<sup>163</sup> In reaching that conclusion, the court assumed the "dominant purposes of the

---

155. Compare *Leventhal v. Knapek*, 266 F.3d 64, 74 (2d Cir. 2001) (holding state agency employee had reasonable expectation of privacy in workplace computer based on employer's infrequent access), with *United States v. Simons*, 206 F.3d 392, 398 (4th Cir. 2000) (holding federal employee had no reasonable expectation of privacy in files downloaded from Internet when employer had published monitoring policy).

156. 206 F.3d 392 (4th Cir. 2000).

157. *Id.* at 396.

158. *Id.*

159. *Id.* at 395–97.

160. *Id.* at 398.

161. *United States v. Simons*, 206 F.3d 392, 398 (4th Cir. 2000).

162. *Id.*

163. *Id.* at 400.

warrantless search . . . was to acquire evidence of criminal activity.”<sup>164</sup> Nevertheless, “FBIS did not lose its special need for ‘the efficient and proper operation of the workplace’ ” simply because a dual purpose of the search was criminal.<sup>165</sup> Thus, the court found the warrantless entry, which was reasonable in its inception and scope, did not violate Mr. Simons’ Fourth Amendment privacy rights.<sup>166</sup>

In *United States v. Angevine*, the Tenth Circuit reached a similar conclusion when it determined a professor at a public university did not have a reasonable expectation of privacy in his university computer from local law enforcement officials.<sup>167</sup> In that case, law enforcement officers obtained a search warrant to seize a work computer belonging to Professor Angevine, who was suspected of possessing child pornography.<sup>168</sup> Numerous images of child pornography were found. At trial, Professor Angevine requested a hearing pursuant to *Franks v. Delaware*<sup>169</sup> to challenge the search warrant.<sup>170</sup> Both the trial and appellate courts denied Professor Angevine’s motion because the university’s computer policy and logon banner prevented finding a reasonable expectation of privacy, thereby making the search warrant unnecessary.<sup>171</sup> Although the circuit court did not expressly discuss the distinction between work-related and law enforcement searches, the court did note the University policy “reserved the right to randomly audit Internet use and to monitor specific individuals suspected of misusing University computers [and] . . . explicitly warned employees that legal action would result from violations of federal law.”<sup>172</sup> The court also found significant the fact that the logon “splash screen” warned users of “ ‘criminal penalties’ for misuse . . . .”<sup>173</sup>

As discussed below, military case law in these areas is still developing. Were courts to apply *Simons* and *Angevine* to the military workplace, however, both could have substantive implications. For military commanders, who are authorized to simultaneously act as supervisors and law enforcement agents, *Simons* could be relied upon to argue that military authorities may conduct a warrantless search of an otherwise protected workplace area if the purpose of the search—which may primarily be criminal in nature—is at least *partly*

---

164. *Id.*

165. *Id.*

166. *Id.* at 401.

167. 281 F.3d 1130 (10th Cir. 2002).

168. *Id.* at 1132.

169. 438 U.S. 154 (1978).

170. 281 F.3d at 1130.

171. *Id.* at 1135.

172. *Id.* at 1134.

173. *Id.*

work-related. For military criminal investigators, *Angevine* could be relied upon to argue that warrantless law enforcement searches of government computers are permissible as long as users receive notice of such searches through an agency-wide policy or logon banner. Unfortunately, as pertaining to searches of government e-mail accounts, neither case is precisely on point. In *Simons*, the initial electronic evidence implicating Mr. Simons was discovered as the result of routine systems monitoring, not a criminal investigation,<sup>174</sup> and in *Angevine* e-mail was not at issue.<sup>175</sup>

### C. E-MAIL PRIVACY IN THE MILITARY WORKPLACE

Considering the unsettled state of civilian case law, one might reasonably expect military cases to reflect a similar schizophrenia. One might also expect, given the skepticism some critics have for the military's protection of individual service members' liberties, that military courts confronting communications privacy in the context of the Fourth Amendment would have approached the issue cautiously, erring on the side of the military's need to ensure "the efficient and proper operation of the workplace"<sup>176</sup> and only reluctantly finding Fourth Amendment protections for individual service members. On both points, one would be wrong.

In general, military courts have been forward leaning in both considering and finding Fourth Amendment privacy expectations in electronic communications, including those maintained privately and in the workplace. In fact, well before any Article III court addressed stored e-mail in the context of the Fourth Amendment, military courts had already ruled on the matter twice.<sup>177</sup> In both cases, which involved law enforcement searches, the CAAF found that service members possessed a reasonable expectation of privacy in their e-mail, and the government had violated service members' Fourth Amendment rights by intruding without a warrant.<sup>178</sup>

The reasons military courts outpaced Article III courts in first addressing privacy protections for stored e-mail may not be immediately apparent, but

---

174. *United States v. Simons*, 206 F.3d 392, 396 (4th Cir. 2000).

175. *Angevine*, 281 F.3d at 1132.

176. *O'Connor v. Ortega*, 480 U.S. 709, 723 (1987) (plurality opinion).

177. *See United States v. Maxwell*, 45 M.J. 406 (C.A.A.F. 1996); *United States v. Long*, 64 M.J. 57 (C.A.A.F. 2006); *see also Freiwald*, *supra* note 22, at 3-4, 4 n.9 (noting in 2007 that "[n]o Article III court has yet established when or even whether users entertain a reasonable expectation of privacy in their e-mails" but that "two military courts have found a reasonable expectation of privacy in stored e-mail and imposed a warrant requirement on government access to it").

178. For a discussion of *Maxwell*, see *infra* Section III.C.1. For a discussion of *Long*, see *infra* Section III.C.3.

several possibilities present themselves. First, as observed earlier, the government assigns qualified counsel without cost to every military defendant in a special or general court-martial, regardless of the indigent circumstances of the accused.<sup>179</sup> Each of these attorneys is licensed to practice law in at least one state, is a graduate of both an American Bar Association (ABA) accredited law school<sup>180</sup> and a service-specific Judge Advocate training course,<sup>181</sup> and is the recipient of extensive and continuing training in trial tactics, trial procedures, and changes in law.<sup>182</sup> Rather than having the enormous caseload commonly carried by public defenders, military defense counsel typically have only a handful of cases, enabling them to spend more time on each case. Additionally, for serious and complex cases, the government may assign military defendants an additional defense attorney free of charge, as well as relevant forensic experts in psychology, pathology, medicine, or computer science, to name a few fields of expertise.<sup>183</sup> The effect of bringing all of these resources to bear in a given case may be that military defense counsel have both the ability and availability to explore and, if warranted, raise novel issues of fact and law for the benefit of their client, a luxury their civilian counterparts may not always enjoy.<sup>184</sup>

---

179. *See supra* text accompanying note 36.

180. The requirement under 10 U.S.C. § 827(b)(1) (2006), is that counsel “must be a judge advocate who is a graduate of an accredited law school or is a member of the bar of a Federal court or of the highest court of a State; or must be a member of the bar of a Federal court or of the highest court of a State.” In the Air Force, for example, the relevant service regulation requires judge advocates to graduate from an ABA accredited law school and “[b]e in good standing and admitted to practice before a Federal court or the highest court of a U.S. state, territory, or the District of Columbia.” U.S. DEP’T OF AIR FORCE, AIR FORCE INSTRUCTION 51-103, DESIGNATION AND CERTIFICATION OF JUDGE ADVOCATES para. 1.1 (2004) [hereinafter AFI 51-103]. The U.S. Army has similar requirements. *See* U.S. DEP’T OF ARMY, ARMY REG. 27-1, JUDGE ADVOCATE LEGAL SERVICES para. 13-2 (1996) [hereinafter ARMY REG. 27-1] (establishing requirements for entry into U.S. Army JAG Corps). The service secretaries have statutory authority to implement the requirements for designation as judge advocates. *See, e.g.*, 10 U.S.C. § 8067(g) (2006) (“Judge advocate functions in the Air Force shall be performed by commissioned officers of the Air Force who are qualified under regulations prescribed by the Secretary, and who are designated as judge advocates.”).

181. For example, the Air Force requires judge advocates to graduate from the Judge Advocate Staff Officer Course prior to being certified as a judge advocate. AFI 51-103, *supra* note 180, at para. 3.1.

182. The Army, Air Force, and Navy each maintain a judge advocate school, which provides accession training, specialty training, and continuing legal education. *See, e.g.*, ARMY REG. 27-1, at para. 10-2 (1996) (discussing the responsibilities of the Commandant of The Judge Advocate General’s School, U.S. Army).

183. Regarding experts, military regulations specifically provide for their employment and payment at government expense. *See, e.g.*, AFI 51-201, *supra* note 9, at para. 6.2.5 (discussing provision of urinalysis experts).

184. The American Bar Association has commented on public defenders’ excessive

Second, civilian law enforcement officials seeking to access a suspect's workplace communications must coordinate with the suspect's employer in accordance with the federal statutory scheme. However, even if disclosure of such communications were to violate the SCA, the lack of a statutory suppression remedy provides little incentive for defense counsel to raise the issue at trial.<sup>185</sup> In the military, however, commanders may search service members' electronic communications without the need to coordinate with, or seek authorization from, a third party.<sup>186</sup> That military commanders act as both supervisors and criminal investigators heightens the possibility that the fruits of their searches will find their way to trial and, as a result, into a motion to suppress. Because such communications may have been seized pursuant to command authority rather than the SCA, defense counsel are free to raise Fourth Amendment concerns without first having to wade through the SCA's omission of statutory suppression.

Third, every defendant convicted at court martial who receives an approved sentence of death, dismissal, dishonorable discharge, bad-conduct discharge, or confinement for more than one year receives an automatic right of appeal, which includes appellate defense counsel and plenary de novo review, all of which are provided at government expense.<sup>187</sup> From the initial

---

caseloads. *See* Eight Guidelines of Public Defense Related to Excessive Workloads, American Bar Association (2009), available at [http://www.abanet.org/legalservices/sclaid/defender/downloads/eight\\_guidelines\\_of\\_public\\_defense.pdf](http://www.abanet.org/legalservices/sclaid/defender/downloads/eight_guidelines_of_public_defense.pdf). Although employed by the military, military defense counsel (MDC) are charged with ethically and vigorously defending their clients regardless of the effect of that representation on the military. The "Fundamental Principles" of the Air Force Military Defense Counsel Charter states:

An MDC's primary responsibility is to his or her client. Constrained only by ethical limits, MDCs are authorized by law to enter into attorney-client relationships and to oppose the government of the United States, in order to promote the individual interests of service members they represent without regard to how their actions might otherwise affect the Air Force as an institution.

AFLSA/JAJD, OPERATING INSTRUCTION 1, AIR FORCE MILITARY DEFENSE COUNSEL CHARTER (2005).

185. *See, e.g.*, *United States v. Kennedy*, 81 F. Supp. 2d 1103, 1110 (D. Kan. 2000) (concluding "Congress clearly intended for suppression not to be an option for a defendant whose electronic communications have been intercepted in violation of the ECPA"); *see also* Kerr, *supra* note 102, at 824 (arguing the lack of case law interpreting the federal surveillance scheme stems from Congress' failure to provide a suppression remedy, which in turn discourages criminal defendants from raising challenges).

186. For a discussion of military commanders' search authority, see *supra* text accompanying notes 44-54. Although a "third party" is not involved, commanders must still ensure their searches comply with the SCA, which prohibits access of stored communications without authorization. *See infra* note 233.

187. *See* 10 U.S.C. § 866(b)-(c) (2006); *see also* *United States v. Roach*, 66 M.J. 410, 412-13 (C.A.A.F. 2008) (discussing three main differences between civilian and military criminal

appellate level, cases also may be appealed to the Court of Appeals for the Armed Forces<sup>188</sup> and ultimately to the Supreme Court.<sup>189</sup> Assigned appellate attorneys—not defense trial counsel—normally represent the defendant on appeal, increasing the likelihood they will catch errors made at trial and raise issues of ineffective assistance of counsel.<sup>190</sup>

The extent to which these factors laid the groundwork for military courts' early entrance into the field of Fourth Amendment communications privacy is open for discussion. What is clear, however, is that military case law, though evolving, has already marked the boundaries for analyzing and determining Fourth Amendment privacy claims in electronic communications in a way that civilian federal courts have not. As discussed below, the guidepost cases are *United States v. Maxwell*,<sup>191</sup> *United States v. Monroe*,<sup>192</sup> *United States v. Long*,<sup>193</sup> and *United States v. Larson*.<sup>194</sup>

1. *United States v. Maxwell—The Fourth Amendment and Commercial E-Mail*

The CAAF first crossed the threshold of Fourth Amendment Internet law in *United States v. Maxwell*, where it addressed whether a U.S. service member possessed a reasonable expectation of privacy in his commercial America Online (AOL) e-mail account.<sup>195</sup> Decided in 1996, the CAAF treated the issue as one of first impression in military law. Reasoning that e-mail transmissions were analogous to traditional communications like mail and telephone calls, the CAAF held a reasonable expectation of privacy in e-mail existed.<sup>196</sup>

Contrary to Colonel Maxwell's pleas, the court martial convicted him of two counts of indecent language, one specification of distributing obscene materials, and one specification of transporting or receiving child pornography in violation of Article 134, UCMJ.<sup>197</sup> The indecent language

---

proceedings: (1) mandatory review; (2) government provided appellate counsel; and (3) de novo review of findings and sentence). Additionally, cases which do not merit mandatory appeal must be reviewed by the office of the Judge Advocate General of the relevant service, who may modify the findings and sentence, refer the case to the appellate courts, or order a rehearing. *See* 10 U.S.C. § 869 (2006).

188. 10 U.S.C. § 867 (2006).

189. 10 U.S.C. § 867(a).

190. *See Gilligan, supra* note 35, at 5.

191. 45 M.J. 406 (C.A.A.F. 1996).

192. 52 M.J. 326 (C.A.A.F. 2000).

193. 64 M.J. 57 (C.A.A.F. 2006).

194. 66 M.J. 212 (C.A.A.F. 2008).

195. 45 M.J. at 416.

196. *Id.* at 410, 417–19.

197. *Id.* at 410.

specifications alleged violations of clauses 1 and 2 of Article 134,<sup>198</sup> which generally prohibit “disorders and neglects to the prejudice of good order and discipline in the armed forces” and “conduct of a nature to bring discredit upon the armed forces,” respectively.<sup>199</sup> The specifications for distributing obscene materials and transporting or receiving child pornography alleged violations of clause 3 of Article 134,<sup>200</sup> which prohibits “crimes and offenses not capital.”<sup>201</sup>

The bases for conviction under Article 134 rested heavily on computer evidence. The conviction of indecent language was principally based on evidence the Air Force Office of Special Investigations (AFOSI) obtained from the Federal Bureau of Investigation (FBI) following execution of a search warrant on AOL. The search, which was directed at Colonel Maxwell’s AOL account “Redde1,” also included e-mails from Colonel Maxwell’s secondary AOL account, “Zirlock.”<sup>202</sup> Among the “Zirlock” e-mails were communications from Colonel Maxwell to another Air Force officer in which Colonel Maxwell discussed his sexual orientation, desires, and preferences. It was these e-mails that led to the conviction for indecent language.<sup>203</sup> Similarly, the conviction for child pornography resulted from seizure of Colonel Maxwell’s on-base personal computer by AFOSI agents, who properly obtained a search warrant from the base military magistrate. The agents found three images depicting minor children, which became the basis for the child pornography conviction.<sup>204</sup>

---

198. *Id.*

199. 10 U.S.C. § 934 (2006). The sweeping language of Article 134 has been upheld despite constitutional criticisms of being “void for vagueness.” *See Parker v. Levy*, 417 U.S. 733, 752–57 (1974) (rejecting “vagueness” claim in light of judicial construction which has “narrow[ed] the . . . literal language of the articles, and at the same time supplying considerable specificity by way of examples of the conduct that they cover”).

200. *United States v. Maxwell*, 45 M.J. 406, 410 (C.A.A.F. 1996).

201. 10 U.S.C. § 934. Although the literal language of Clause 3 is obscure, its meaning in practice is clear. Service members may be punished under Clause 3 for violating federal crimes of unlimited application, such as counterfeiting or various frauds, and crimes of local application, which includes both federal crimes and state crimes adopted as part of the Federal Assimilative Crimes Act (18 U.S.C. § 13). *See* 10 U.S.C. § 934.

202. *Maxwell*, 45 M.J. at 414. In the search warrant, Colonel Maxwell’s AOL account was erroneously listed as “REDDEL” rather than “Redde1” (pronounced “Ready One”), and his alternate account “Zirlock” was not listed at all. AOL collected records for both “Redde1” and “Zirlock,” however, because they belonged to the user. Colonel Maxwell challenged both searches, but was unsuccessful as concerning “Redde1,” which the CAAF held was a minor scrivener’s error. *Id.* at 413, 420.

203. *Id.* at 414.

204. *Id.*

On appeal, defense counsel raised numerous issues of error, ranging from the constitutionality of the applicable federal statutory scheme to the validity of both the AOL and on-base search warrants.<sup>205</sup> Before reaching these issues, however, the CAAF explored as an initial matter whether appellant's expectation of privacy in his AOL e-mail account was reasonable.<sup>206</sup> The CAAF relied heavily on the testimony of AOL's vice president of marketing, who stated that "AOL's policy was not to read or disclose subscribers' e-mail to anyone except authorized users" and "[i]t was AOL's practice to guard these 'private communications' and only disclose them to third parties if given a court order . . . ."<sup>207</sup> From these statements, the CAAF concluded that the "appellant possessed a reasonable expectation of privacy, albeit a limited one, in the e-mail messages that he sent and/or received on AOL."<sup>208</sup>

The CAAF then took an interesting and important turn in its analysis. Acknowledging the varying degrees of Fourth Amendment privacy expectations, the CAAF distinguished between privacy interests in the computer itself and the messages sent or e-mailed from that computer:

We are satisfied that the Constitution requires that the FBI and other police agencies establish probable cause to enter into a personal and private computer. However, when an individual sends or mails letters, messages, or other information on the computer, the Fourth Amendment expectation of privacy diminishes incrementally. Moreover, the more open the method of transmission, such as the 'chat room,' the less privacy one can reasonably expect. This case alone presents a spectrum of privacy expectations . . . .<sup>209</sup>

After acknowledging the "spectrum of privacy expectations," the CAAF analogized to existing technologies. The court noted that in making an initial transmission over e-mail, first class mail, or telephone, a sender has a

---

205. *Id.* at 416.

206. *Id.*

207. *Id.* at 417. In *United States v. Long*, the CAAF similarly relied on network administrator testimony in determining whether the network provider had, through its policies and actions, created a reasonable expectation of privacy in its computers. 64 M.J. 57 (C.A.A.F. 2006); see also *infra* text accompanying note 234.

208. *Maxwell*, 45 M.J. at 417. The CAAF also observed the AOL's system differed from "less secure" e-mail systems on the Internet. *Id.* (citing *American Civil Liberties Union v. Reno*, 929 F. Supp. 824, 830–44 (E.D. Pa. 1996)). Whether the technical differences that led CAAF to draw this distinction still exist is questionable, especially given the evolution from proprietary platforms to web-based e-mail applications. See generally Couillard, *supra* note 128, at 2205 (discussing application of Fourth Amendment privacy expectations to electronic communications stored on Internet cloud).

209. *Maxwell*, 45 M.J. at 417.

reasonable expectation of privacy from unauthorized interceptions by law enforcement authorities. This expectation, however, does not extend to secondary and tertiary transmissions over which the original sender has no control.<sup>210</sup> Having found parallels in other mediums of communication, the CAAF had little difficulty in finding that “the transmitter of an e-mail message enjoys a reasonable expectation that police officials will not intercept the transmission without probable cause and a search warrant.”<sup>211</sup> Accordingly, the CAAF held that the “Zirlock” e-mail messages supporting the indecent language specifications were beyond the scope of the search warrant (which included only “Redde1” e-mails) and thus had to be suppressed.<sup>212</sup>

Although *Maxwell* addressed e-mails sent over commercial, as opposed to government networks, it provides a helpful backdrop in identifying the first principles underscoring the CAAF’s subsequent privacy expectation decisions in *Monroe* and *Long*—decisions which, at first blush, appear to reach opposite conclusions concerning Fourth Amendment privacy expectations in government e-mail. The strength of the CAAF’s approach in *Maxwell*, which the court drew upon in subsequent cases, lies in its willingness to base its privacy expectation analysis on the vital nature of the technology at issue, the methods of storage, retrieval, and transmission, and the distinction between non-law enforcement and law enforcement searches. As the CAAF concluded in *Maxwell*, “[e]xpectations of privacy in e-mail transmissions depend in large part on the type of e-mail involved and the intended recipient.”<sup>213</sup> These distinctions would define the outcomes of later CAAF decisions.

2. United States v. Monroe—*The Fourth Amendment and Government E-Mail Monitoring*

In *United States v. Monroe*, the CAAF’s next milestone communications privacy case, the court considered whether e-mails transmitted over a government network warranted protection from systems monitoring.<sup>214</sup> By drawing a line between routine systems monitoring and purposeful law enforcement activity, the CAAF held the e-mails at issue did not qualify for Fourth Amendment protection.<sup>215</sup>

---

210. *Id.* at 417–18.

211. *Id.* at 418. The possibility of unauthorized interceptions by others, such as hackers, “does not diminish the legitimate expectation of privacy in any way.” *Id.*

212. *Id.* at 424.

213. *Id.* at 418–19.

214. 52 M.J. 326, 330 (C.A.A.F. 2000).

215. *Id.* at 329–31.

In late 1995, during the course of routine systems maintenance, two Air Force systems administrators at Osan Air Base, Korea discovered the local government e-mail server had fifty-nine undeliverable messages addressed to Staff Sergeant Monroe.<sup>216</sup> To learn why the files had not been routed correctly, the e-mails were copied to another computer and opened by the systems administrators, who discovered multiple sexually explicit images.<sup>217</sup> The systems administrators then opened Staff Sergeant Monroe's e-mail account "to find out whether this material had been requested or whether Monroe had been the victim of a prank."<sup>218</sup> After finding an e-mail from Staff Sergeant Monroe requesting a file from the sender of the pornographic images, the systems administrators turned the material over to AFOSI agents, who consulted with the local judge advocate and base military magistrate to obtain probable cause search authorization.<sup>219</sup> AFOSI agents then searched Staff Sergeant Monroe's on-base dormitory room, seized his personal computer, and found several images of child pornography.<sup>220</sup>

At trial, Staff Sergeant Monroe argued his Fourth Amendment rights had been violated and moved to suppress both the initial evidence found by the systems administrators and the evidence the AFOSI later seized.<sup>221</sup> The trial court denied his motion, and Staff Sergeant Monroe entered a conditional guilty plea to preserve the suppression issue on appeal.<sup>222</sup> He was convicted by a general court-martial of violating a general order, possessing child pornography, and transmitting and receiving obscene writings and computer graphics.<sup>223</sup>

On appeal, the CAAF began by comparing the facts at issue in *Monroe* to those at issue in *Maxwell*.<sup>224</sup> The court noted that in *Maxwell* the e-mails were

---

216. *Id.* at 328–29.

217. *Id.*

218. *Id.*

219. *Id.* at 326–29.

220. *Id.* at 329. Although not explicitly stated in the court's opinion, it appears the e-mails had been sent to and from Staff Sergeant Monroe's government e-mail account, which he accessed using his personal computer in his dormitory room. Staff Sergeant Monroe did not have access to government computers at his workplace capable of sending or receiving e-mail. *Id.*

221. *Id.*

222. *Id.*

223. *Id.*

224. *Id.* at 330. At the intermediate appellate level, the Air Force Court of Criminal Appeals had affirmed the military judge's denial of the motion to suppress, finding Staff Sergeant Monroe had no reasonable expectation of privacy in his e-mails "at least as regards his superiors and the [systems] administrator and his/her superiors." *Id.* at 329 (quoting *United States v. Monroe*, 50 M.J. 550, 559 (A.F. Ct. Crim. App. 1999)) (internal quotation marks omitted). The Air Force appellate court also sustained the probable cause search authorization issued by the base magistrate. *Id.*

located on a privately owned AOL server and that AOL had contractually agreed to provide limited subscriber privacy.<sup>225</sup> By contrast, the e-mails in *Monroe* were located on a government-owned server, and the government had a specific notice that “users logging on to this system consent to monitoring by the administrator.”<sup>226</sup> Implicitly applying the *O’Connor* totality of the circumstances test,<sup>227</sup> the court held that “Monroe had no reasonable expectation of privacy in his e-mail messages or e-mail box at least from the personnel charged with maintaining the . . . system.”<sup>228</sup>

The limited nature of the CAAF’s holding is important. By adopting the intermediate appellate court’s reasoning that Staff Sergeant Monroe had no privacy expectation “from the personnel charged with maintaining the . . . system,”<sup>229</sup> the CAAF was able to embrace both its reasoning in *Maxwell* and the work-related/law enforcement distinction underlying federal workplace privacy case law. Citing *Maxwell*, the court explicitly stated, “[t]he transmitter of an e-mail message enjoys a reasonable expectation of privacy that police officials will not intercept the transmission without probable cause and a search warrant.”<sup>230</sup> No such expectation, however, is available as against those tasked with routine systems maintenance.<sup>231</sup>

Having found the systems administrators’ search did not violate the Fourth Amendment, the CAAF next considered whether the systems administrators had disclosed the messages’ contents in violation of the SCA.<sup>232</sup> Noting the distinction Congress had drawn between intercepted and stored electronic communications, the CAAF found, without explanation, “the e-mail messages in this case were accessed from storage and not intercepted in transit,” and therefore the SCA’s disclosure provisions (as opposed to the Wiretap provisions) applied.<sup>233</sup> Under the SCA, the court

---

225. *Id.* at 330.

226. *Id.*

227. *O’Connor v. Ortega*, 480 U.S. 709, 725–26 (1987) (plurality opinion).

228. *Monroe*, 52 M.J. at 330.

229. *Id.*

230. *Id.* (citing *United States v. Maxwell*, 45 M.J. 406, 418 (C.A.A.F. 1996)).

231. *Id.*

232. *Id.* at 330–31. Interestingly, *Monroe* is the only one of the four seminal communications privacy cases in which CAAF considered the SCA. For a discussion of the SCA’s limitations on disclosure, see *supra* text accompanying notes 121–23123.

233. *Monroe*, 52 M.J. at 330–31. It would have been interesting if the CAAF had explored the distinction between “in transit” and “stored” communications before concluding the SCA was the applicable statutory scheme, especially since e-mails that have not been delivered to the recipient due to a technical malfunction are, in a sense, intercepted while “in transit.” The Wiretap Act prohibits unauthorized “interception” of electronic communications, while the SCA prohibits the unauthorized access of an electronic communication while it is in “electronic storage.” Under 18 U.S.C. § 2510(4) (2006),

observed, disclosure by a communications systems provider (such as the U.S. Air Force) to a law enforcement agency is permissible “if such contents (A) were inadvertently obtained by the service provider; and (B) appear to pertain to the commission of a crime.”<sup>234</sup> Because the systems administrators had obtained the e-mail inadvertently without a law enforcement purpose, the disclosure was appropriate.<sup>235</sup>

In *Maxwell*, the CAAF laid the foundation for finding a reasonable expectation of privacy in personal e-mail. In *Monroe*, the CAAF built on that foundation by holding that a user sending e-mail over a government network does not possess a reasonable expectation of privacy against systems monitoring, but also stating in dicta that users do enjoy a reasonable

---

“‘intercept’ means the aural or other acquisition of the contents of any wire, electronic, or oral communication through the use of any electronic, mechanical, or other device.” Under 18 U.S.C. § 2510(17)(A), “‘electronic storage’ means any temporary, intermediate storage of a wire or electronic communication incidental to the electronic transmission thereof.” Based on these definitions, the e-mails at issue in *Monroe* most likely fall under the SCA because they were in “temporary” storage even while in transit. For a discussion of the distinction between “in transit” and “stored” communications, see *supra* text accompanying note 123.

234. *Monroe*, 52 M.J. 326, 331 (citing 18 USC § 2702(b) (1994)) (internal quotation marks omitted). By design, the SCA distinguishes between “public” and “nonpublic” service providers. Commentators disagree about the SCA’s applicability to the military as a “public” service provider. Compare *Coacher*, *supra* note 152, at 178 (arguing SCA is not applicable to U.S. Air Force), with *Edell*, *supra* note 9, at 11 (arguing DoD regulations and military practice make SCA applicable U.S. Army). Either way, the court in *Monroe* applied the SCA to the Air Force, 52 M.J. at 330, and the DoD requires electronic communications interceptions be conducted in accordance with the federal statutory scheme. See DEPT OF DEFENSE, DIR. 5505.9, INTERCEPTION OF WIRE, ELECTRONIC, AND ORAL COMMUNICATIONS FOR LAW ENFORCEMENT (1995). On May 21, 2009 the Air Force Network Operations Commander, Major General David Senty, issued an Air Force-wide memorandum stating, “[a]s a non-public communications service provider, the Electronic Communications Privacy Act (ECPA) authorizes Air Force systems administrators to access stored communications stored [sic] if they are acting within their authority.” Memorandum from the Air Force Network Operations for All Installation Commanders and All Network Control Centers (May 21, 2009) (on file with author) (emphasis added); see also Rich Ladue, *Bullet Background Paper on the Voluntary Provision of Stored Communications and Data, the Designated Accrediting/Approval Authority, & Computer Trespasser Monitoring*, U.S. AIR FORCE, May 26, 2009 (advising AFOSI agents need only obtain Air Force approval—not user approval—before searching service member e-mails under SCA). Whether the Air Force’s claim of its non-public service provider status will survive judicial scrutiny, especially given that the Air Force portal allows non-military family members to establish e-mail and instant messaging accounts, remains to be seen. See Air Force Portal: Friends & Family Instant Messenger, <https://guest.my.af.mil/> (last visited Sept. 4, 2009); see also Jim Tice, *Stricter Security Won’t Interfere with AKO Access*, ARMY TIMES, May 29, 2006, at 16 (discussing “Army Knowledge Online” access for service members, contractors, retirees, and military family and friends).

235. *Monroe*, 52 M.J. at 331. Even if the disclosure had been in violation of the SCA, Staff Sergeant Monroe still would have been unsuccessful in suppressing the e-mails because the SCA does not provide a statutory exclusionary remedy, and the court had already determined a constitutional violation did not occur. See *supra* text accompanying note 122.

expectation of privacy against the police. In *Long*, the CAAF had occasion to test that proposition by considering a “police” intrusion in the form of a command-directed criminal investigation. The question at issue was whether and to what extent military members enjoy a reasonable privacy expectation in their workplace e-mail when the purpose of the intrusion is law enforcement.<sup>236</sup>

3. United States v. Long—*The Fourth Amendment and Government E-Mail Searches*

In 2006, six years after its decision in *Monroe* and ten years after its decision in *Maxwell*, the CAAF heard *United States v. Long*, a case in which “the reasonable expectation of privacy a military person has in e-mail messages sent and stored on a government computer system” was placed front and center.<sup>237</sup> Because the e-mails at issue were searched for law enforcement purposes, rather than routine systems monitoring, the CAAF found a reasonable expectation of privacy, confirming in *Long* what it had stated in dicta in *Monroe* regarding the distinction between work-related and law enforcement intrusions.<sup>238</sup>

Contrary to her pleas, Lance Corporal Long was convicted of unlawful drug use in violation of Article 112a, UCMJ.<sup>239</sup> The evidence against her was based in part on e-mails obtained from her government e-mail account in which she had indicated “a fear that her drug use would be detected by urinalysis testing and the steps she had taken in an attempt to avoid such detection.”<sup>240</sup> The e-mails were found during the course of an investigatory search conducted by the systems administrator at the direction of an investigator from the Marine Corps Inspector General, who was investigating—not possible drug use—but “allegations of an improper

---

236. *United States v. Long*, 64 M.J. 57 (C.A.A.F. 2006). In her dissenting opinion in *Long*, Judge Crawford, who had joined the majority in *Monroe*, recognized that the court left open in *Monroe* the issue presented in *Long*, stating, “[i]n *United States v. Monroe*, we held that a defendant does not have an expectation of privacy in his e-mail, ‘at least from the personnel charged with maintaining the . . . [electronic mail host] system.’ We left open the issue presented in this particular case.” *Id.* at 67 (Crawford, J., dissenting).

237. *Id.* at 57–58. In *Long*, Judge Gierke, who had concurred in part and concurred in full in *Maxwell* and *Monroe*, respectively, delivered the court’s opinion. Judge Effron, who was not on the court for *Maxwell* but had concurred in *Monroe*, joined him. The other majority members, Judge Baker and Judge Erdmann, had not been on the court for either *Maxwell* or *Monroe*. *Id.*

238. *Id.* at 65.

239. *Id.* at 59. 10 U.S.C. § 912 (2006) prohibits the wrongful use, possession, manufacturing, distribution, or importation of controlled substances.

240. *Long*, 64 M.J. at 57–59.

relationship between [Lance Corporal Long] and an officer.”<sup>241</sup> After the search, the drug-related e-mails were provided to officials and subsequently introduced against Lance Corporal Long at trial.<sup>242</sup>

At trial, the defense moved to suppress the e-mails as violating the Fourth Amendment.<sup>243</sup> The systems administrator testified for the government in opposition to the suppression motion, stating (a) the search had been conducted to find evidence of misconduct, not as part of routine systems monitoring; (b) each computer user “had his or her own unique password known only to them”; (c) the systems administrator could access users’ e-mails, although “it was general policy to avoid examining e-mails and their content because it was a ‘privacy issue’”; (d) the policy regarding personal e-mails “had always been lenient and that such use of the network was considered authorized”; and (e) a logon banner appeared every time a user logged onto the network, informing the user that using the network constituted consent to monitoring.<sup>244</sup> The banner stated:

This is a Department of Defense computer system. This computer system, including all related equipment, networks and network devices (specifically including Internet access), are provided only for authorized U.S. Government use. DoD computer systems may be monitored for all lawful purposes, including to ensure that their use is authorized, for management of the system, to facilitate protection against unauthorized access, and to verify security procedures, survivability and operational security. Monitoring includes active attacks by authorized DoD entities to test or verify the security of this system. During monitoring, information may be examined, recorded, copied and used for authorized purposes. All information, including personal information, placed on or sent over this system may be monitored. Use of this DoD computer system, authorized or unauthorized, constitutes consent to monitoring of this system. Unauthorized use may subject you to criminal prosecution. Evidence of unauthorized use collected during monitoring may be used for administrative, criminal, or other adverse action. Use of this system constitutes consent to monitoring for these purposes.<sup>245</sup>

Although the trial judge determined the search was a non-consensual search for evidence conducted without search authorization, the judge denied the

---

241. *Id.* at 59.

242. *Id.*

243. *Id.*

244. *Id.* at 59–60, 64.

245. *Id.* at 60.

motion to suppress on the basis that Lance Corporal Long “had no expectation of privacy in the e-mails stored on the government server.”<sup>246</sup>

The intermediate appellate court, the Navy–Marine Corps Court of Criminal Appeals, disagreed, finding the logon banner removed Lance Corporal Long’s expectation of privacy only as to systems monitoring, not law enforcement activity.<sup>247</sup> The court sustained the conviction, however, on the basis of harmless error.<sup>248</sup>

Employing the *Katz* two-part inquiry in the context of the *O’Connor* workplace privacy analysis, the CAAF first reviewed whether Lance Corporal Long had a subjective expectation of privacy in her government e-mail account. The CAAF noted Lance Corporal Long had her own password, the systems administrator had “repeatedly emphasized the agency practice of recognizing the privacy interests of users in their e-mail,” and the banner described monitoring, not law enforcement, activity.<sup>249</sup> Finding a subjective privacy expectation, the CAAF then reviewed the “operational realities of the workplace” to determine whether Lance Corporal Long’s expectation was objectively reasonable.<sup>250</sup> While acknowledging the e-mails were sent from a government computer, over a government computer network, and stored on a government server, the court found that actual office practices “reaffirm[ed] rather than reduce[d] the expectations regarding privacy on office computers.”<sup>251</sup> The court then contrasted *Monroe*, in which the e-mails were discovered in the course of routine monitoring, not law enforcement.<sup>252</sup> In light of the banner’s limited monitoring notice and actual workplace practices, the CAAF found Lance Corporal Long’s expectation of privacy reasonable.<sup>253</sup> Because the search was for a law enforcement—not work-related—reason, the search required probable cause, which had not been obtained.<sup>254</sup> The e-mails were accordingly suppressed, the error was deemed not harmless, and the conviction was set aside.<sup>255</sup>

The CAAF’s opinion in *Long* sent ripples through the military legal and law enforcement communities, with critics expressing concern that such privacy expectations could deprive the military “of the ability to effectively

---

246. *Id.* at 59–60.

247. *Id.* at 60.

248. *Id.*

249. *Id.* at 63.

250. *Id.*

251. *Id.* at 64.

252. *Id.* at 64–65.

253. *Id.*

254. *Id.*

255. *Id.* at 66.

monitor government communications and respond to threats to national security.”<sup>256</sup> The military responded by issuing guidance for military attorneys and investigators about the impact of *Long* on criminal military investigations,<sup>257</sup> and the Joint Task Force–Global Network Operation began drafting a revised logon banner for the DoD Chief Information Officer to “preclude any claims to an expectation of privacy on base network servers for e-mails or any other digital evidence.”<sup>258</sup> To some critics, it may have seemed ludicrous that a military member could possess a reasonable expectation of privacy in e-mail sent from a government computer over a government network, especially when the government computer was issued for official use and contained a logon banner to that effect.<sup>259</sup> Two years later, the concerns of those critics would find expression in the CAAF’s next major computer privacy case, *United States v. Larson*.<sup>260</sup>

---

256. Mary L. Walker, *Lack of Privacy in Government Computer Systems*, GEN. COUNSEL’S Q. (2007).

257. See Edell, *supra* note 9, at 24 n.325 (citing e-mail from Deputy Assistant Staff Judge Advocate Gen. (Criminal Law) to All Navy and Marine Corps Judge Advocates, subject: Search Authorization for Computer Files in Light of *United States v. Long*, 64 M.J. 57, Part II (1 June 2007)); Rich Ladue, *Bullet Background Paper—U.S. v. Long*, 4 TJAG ONLINE NEWS, Oct. 4, 2006 [hereinafter Ladue, *U.S. v. Long*]; Rich Ladue, *Bullet Background Paper on U.S. v. Long and Its Impact on Investigations*, U.S. AIR FORCE, Oct. 3, 2006 [hereinafter Ladue, *Impact on Investigations*]; Mary L. Walker, *Expectation of Privacy in Computer Systems: Follow-Up*, GEN. COUNSEL’S Q. (2007); Memorandum from W. Kipling At Lee, Jr., Deputy Gen. Counsel (Nat’l Sec. & Military Affairs), to Air Force Office of Special Investigations Judge Advocate, subject: Computer Privacy (Dec. 14, 2006) (on file with author).

258. Ladue, *Impact on Investigations*, *supra* note 257.

259. In fact, these were largely the concerns raised by Judge Crawford in her vigorous dissent. *Long*, 64 M.J. at 67 (Crawford, J., dissenting). In her dissent, Judge Crawford argued that the majority’s decision rested on the erroneous assumption that “an objective reasonable expectation of privacy can be preserved for some forms of seizure despite being nonexistent for others.” *Id.* In Judge Crawford’s view,

Once [Lance Corporal Long] was given notice of and consented to monitoring of any kind, she could not maintain a reasonable expectation of privacy against other forms of intrusion . . . . That the communications were obtained specifically for law enforcement purposes has no bearing on her expectation of privacy.

*Id.* at 67–68. Judge Crawford also argued the search was lawful on the basis of consent. *Id.* at 70. Judge Crawford’s opinion that the purpose of the search is irrelevant is difficult to square with the Supreme Court’s view in *O’Connor*, which explicitly stated the purpose of the search was relevant. See *O’Connor v. Ortega*, 480 U.S. 709, 716 (1987) (plurality opinion) (“Within the workplace context, this Court has recognized that employees may have a reasonable expectation of privacy against intrusions by police.” (citation omitted)). For additional discussion of user consent, see *infra* Section IV.B.3.

260. 66 M.J. 212 (C.A.A.F. 2008). I use the term “computer privacy” case purposefully, as *Larson* involves internet files rather than e-mails. On this basis alone, future courts may find *Larson* distinguishable from its predecessors.

4. United States v. Larson—*The Fourth Amendment and Government Computer Searches*

Unlike *Maxwell*, *Monroe*, and *Long*, which all involved e-mail, *United States v. Larson* involved a service member's privacy expectation in electronic files residing on his government computer hard drive.<sup>261</sup> In an opinion that, in some ways, stands analytically apart from the three prior communications privacy cases, the CAAF<sup>262</sup> held that the facts presented did not provide a basis to find either a subjective or objective expectation of privacy.<sup>263</sup>

Contrary to his pleas, a general court-martial convicted Major Larson of attempted carnal knowledge, attempted indecent acts with a minor, violation of a general regulation, communicating indecent language, and attempting to entice a minor to engage in sexual activity.<sup>264</sup> Evidence was presented that Major Larson, a 36-year-old reservist at Schriever Air Force Base, Colorado Springs, Colorado, was temporarily occupying the office of a deployed activity duty member when he began using his government computer to send instant messages (IM) to a fourteen-year-old girl named "Kristin."<sup>265</sup> "Kristin" was, in fact, a civilian detective posing online as a young girl to catch sexual predators.<sup>266</sup> After multiple salacious conversations, Major Larson and "Kristin" agreed to meet at a nearby mall to have sex.<sup>267</sup> Major Larson was met instead by local police, who arrested him and alerted the

---

261. *Id.* at 214. The prior CAAF case most factually analogous to *Larson* is *United States v. Tanksley*, which also addressed a service member's privacy interest in his government workplace computer. *United States v. Tanksley*, 54 M.J. 169, 171–72 (C.A.A.F. 2007) (finding no reasonable privacy expectation from work-related intrusion uncovering criminal evidence in document open on computer screen). However, *Tanksley* did not involve a law enforcement search. *Id.*

262. In a unanimous decision, Judge Ryan, who was not on the court for any of the three prior decisions, issued the opinion of the court in *Larson*. She was joined by Judge Baker, Judge Erdmann, and Judge Stucky. Judge Effron, filing a separate concurring opinion here, had also joined the majority opinion in *Long*. 66 M.J. at 213.

263. *Long*, 66 M.J. at 216. In addition to the Fourth Amendment challenge, the CAAF also addressed a claim of ineffective assistance of counsel. The CAAF found that the defendant's civilian defense counsel had been deficient in conceding guilt to one of the charged offenses, but that the defendant was not prejudiced by his attorney's tactical decision. *Id.* at 219. Because the ineffective assistance of counsel claim is not relevant to this Article, I do not discuss it here.

264. *Id.* at 213. The specifications alleged violations of Articles 80, 92, and 134, UCMJ. Major Larson was sentenced to nine years confinement, dismissal from the military, and forfeiture of all pay and allowance. The convening authority reduced his confinement to six years. *Id.*

265. *United States v. Larson*, 64 M.J. 559, 561 (A.F. Ct. Crim. App. 2006).

266. *United States v. Larson*, 66 M.J. 212, 214 (C.A.A.F. 2008).

267. *Id.* at 214.

AFOSI.<sup>268</sup> Continuing the investigation, the AFOSI contacted Major Larson's commander, who opened Major Larson's office using a master key and allowed the AFOSI to seize Major Larson's government computer.<sup>269</sup> A search of the computer "revealed stored pornographic material, a web browser history that showed [Major Larson] visited pornographic websites and engaged in sexually explicit chat sessions in his office on his government computer, and other electronic data implicating [Major Larson] in the charged offenses."<sup>270</sup>

At trial, Major Larson unsuccessfully moved to suppress all evidence seized from his government computer, arguing a violation of his Fourth Amendment privacy rights.<sup>271</sup> Unlike *Long*, in which the systems administrator testified on behalf of the government, the government placed Major Larson's commander on the stand.<sup>272</sup> Based on his testimony, the military judge found that (a) Major Larson occupied a private, lockable office, though others had keys, (b) the government computer was for official use, (c) the computer was password protected, although a systems administrator still could access the hard drive, and (d) a logon banner appeared every time a user logged onto the network, informing the user "the computer was DoD property, was for official use, and that [the user] consented to monitoring."<sup>273</sup> In accordance with these findings, the military judge denied the defense's suppression motion.<sup>274</sup>

On appeal, the Air Force Court of Criminal Appeals characterized the case as one "of first impression in some respects."<sup>275</sup> Citing both *Long* and *Monroe*, the appellate court noted that prior military cases involving computers and the Fourth Amendment focused on e-mail communications, whereas "[t]he search of the government computer here did not focus on such communications."<sup>276</sup> In this case, the AFOSI's investigative search involved "certain data files, created as part of the 'normal operating procedure' of the Microsoft Windows operating system . . . ." <sup>277</sup> Because the

---

268. *Id.*; *Larson*, 64 M.J. at 562.

269. *Larson*, 66 M.J. at 214.

270. *Id.* at 214–15. The CAAF opinion characterizes the evidence as "files" stored on Major Larson's workplace computer. *Id.* The intermediate appellate court opinion clarifies, however, that the files were all temporary Internet files, cached from Major Larson's Internet browsing. *Larson*, 64 M.J. at 563.

271. 66 M.J. at 214.

272. *Larson*, 66 M.J. at 214.

273. *Id.*

274. *Id.* at 215; *Larson*, 64 M.J. at 563.

275. *Larson*, 64 M.J. at 563.

276. *Id.*

277. *Id.*

“data files” were created as part of a “normal operating procedure,” they were more akin to routine systems monitoring than active law enforcement efforts:

There is no evidence the appellant was aware the Internet history files existed, and we are unconvinced the appellant could entertain a subjective expectation of privacy in them without such knowledge. Moreover, we conclude such an expectation, even if it existed, would on these facts not be reasonable. *The data in question was recorded automatically, not for law enforcement purposes, but as part of the computer’s operating system.* The appellant could not expect to keep private automatically-recorded data stored on government property he would reasonably have known would be turned over to another officer on that officer’s return from deployment.<sup>278</sup>

Finding Major Larson had neither a subjective nor an objective expectation of privacy in such data, the intermediate appellate court held the trial judge had not abused his discretion.<sup>279</sup>

Surprisingly, the CAAF took an entirely different tack from the lower appellate court, avoiding any discussion of the case as one of first impression. Further, the court did not analyze either the impact of the deployed member’s departure or return on Major Larson’s privacy interest or the nature of the data files as being “recorded automatically, not for law enforcement purposes.”<sup>280</sup> In fact, the CAAF even took a different analytical tack from its previous approaches in *Maxwell*, *Monroe*, and *Long*, focusing heavily on Major Larson’s privacy interest in the situs of the search rather than the data files on the computer: “[i]n addressing *Fourth Amendment* privacy claims, the threshold issue is whether the person has a legitimate expectation of privacy in the invaded *place*.”<sup>281</sup> The CAAF continued its emphasis on “place” by quoting MRE 314(d), which, in CAAF’s view, created a rebuttable presumption that individuals cannot have a reasonable expectation of privacy in government property:

Government property may be searched under this rule unless the person to whom the property is issued or assigned has a reasonable expectation of privacy therein at the time of the search. Under

---

278. *Id.* (emphasis added).

279. *Id.*

280. *Compare Larson*, 66 M.J. at 214–16, *with Larson*, 64 M.J. at 563.

281. *Larson*, 66 M.J. at 215 (final emphasis added). The *Larson* court’s recitation of the law was not in any way incorrect. In fact, the court’s statement was taken almost verbatim from *Rakas v. Illinois*, 439 U.S. 128, 143 (1978), a Supreme Court opinion relying on *Katz*. However, *Rakas* was not a communications privacy case. It involved search of a vehicle and heavily emphasized property and possessory interests. 439 U.S. at 143.

normal circumstances, a person does not have a reasonable expectation of privacy in government property that is not issued for personal use . . . .<sup>282</sup>

The CAAF noted that Major Larson had utilized a government computer that was accessible by both the commander and systems administrator, and located in an office that was also accessible by the commander and others.<sup>283</sup> Finding no evidence that Major Larson had a subjective expectation of privacy in his government computer, and finding evidence based on the commander's testimony and logon banner to support "the validity of the presumption that [Major Larson] had no reasonable expectation of privacy in the government computer," the CAAF sustained the lower courts' rulings.<sup>284</sup>

What is surprising in *Larson* is not the outcome reached by the CAAF (after all, the lower courts had reached the same conclusion, and the CAAF had also found no reasonable expectation of privacy in *Monroe*), but rather the path it took to get there. First, as discussed below, the court did not utilize the *O'Connor* workplace analysis. Second, the court did not discuss the distinction between work-related and law enforcement searches. Finally, the court did not analyze (as the intermediate appellate court had done) whether Major Larson had any privacy expectation in the data files on his computer as opposed to the computer itself. The reasoning in *Larson* thus differed substantially from that the CAAF adopted in *Maxwell*, *Monroe*, and *Long*, all of which are fundamentally based on the normative "content" inquiry the Supreme Court set forth in *Katz*.<sup>285</sup>

---

282. *Larson*, 66 M.J. at 215 (quoting MIL. R. EVID. 314(d)).

283. *Id.* at 215–16. Although the CAAF opinion did not mention it, the intermediate appellate court relied on the fact that the office belonged to another service member who was deployed, still had the deployed member's personal things in it, and would be used (including the computer) upon the deployed member's return. *Larson*, 64 M.J. at 563.

284. *Larson*, 66 M.J. at 216. To a certain extent, in *Larson* the CAAF balanced its current and prior rulings on the head of a pin by referring to the scope of judicial review, leaving open the possibility of continuing developments. Military appellate courts review a military judge's decision to admit or exclude evidence for an abuse of discretion. *See* *United States v. McCollum*, 58 M.J. 323, 335 (C.A.A.F. 2003). In *Larson*, the CAAF reiterated its limited holding in *Long*, which only affirmed "the lower court was not clearly erroneous" in finding a subjective expectation of privacy, and stated in its present holding, "we agree with the [Court of Criminal Appeals] that the military judge did not abuse his discretion" in finding no expectation of privacy in Major Larson's government computer. 66 M.J. at 216.

285. Professor Freivald argues that the Supreme Court, in *Smith v. Maryland*, 442 U.S. 735 (1979), similarly "avoided the normative inquiry required by *Katz*" by failing to discuss "the vital nature of the telephone system" and "whether telephone users should be entitled to expect their telephone numbers to remain protected by the Fourth Amendment." Freivald, *supra* note 22, at 47.

In *O'Connor*, the Supreme Court relied on actual workplace practices to determine the reasonableness of the employee's privacy expectation.<sup>286</sup> It was not enough to aver summarily that a particular office was accessible to or subject to search by others.<sup>287</sup> What mattered was whether others actually accessed the office, how often, and in what manner.<sup>288</sup> In *Larson*, however, the CAAF cited the fact that "fire department and the command's facility manager" had keys to Major Larson's office, but never queried how often they actually used those keys to enter the office in Major Larson's absence.<sup>289</sup> Similarly, the court found that both the systems administrator and commander could access the computer's hard drive without Major Larson's password, but did not discuss whether they had ever done so.<sup>290</sup> Further, while it was important to the intermediate appellate court's opinion that Major Larson occupied a deployed member's office, the CAAF's opinion did not mention that the office was only temporarily assigned to Major Larson, still had the deployed officer's personal things in it, and would be reoccupied by the deployed officer upon his return.<sup>291</sup>

---

286. See *supra* text accompanying notes 68–73.

287. In *United States v. Kaban*, the district court expressly found that government employees have an expectation of privacy from government criminal investigators and government supervisors may not provide third-party consent to private office areas. 350 F. Supp. 784, 795–96 (S.D.N.Y. 1972); see also *supra* text accompanying notes 74–80 (discussing *Kaban*). In *Larson*, the CAAF completely overlooked *Kaban* and the Supreme Court's endorsement of it in *O'Connor*. *Larson*, 66 M.J. at 215–16. But see *United States v. Muniz*, 23 M.J. 201, 205–06 (C.M.A. 1987) (citing *Mancusi v. DeForte*, 392 U.S. 364, 369–70 (1968), for the proposition that "a business supervisor [can] consent to the search of company property in the custody of a subordinate").

288. See *supra* text accompanying notes 68–73.

289. *Larson*, 66 M.J. at 214. As one blogger noted on CAAFlog, a blog about military law, in response to *Larson*:

Does the fact that a fire dept (or a cleaning person) has [sic] access to a space means it is not private? Of course not. Society still expects that space to be private, notwithstanding who might have keys. Otherwise, the fact that an apartment building manager has keys to your apartment would destroy your expectation of privacy.

Posting of Dwight Sullivan to CAAFlog, Government Computers and Expectations of Privacy, <http://caaflog.blogspot.com/2008/04/government-computers-and-expectation-of.html> (Apr. 27, 2008 11:09 PM).

290. *Larson*, 66 M.J. at 214.

291. Compare *Larson*, 66 M.J. at 214–16, with *Larson*, 64 M.J. at 563. With no exigent circumstances presented as to why the AFOSI did not obtain search authorization when it clearly could have done so, the Supreme Court's reasoning in *Chapman v. United States* is especially applicable: "We think it must be concluded here, as it was in *Johnson*, that 'If the officers in this case were excused from the constitutional duty of presenting their evidence to a magistrate, it is difficult to think of a case in which it should be required.'" 365 U.S. 610 (citing *Johnson v. United States*, 333 U.S. 10, 15 (1948)).

Similarly, the CAAF chose not to inquire into whether and to what extent the purpose of the intrusion—law enforcement—factored into the reasonableness of Major Larson’s privacy expectation. In *Monroe* and *Long*, the court’s holdings largely turned on the distinction raised by the Supreme Court in *O’Connor* regarding the purpose of the search. In *O’Connor*, the Supreme Court held that employers require wide latitude to ensure “the efficient and proper operation of the workplace,” but distinguished between employer intrusion for work-related and law enforcement purposes.<sup>292</sup> Thus, employees may not have a reasonable privacy expectation against work-related intrusion by fellow employees and supervisors, but may have a reasonable privacy expectation against police intrusion. In *Larson*, however, the CAAF avoided this inquiry, even though AFOSI agents who were cooperating with civilian authorities in a clearly criminal investigation conducted the search.

Finally, the CAAF failed to explore the rather technologically savvy analytical hook the Air Force Court of Criminal Appeals used in resolving the case’s work-related/law enforcement conundrum. The conundrum may be approached as follows:

- The search of Major Larson’s office was conducted pursuant to a criminal investigation. Although it was work-related in the sense that, in the military, workplace misconduct may constitute criminal misconduct, the crimes at issue generally did not pertain to Major Larson’s duties.
- Based on the testimony of Major Larson’s commander, who testified that he and others had access to Major Larson’s office and computer, Major Larson probably did not have a reasonable expectation of privacy against his work associates for work-related intrusions. The fact that Major Larson was occupying a deployed member’s office also militates against finding a work-related privacy expectation.
- The conclusion that Major Larson did not have a reasonable expectation of privacy for work-related intrusions does not mean that he did not have a reasonable expectation of privacy against intrusions for law enforcement purposes. If the logon banner was limited to “monitoring,” and if Major Larson’s office and computer were not

---

292. *O’Connor v. Ortega*, 480 U.S. 709, 721 (1987) (plurality opinion) (“While police, and even administrative enforcement personnel, conduct searches for the primary purpose of obtaining evidence for use in criminal or other enforcement proceedings, employers most frequently need to enter the offices and desks of their employees for legitimate work-related reasons wholly unrelated to illegal conduct.”).

routinely monitored and searched, arguably, he could have had a reasonable expectation of privacy against a law enforcement search of his computer files.<sup>293</sup>

To avoid this conundrum, and the conclusion to which it may have lead, the intermediate appellate court relied on the fact that the files downloaded by Major Larson were temporary Internet files, “recorded automatically, not for law enforcement purposes” as part of “‘normal operating procedure’ of the Microsoft Windows operating system.”<sup>294</sup> By finding they were automatically recorded files, which the deployed service member could have found upon his return, the court sidestepped the law enforcement issue.<sup>295</sup> This approach does not completely resolve the issue as the search itself was still for law enforcement purposes, but it at least acknowledges the notion that reasonableness depends on the purpose of the search.<sup>296</sup>

Whatever the shortcomings of *Larson*, there is little doubt that those who had expressed dismay over the holding reached by the CAAF in *Long* were relieved by the apparent turnabout in *Larson*.<sup>297</sup> Not everyone, however, viewed *Larson* as a complete panacea for the problems caused by the court’s holding in *Long*.<sup>298</sup> Perhaps several reasons support such caution. First, as the Air Force Court of Criminal Appeals noted in its intermediate appellate opinion, *Larson* is almost a matter of first impression.<sup>299</sup> Second, unlike *Maxwell*, *Monroe*, and *Long*, which all addressed communications privacy as it

---

293. One downside of emphasizing “actual” workplace practices is that it creates a perverse incentive for employers to routinely invade their employees’ workspaces to nip any expectation of privacy in the bud before it has a chance to grow.

294. *Larson*, 64 M.J. at 563.

295. *Id.*

296. The *Larson* court could have drawn on *United States v. Muniz* for guidance. *United States v. Muniz* 23 M.J. 201 (C.M.A. 1987); see also *supra* text accompanying notes 82–87. As discussed above, in *Muniz* the Court of Military Appeals addressed whether a service member had a reasonable expectation of privacy in the address on an envelope kept in a locked credenza in a lockable private office. 23 M.J. at 205 n.5. In discussing the service member’s reasonable privacy expectations, the court observed, “[t]he government property was the drawer itself. Assuming the legitimacy of that entry, the return address, on what was undeniably private property, could apparently be seen in plain view.” *Id.* (citation omitted). In *Larson*, the temporary Internet files were stored on Major Larson’s hard drive, a locked but accessible “container.” *Larson*, 64 M.J. at 563. Like the envelope in *Muniz*, the CAAF could have found that the Internet files were “envelope” information available for public view (though whether temporary Internet files are “envelope” or “content” information is still unresolved). See Tokson, *supra* note 21, at 2109.

297. See *supra* Section III.C.4.

298. See Mendelson, *supra* note 9, at 9 (urging judge advocates to continue seeking probable cause search authorization for government e-mail accounts even after *Larson*).

299. *Larson*, 64 M.J. at 563.

pertained to e-mail, *Larson* involved the government computer itself, making it much more like earlier military cases addressing privacy interests in a government desk or credenza than a true communications privacy case. Finally, *Larson* pre-dated the DoD's new logon banner and user agreement policy, which has yet to be reviewed by military appellate courts.<sup>300</sup> As discussed in Part IV below, to the extent that military authorities and investigators rely on *Larson* for the proposition that the government may conduct warrantless searches of government e-mail accounts for law enforcement purposes, they may be doing so at their own risk.

#### IV. THE LANDSCAPE AFTER *LONG* AND *LARSON*

Having surveyed the law of workplace privacy in the military and the guidepost communications privacy cases issued by the CAAF, this Part turns to the DoD's response to the *Long* decision. This Part first reviews the DoD's new logon banner and user agreement policy. It then explores the possibility that the DoD's new logon banner constitutes an unconstitutional violation of U.S. service members' Fourth Amendment rights by functioning as a general warrant. I advocate for a revised policy based on a normative determination that electronic communications—even in the workplace—are sufficiently vital to warrant limited workplace protections. This Part then concludes by suggesting that both military interests and individual service member interests are best protected by a legal standard that requires probable cause search authorization whenever the primary purpose of a search is law enforcement.

##### A. THE DOOD'S NEW LOGON BANNER AND CONSENT AGREEMENT

For military practitioners struggling to reconcile the holdings of *Long* and *Larson*, two dissimilar factual findings must have seemed particularly thorny. First, the *Long* court placed significant weight on the systems administrator's testimony that "it was a general policy to avoid examining e-mails and their content because it was a 'privacy issue.'"<sup>301</sup> The court noted that the systems administrator "repeatedly emphasized the agency practice of recognizing the privacy interests of users in their e-mail"<sup>302</sup> and that it found "the testimony of the network administrator, describing the agency practices and policies to be most persuasive."<sup>303</sup> In *Larson*, however, the military commander testified "he could log onto [Major Larson's] computer with his own password and

---

300. See *infra* text accompanying note 323.

301. *United States v. Long*, 64 M.J. 57, 60 (C.A.A.F. 2006).

302. *Id.* at 63.

303. *Id.* at 64.

access all portions of the hard drive unless [Major Larson] protected something with his own password.”<sup>304</sup> Although the court in *Larson* did not state whether the commander had actually logged onto Major Larson’s computer, it found the possibility that he could do so persuasive.<sup>305</sup> Comparing these two approaches, it appears that the CAAF was concerned with the *practice* of intrusion in *Long*, but only the *possibility* of intrusion in *Larson*.

The second prickly point in both cases is the logon banner. In *Long*, the court relied heavily on the language of the logon banner, which “described access to ‘monitor’ the computer system, [but] not to engage in law enforcement intrusions by examining the contents of particular e-mails in a manner unrelated to maintenance of the e-mail system.”<sup>306</sup> As a result, the court found that military authorities overstepped their bounds by intruding for law enforcement purposes.<sup>307</sup> In *Larson*, however, the logon banner appeared to say much the same thing, but it led the court to a different conclusion. The logon banner “state[d] that it was a DoD computer, it [was] for official use, [and] not to be used for illegal activity.”<sup>308</sup> The banner “also had a statement that users of the computer consent to monitoring.”<sup>309</sup> Yet, the *Larson* court found the banner was sufficient to place Major Larson “on notice that the computer was not to be used for illegal activity and that there could be third-party monitoring.”<sup>310</sup>

---

304. *United States v. Larson*, 66 M.J. 212, 215 (C.A.A.F. 2008).

305. To me, it is troubling that the court was persuaded by the fact that the commander could (but not necessarily did) log onto the network using Major Larson’s computer. First, it is not surprising that the commander—who apparently had a network account—could access the network by logging onto it from Major Larson’s networked computer. Second, analogizing the computer to an office, the mere fact that a supervisor can access an employee’s office is hardly grounds, at least under *O’Connor*, an employee does not have reasonable expectation of privacy in the computer. *O’Connor* seems to stand for the proposition that it is office practices—not possibilities—that define what is and is not reasonable for Fourth Amendment workplace privacy protections. *See O’Connor v. Ortega*, 480 U.S. 709 (1987) (plurality opinion). Third, there is no evidence that Major Larson even knew the commander could log onto his computer.

306. *Long*, 64 M.J. at 63.

307. *Id.* at 63, 65.

308. *Larson*, 66 M.J. at 216.

309. *Id.*

310. *Id.* As far as I know, no government property—or any other property—is intended to be used for illegal activity. To the extent the court relied on this provision to negate a subjective or objective expectation of privacy, its reliance appears to be misplaced. A prohibition against illegal use hardly seems the same as a knowing consent to a search for illegal use.

The questions left by these cases are tangled. With respect to access, does a user retain a reasonable privacy expectation in his e-mail (or computer) when a systems administrator can—but normally does not—access the user’s e-mail account? Or is the mere possibility that a third party can access the e-mail (or computer) enough to extinguish the user’s privacy interest? With respect to monitoring, does a user have to be on notice that his computer activity may be monitored for both work-related and law enforcement purposes, or is it sufficient to simply instruct the user that monitoring may occur and the system may not be used for illegal purposes?

The DoD’s response to these questions was to revise and broaden its logon banner.<sup>311</sup> On May 8, 2008, two years after *Long* and a few weeks after *Larson*, the DoD issued its “Standard Consent Banner and User Agreement” policy, notifying users that DoD computer communications were not private and could be monitored, searched, inspected and seized at any time and for any purpose.<sup>312</sup> The new policy required all DoD computer systems to adopt the new banner within sixty days, and encouraged widespread training, publication, and security awareness briefings to inform DoD users of the policy. Attachment 1 to the Banner and User Agreement policy contained the new Standard Mandatory DoD Notice and Consent Banner, which DoD users were to acknowledge each time they logged onto their computers:

---

311. See Ladue, *Impact on Investigations*, *supra* note 257 (noting the “Joint Task Force–Global Network Operations is updating the banner and requesting approval from the DoD Chief Information Officer so that the DoD Notice and Consent Banner will preclude any claims to an expectation of privacy on base network servers for e-mails or any other digital evidence”).

312. Memorandum from John Grimes, Chief Info. Officer, Dep’t of Defense to Sec’y of the Military Dep’ts (May 9, 2008). The new policy directly impacted service regulations. For example, a prior version of the U.S. Army’s Information Assurance regulation “specifically stated that computer users had a reasonable expectation of privacy.” Edell, *supra* note 9, at 1 n.7 (citing U.S. DEPARTMENT OF ARMY, REGULATION 25-2, INFORMATION ASSURANCE para. 4–5r (Nov. 14, 2003)).

Figure 1: Standard Mandatory DoD Notice and Consent Banner

ATTACHMENT 1  
STANDARD MANDATORY  
DOD NOTICE AND CONSENT BANNER

[A. Use this banner for desktops, laptops, and other devices accommodating banners of 1300 characters. The banner shall be implemented as a click-through banner at logon (to the extent permitted by the operating system), meaning it prevents further activity on the information system unless and until the user executes a positive action to manifest agreement by clicking on a box indicating “OK.”]

You are accessing a U.S. Government (USG) Information System (IS) that is provided for USG-authorized use only.

By using this IS (which includes any device attached to this IS), you consent to the following conditions:

- The USG routinely intercepts and monitors communications on this IS for purposes including, but not limited to, penetration testing, COMSEC monitoring, network operations and defense, personnel misconduct (PM), law enforcement (LE), and counterintelligence (CI) investigations.
- At any time, the USG may inspect and seize data stored on this IS.
- Communications using, or data stored on, this IS are not private, are subject to routine monitoring, interception, and search, and may be disclosed or used for any USG authorized purpose.
- This IS includes security measures (e.g., authentication and access controls) to protect USG interests—not for your personal benefit or privacy.
- Notwithstanding the above, using this IS does not constitute consent to PM, LE or CI investigative searching or monitoring of the content of privileged communications, or work product, related to personal representation or services by attorneys, psychotherapists, or clergy, and their assistants. Such communications and work product are private and confidential. See User Agreement for details.

OK

[B. For Blackberries and other PDAs/PEDs with severe character limitations:]  
I've read & consent to terms in IS user agreem<sup>313</sup>t.

---

313. Memorandum from John Grimes, *supra* note 312, at Attachment 1.

Attachment 2 to the Banner and User Agreement policy contained the new mandatory User Agreement, which DoD users were required to sign:

**Figure 2: Standard Mandatory Notice and Consent Provision**

ATTACHMENT 2  
STANDARD MANDATORY NOTICE AND CONSENT PROVISION  
FOR ALL DOD INFORMATION SYSTEM USER AGREEMENTS

By signing this document, you acknowledge and consent that when you access

Department of Defense (DoD) information systems:

- You are accessing a U.S. Government (USG) information system (IS) (which includes any device attached to this information system) that is provided for U.S. Government authorized use only.
- You consent to the following conditions:
  - The U.S. Government routinely intercepts and monitors communications on this information system for purposes including, but not limited to, penetration testing, communications security (COMSEC) monitoring, network operations and defense, personnel misconduct (PM), law enforcement (LE), and counterintelligence (CI) investigations.
  - At any time, the U.S. Government may inspect and seize data stored on this information system.
  - Communications using, or data stored on, this information system are not private, are subject to routine monitoring, interception, and search, and may be disclosed or used for any U.S. Government-authorized purpose.
  - This information system includes security measures (e.g., authentication and access controls) to protect U.S. Government interests—not for your personal benefit or privacy.
    - Notwithstanding the above, using an information system does not constitute consent to personnel misconduct, law enforcement, or counterintelligence investigative searching or monitoring of the content of privileged communications or data (including work product) that are related to personal representation or services by attorneys, psychotherapists, or clergy, and their assistants. Under these circumstances, such communications and work product are private and confidential, as further explained below: Nothing in this User Agreement shall be interpreted to limit the user's consent to, or in any other way restrict or affect, any U.S. Government actions for purposes of network administration, operation, protection, or defense, or for communications security. This includes all communications and data on an information system, regardless of any applicable privilege or confidentiality.

- The user consents to interception/capture and seizure of ALL communications and data for any authorized purpose (including personnel misconduct, law enforcement, or counterintelligence investigation). However, consent to interception/capture or seizure of communications and data is not consent to the use of privileged communications or data for personnel misconduct, law enforcement, or counterintelligence investigation against any party and does not negate any applicable privilege or confidentiality that otherwise applies.
- Whether any particular communication or data qualifies for the protection of a privilege, or is covered by a duty of confidentiality, is determined in accordance with established legal standards and DoD policy. Users are strongly encouraged to seek personal legal counsel on such matters prior to using an information system if the user intends to rely on the protections of a privilege or confidentiality.
- Users should take reasonable steps to identify such communications or data that the user asserts are protected by any such privilege or confidentiality. However, the user's identification or assertion of a privilege or confidentiality is not sufficient to create such protection where none exists under established legal standards and DoD policy.
- A user's failure to take reasonable steps to identify such communications or data as privileged or confidential does not waive the privilege or confidentiality if such protections otherwise exist under established legal standards and DoD policy. However, in such cases the U.S. Government is authorized to take reasonable actions to identify such communication or data as being subject to a privilege or confidentiality, and such actions do not negate any applicable privilege or confidentiality.
- These conditions preserve the confidentiality of the communication or data, and the legal protections regarding the use and disclosure of privileged information, and thus such communications and data are private and confidential. Further, the U.S. Government shall take all reasonable measures to protect the content of captured/seized privileged communications and data to ensure they are appropriately protected.
- In cases when the user has consented to content searching or monitoring of communications or data for personnel misconduct, law enforcement, or counterintelligence investigative searching, (i.e., for all communications and data other than privileged communications or data that are related to personal representation or services by attorneys, psychotherapists, or clergy, and their assistants), the U.S. Government may, solely at its discretion and in

accordance with DoD policy, elect to apply a privilege or other restriction on the U.S. Government's otherwise-authorized use or disclosure of such information.

- All of the above conditions apply regardless of whether the access or use of an information system includes the display of a Notice and Consent Banner ("banner"). When a banner is used, the banner functions to remind the user of the conditions that are set forth in this User Agreement, regardless of whether the banner describes these conditions in full detail or provides a summary of such conditions, and regardless of whether the banner expressly references this User Agreement.<sup>314</sup>

Under the terms and conditions of the new Banner and User Agreement policy, use of a government information system constitutes consent for the government to access users' communications and data for any purpose, including "penetrations testing, communications security (COMSEC) monitoring, network operations and defense, personnel misconduct (PM), law enforcement (LE), and counterintelligence (CI) activities."<sup>315</sup> In addition, users consent that "[c]ommunications using, or data stored on, this information system are not private, are subject to routine monitoring, interception, and search, and may be disclosed or used for any U.S. Government-authorized purpose."<sup>316</sup>

By addressing both "privacy" and "law enforcement" simultaneously, the new Banner and User Agreement directly responded to the facts of *Long*. No longer could systems administrators testify of a "general policy to avoid examining e-mails and their content because it was a 'privacy issue.'"<sup>317</sup> No longer could courts find logon banners insufficient in notifying users of law enforcement monitoring and searching.<sup>318</sup> On its face, the policy extinguished any reasonable privacy expectation users had in their government computer activity, with the exception of activity related to privileged communications (i.e., attorneys, psychotherapists, clergy, and their assistants).<sup>319</sup> In addition, to

---

314. *Id.* at Attachment 2.

315. *Id.*

316. *Id.*

317. *United States v. Long*, 64 M.J. 57, 60 (C.A.A.F. 2006).

318. *See id.* at 64.

319. Memorandum from John Grimes, *supra* note 312, at Attachments 1 and 2. A discussion of the potential impact of the new Banner and User Agreement on privileged communications is beyond the scope of this Article. However, some commentators have expressed concern. *See, e.g.*, Edell, *supra* note 9, at 17; William H. McMichael, *Lanyers: DoD rule threatens confidential client e-mails*, ARMY TIMES, June 16, 2008, at 22. For a discussion of workplace e-mails and the attorney-client privilege, see Kelcey Nichols, *Hiding Evidence from the Boss: Attorney-Client Privilege and Company Computers*, 3 SHIDLER J.L. COM. & TECH. 6 (2006).

the extent a limited reasonable expectation of workplace privacy in a government computer still existed under *O'Connor*, the Banner and User Agreement policy established that law enforcement searches were based on user consent and, thus, did not require search authorization.<sup>320</sup>

The core question, of course, is whether the Banner and User Agreement policy structurally changed the communications privacy underpinnings that led the *Long* court to its holding, or if it simply rearranged the factual furniture. Certainly, response to the new logon banner was mixed. The legal division of the Air Force Office of Special Investigation announced that criminal search authorization was no longer required for government computers,<sup>321</sup> while attorneys from the Air Force Government Trial and Appellate Operations Division (who, interestingly, were the ones who successfully argued *Larson* before the CAAF) urged practitioners not to read *Larson* too broadly and to continue seeking search authorization for criminal searches of government information systems.<sup>322</sup> As discussed below, the jury is still out on this issue—and for good reason.

320. In her dissent in *Long*, Judge Crawford argued the search was permissible even assuming a reasonable expectation of privacy, because Lance Corporal Long had consented to the search by clicking the logon banner. *Long*, 64 M.J. at 70 (Crawford, J., dissenting). As discussed below, however, it is debatable whether clicking on a mandatory logon banner constitutes consent under MRE 314(e)(4). See *infra* text accompanying notes 352–54354.

321. In a background paper issued for Air Force attorneys, the legal division of AFOSI stated:

[W]ith exceptions for privileged communications, . . . [law enforcement] . . . investigators seizing, searching, and intercepting communications or data stored on 1) a bannered government computer in a government office lacking an expectation of privacy, or 2) a base network control center (BNCC) storing data transmitted from a bannered government computer, may do so without a criminal search authorization . . . in accordance with the revised DoD Notice and Consent Banner/User Agreement and findings in *Larson*.

AFOSI/JA, *Background Paper on the Revised DoD Notice and Consent Banner and U.S. v. Larson*, U.S. AIR FORCE, Jun. 19, 2008 (on file with author), *quoted in* Mendelson, *supra* note 9, at 9.

322. In her column, Captain Mendelson, an appellate attorney for the Air Force, cautioned practitioner's against adopting AFOSI's blanket approach:

We strongly advise against any blanket policy that a search authorization need not be sought. To the contrary, we advise that in every case a search authorization should be sought. We fully appreciate that *Larson* was a major victory for the Government, and that the newly revised DoD consent banner will address some of the issues present in *United States v. Long*, 64 M.J. 57 (C.A.A.F. 2006) (holding the servicemember had a reasonable expectation of privacy in her government email stored on a government server). However, we caution that *Larson* can be distinguished in future cases, particularly with respect to the Court's reliance on the fact that *Larson* presented no evidence that he enjoyed a *subjective* expectation

## B. GENERAL WARRANTS AND FOURTH AMENDMENT PROTECTIONS

Military courts have yet to address a case involving the new Banner and User Agreement.<sup>323</sup> However, a number of factors militate in favor of seeking search authorization for search and seizure of service members' workplace communications.<sup>324</sup> First, as previously discussed, the analytical shortcomings of the court's reasoning in *Larson*—even if ultimately leading to the correct conclusion—caution against giving this case too much precedential value. The court's opinion simply omits too many factual and legal signposts, including the *O'Connor* workplace privacy analysis, which could have led the court (or may lead the next court) to an altogether different destination. Second, as discussed below, it is at least questionable whether courts will uphold searches based on consent when, for military members, consent is arguably an unavoidable term and condition of employment. Like “general warrants” the Supreme Court has previously struck down, the new Banner and User Agreement policy could be argued to bestow *a priori* authority on DoD law enforcement officials to conduct broad, general searches to discover criminal misconduct. Third, as also discussed below, the Fourth Amendment requires, absent exigent circumstances, that “a neutral and detached” authority be interposed between the police and the public.<sup>325</sup> The new Banner and User Agreement policy may be viewed as violating this principle by authorizing *ex parte* criminal searches of potentially protected privacy expectations without judicial oversight or a commander's probable cause review. Finally, the voluntariness of service members' consent in agreeing to the new Banner and User Agreement policy is debatable. By virtue of the nature of military service, service members could claim little agency in choosing whether to accept or reject the policy's prospective, unbounded intrusions.

### 1. *General Warrants and the Particularity Requirement*

The Fourth Amendment, in addition to prohibiting “unreasonable searches and seizures,” states that “no [w]arrants shall issue, but upon probable cause, supported by [o]ath or affirmation, and particularly

---

of privacy in the contents of his government computer. . . . While *Larson* is a very favorable case for the Government, it does not create a blanket rule obviating the need for search authorization and it should not be treated as such.

Mendelson, *supra* note 9, at 9.

323. At the time of this writing, a search revealed only five cases citing to *United States v. Larson*, 66 M.J. 212 (C.A.A.F. 2008), none of which addressed communications privacy.

324. This is the same approach recommended by attorneys from the Air Force Trial and Appellate Division. See Mendelson, *supra* note 9, at 9.

325. *Johnson v. United States*, 333 U.S. 10, 14 (1948).

describing the place to be searched, and the person or things to be seized.”<sup>326</sup> The requirement for particularity “makes general searches . . . impossible and prevents the seizure of one thing under a warrant describing another. As to what is to be taken, nothing is left to the discretion of the officer executing the warrant.”<sup>327</sup> Both courts and commentators have observed that the inclusion of the particularity requirement in the Fourth Amendment was in response to the pre-Revolutionary use of “general warrants,” which, among other things, gave customs officials “blanket authority to conduct general searches for goods imported to the Colonies in violation of the tax laws of the Crown.”<sup>328</sup>

Judicial consideration of general warrants extends back at least as far as 1765, when, in *Entick v. Carrington*, Lord Camden reviewed whether a general warrant by Earl Halifax “to search for and seize the [papers of Mr. Entick]” was issued in contravention to law.<sup>329</sup> The warrant—which included all of Mr. Entick’s papers—had been issued under Earl Halifax’s executive authority as a principal secretary of state for England to investigate suspected cases of seditious libel.<sup>330</sup> In his opinion, Lord Camden first held that the secretary of state was not empowered to issue warrants, an authority reserved by law to justices of the peace.<sup>331</sup> Then, in an analysis at once both colorful and brilliant, Lord Camden (a) articulated the dangers of such unchecked power, (b) dissected the argument that “this power is essential to government, and the only means of quieting clamors and sedition,” and (c) explained the normative value inherent in individual privacy.<sup>332</sup> Although reflective of the strong connection between property interests and privacy rights in existence at the time, Lord Camden’s reasoning is instructive in our consideration of electronic communications:

Papers are the owner’s goods and chattels: they are his dearest property; and are so far from enduring a seizure, that they will hardly bear an inspection; and though the eye cannot by the laws of England be guilty of a trespass, yet where private papers are

---

326. U.S. CONST. amend. IV.

327. *Berger v. New York*, 388 U.S. 41, 58 (1967) (quoting *Marron v. United States*, 275 U.S. 192, 196 (1927)) (internal quotation marks omitted).

328. *Id.* (noting the use of general warrants “was a motivating factor behind the Declaration of Independence” and the Fourth Amendment “repudiated these general warrants”) (citations omitted); see also Freiwald, *supra* note 22, at 59 (2007) (discussing the framers’ concern with general warrants and citing Thomas Y. Davies, *Recovering the Original Fourth Amendment*, 98 MICH. L. REV. 547, 619–68 (1999)).

329. *Entick v. Carrington*, (1765) 95 Eng. Rep. 807 (K.B.).

330. *Id.*

331. *Id.*

332. *Id.*

removed and carried away, the secret nature of those goods will be an aggravation of the trespass, and demand more considerable damages in that respect. Where is the written law that gives any magistrate such a power? I can safely answer, there is none; and therefore it is too much for us without such authority to pronounce a practice legal, which would be *subversive of all the comforts of society*.<sup>333</sup>

Lord Camden then held that the general warrant to “seize and carry away the party’s papers in the case of a seditious libel [was] illegal and void.”<sup>334</sup>

Taking notice of *Entick v. Carrington*, the Supreme Court in *Berger v. New York* quoted Lord Camden’s reference to privacy intrusions as “subversive of all the comforts of society.”<sup>335</sup> Decided in 1967, *Berger* involved a New York statute which authorized “the issuance of the order, or warrant for eavesdropping, upon the oath of the attorney general, the district attorney or any police officer above the rank of sergeant stating that ‘there is reasonable ground to believe that evidence of crime may be thus obtained . . . .’”<sup>336</sup> Pursuant to the statute, New York law enforcement officials had obtained a warrant for electronic surveillance that led to the discovery of evidence of a conspiracy relating to the issuance of state liquor licenses.<sup>337</sup>

In striking down the New York law, the Supreme Court held it failed to meet the Fourth Amendment’s requirement for “particularization,” observing, “[i]t lays down no requirement for particularity in the warrant as to what specific crime has been or is being committed, nor ‘the place to be searched,’ or ‘the persons or things to be seized’ as specifically required by the Fourth Amendment.”<sup>338</sup> The Court then went on to explain that the purpose of the Fourth Amendment’s probable cause requirement is “to keep the state out of constitutionally protected areas until it has reason to believe that a specific crime has been or is being committed. . . .”<sup>339</sup> Like general warrants, the New York statute left “too much to the discretion of the officer executing the order.” Finally, the Court found additional cause for concern with the length of the two-month surveillance authorization of the warrant that effectuated “the equivalent of a series of intrusions, searches, and seizures pursuant to a single showing of probable cause,” the lack of a

---

333. *Id.* (emphasis added).

334. *Id.*

335. 388 U.S. 41, 49 (1967). Both *Berger* and *Katz* were both issued during the Court’s 1967 term.

336. *Id.* at 54.

337. *Id.* at 45.

338. *Id.* at 56.

339. *Id.* at 59.

termination date “once the conversation sought is seized,” and the lack of notice to the subject of the intrusion.<sup>340</sup>

For all intents and purposes, the Banner and User Agreement policy could be argued to function as a standing general warrant issued by an executive agency to search for evidence of general criminal misconduct. It goes well beyond the “monitoring” policies previously upheld in both civilian and military courts, which are soundly based on the “operational realities of the workplace” described by the Supreme Court in *O’Connor*<sup>341</sup> and easily connected to the authority in MRE 313 to conduct work-related inspections.<sup>342</sup> In effect, it permits the government to conduct continuous and unending law enforcement searches without meeting any of the requirements for particularity. To the extent that service members have a reasonable expectation of privacy in their workplace e-mail—and I do not argue that they always do—the policy appears to permit a privacy intrusion of the sort which Lord Camden found in 1765 to be subversive of “all the comforts of society,”<sup>343</sup> and which the Framers of the Constitution specifically intended to avoid.<sup>344</sup>

On the other hand, distinctions clearly exist between general warrants and the Banner and User Agreement policy. For example, one could argue there are substantive differences between entering a home to generally search and seize papers, and entering a military member’s e-mail account to generally monitor, intercept, search, and seize e-mails.<sup>345</sup> This is especially true given the different complexion of constitutional guarantees for military members and the distinctiveness of military service.<sup>346</sup> However, we should ask whether the difference in privacy expectations held by the individual being intruded upon is a matter of kind or simply one of degree. As previously discussed, service members do not cast aside their cloak of constitutional protections simply by entering military service.<sup>347</sup> Factoring in the possibility that (a) with web-based e-mail, service members may be

---

340. *Id.*

341. *O’Connor v. Ortega*, 480 U.S. 709, 717 (1987) (plurality opinion).

342. For a discussion of MRE 313, see *supra* text accompanying notes 48–51. *See also* Edell, *supra* note 9, at 23 (arguing “[t]he monitoring policy was consistent with an inspection under MRE 313 [as] an inspection directed at everyone using the network and subjecting everyone to the same level of scrutiny.”).

343. *Entick v. Carrington*, (1765) 95 Eng. Rep. 807 (K.B.).

344. *See supra* note 22.

345. After all, homes retain the highest Fourth Amendment protections, see *Silverman v. United States*, 365 U.S. 505, 511 (1961), while the workplace enjoys only limited Fourth Amendment protections. *See O’Connor*, 480 U.S. at 717–18.

346. *See supra* text accompanying notes 25–34.

347. *See supra* text accompanying notes 40–41.

accessing precisely the same e-mails from both home and work, (b) when deployed, service members may have no means of communication with family and friends other than computers and Internet access provided by the government,<sup>348</sup> and (c) military service regulations expressly authorize service members to use government information systems for limited personal use,<sup>349</sup> and the normative divide between communications privacy in the home and workplace—especially the military workplace—narrows considerably.

Analogizing the Banner and User Agreement policy to general warrants may be imperfect because service members are not subject to the same warrant requirements of the Fourth Amendment, and by extension, the particularity requirement. However, while the military is not subject to federal civilian warrant requirements,<sup>350</sup> MRE 315 does require both a probable cause review and a particular articulation of the property, evidence, or person to be seized, neither of which are incorporated into the Banner and User Agreement policy.<sup>351</sup>

Finally, one could argue that even if the Banner and User Agreement policy functioned as a standing general warrant, national security concerns provide enough particularity for its enforcement.<sup>352</sup> This argument is especially persuasive in light of the Fourth Circuit's holding in *United States v. Simons*, in which the internet policy of a division of the Central Intelligence Agency (a workplace vital to national security) effectively removed Mr. Simons' limited privacy interest in his workplace computer.<sup>353</sup> However, the Banner and User Agreement policy authorizes searches for any law enforcement purposes, not simply those linked to national security concerns. Moreover, as previously described, military courts in other contexts have

348. See Edell, *supra* note 9, at 23 (“Often e-mail is the only means of communication for deployed Soldiers.”)

349. See *supra* text accompanying note 11. Edell observes that [t]he Army and DOD further reinforce a Soldier's expectation of privacy in government e-mail by allowing personal use. . . . The Army has even touted [Army Knowledge Online] as a means for Soldiers to communicate with their families by offering spouses e-mail addresses and informing Soldiers how to send video messages with their e-mail accounts.

Edell, *supra* note 9, at 23.

350. See, e.g., *United States v. Chapman*, 954 F.2d 1352, 1367–70 (7th Cir. 1992) (finding while military procedures differ from federal civilian procedures, Fourth Amendment constitutional guarantees were upheld).

351. MIL. R. EVID. 315(b)(1), (f); see also *United States v. Hester*, 47 M.J. 461, 463 (C.A.A.F. 1998) (“To satisfy the *Fourth Amendment* and the Military Rules of Evidence, there must be probable cause for a search authorization, Mil.R.Evid. 315(f)(1), and the search authorization must be specific. Mil.R.Evid. 315(b)(1).”).

352. For a brief discussion of national security concerns requiring stringent DoD communications monitoring, see *supra* text accompanying notes 256–60.

353. *United States v. Simons*, 206 F.3d 392 (4th Cir. 2000).

rejected a bright line rule that service members have no workplace privacy expectations, effectively establishing that national security concerns alone do not obviate all military workplace privacy expectations.<sup>354</sup>

## 2. *Neutral and Detached Interposition*

Absent exigent circumstances, both the Fourth Amendment<sup>355</sup> and MRE 315(d) require law enforcement officials to obtain authorization from a neutral and detached authority before conducting a search requiring probable cause.<sup>356</sup> As the Supreme Court noted in *Johnson v. United States*,

[t]he point of the Fourth Amendment, which often is not grasped by zealous officers, is not that it denies law enforcement the support of the usual inferences which reasonable men draw from evidence. Its protection consists in requiring that those inferences be drawn by a neutral and detached magistrate instead of being judged by the officer engaged in the often competitive enterprise of ferreting out crime.<sup>357</sup>

In the military, such search authority may be issued by a civilian magistrate in the form of a warrant, by a commander “who has control over the place where the property or person to be searched is situated or found,” or by a military judge.<sup>358</sup>

In certain exigent circumstances, of course, law enforcement officials may dispense with the requirement to obtain a search warrant. In *Chapman v. United States*, the Supreme Court considered whether Georgia law enforcement officials properly entered a rented home when they did so with the landlord’s consent but without a warrant.<sup>359</sup> Holding that the tenant’s Fourth Amendment rights had been violated, the Court acknowledged that exceptional circumstances exist in which the warrant requirement may be obviated.<sup>360</sup> On the facts of the case, however, the Court found that “inconvenience to the officers and some slight delay necessary to prepare papers and present the evidence to a magistrate” did not constitute exigent circumstances.<sup>361</sup>

---

354. *See supra* text accompanying notes 83–88.

355. *See Johnson v. United States*, 333 U.S. 10, 14 (1948).

356. *See* MIL. R. EVID. 315(d).

357. 333 U.S. at 13–14, *quoted in* *Chapman v. United States*, 365 U.S. 610, 614–15 (holding warrantless search of rented home unlawful).

358. MIL. R. EVID. 315(b)(2), (d)(1)–(2).

359. 365 U.S. at 610–11.

360. *Id.* at 615.

361. *Id.* (quoting *Johnson v. United States*, 333 U.S. 10, 15 (1948)) (internal quotation marks omitted). The Supreme Court noted, “[n]o suspect was fleeing or likely to take flight,”

With respect to the new Banner and User Agreement policy, it is difficult to see what exigent circumstances warrant the DoD to extend its authority from routine work-related monitoring of its information systems to warrantless law enforcement searches.<sup>362</sup> In support of the Banner and User Agreement policy, perhaps it could be argued that the viability of electronic evidence is threatened by the logistical difficulty in obtaining search authorization. This seems unlikely. The scope of persons authorized to conduct probable cause reviews in the military (civilian magistrates, commanders, and military judges) is broader than it is in the civilian world, meaning there are more people to provide authorization in emergency situations.<sup>363</sup> By virtue of their military association, commanders and military judges are normally accessible to military law enforcement officials twenty-four hours a day, 365 days a year, regardless of duty location. Moreover, unlike civilian jurisprudence, the MRE do not require a sworn affidavit (although it is recommended) before search authorization is granted, and the authorization itself may be issued orally or in writing.<sup>364</sup>

Proponents of the Banner and User Agreement policy also could argue that the “operational realities” of military service require unrestricted access to otherwise constitutionally protected privacy interests for purposes of criminal investigations involving national security. As previously noted, constitutional protections do “take on a different complexion” when applied to members of the military,<sup>365</sup> in part because the military workplace “can range from an office building to a bunker or tent in a combat zone.”<sup>366</sup> Furthermore, protection of national security assets from state and non-state actors is justifiably a critical concern.<sup>367</sup> However, the Banner and User Agreement policy goes beyond national security concerns by permitting

---

the search “was of permanent premises, not of a movable vehicle,” and “[n]o evidence or contraband was threatened with removal or destruction. . . .” *Id.* (quoting *Johnson v. United States*, 333 U.S. at 15) (internal quotation marks omitted).

362. Some may criticize the characterization that the Banner and User Agreement policy extended, rather than reaffirmed, the scope of search authority which existed prior to *Long*. Certainly, the Air Force General Counsel Memorandum took the position that search authorization was not required prior to *Long* based on users’ consent to the logon banner. *See* Memorandum from W. Kipling At Lee, Jr., *supra* note 257. Other branches of the service disagreed, however. Edell, *supra* note 9, at 23 (“Prior to the decision in [*United States v. Long*], the Army specifically ensured that Soldiers had a reasonable expectation of privacy from law enforcement during systems monitoring.”).

363. MIL. R. EVID. 315(b)(2), (d)(1)–(2).

364. MIL. R. EVID. 315(b)(1).

365. *United States v. Allen*, 1999 CCA LEXIS 116, at \*11 (A.F.C.C.A. April 22, 1999); *see also supra* text accompanying notes 25–34.

366. *Long*, 64 M.J. at 62.

367. *See supra* text accompanying note 152.

warrantless law enforcement searches for any purpose, essentially making permissible what arguably had been impermissible prior to the new policy. Moreover, even some military attorneys, who certainly are aware of national security needs of the military, argue for continued reliance on search authorization.<sup>368</sup> While “operational realities” may certainly justify stringent monitoring, perhaps we should ask whether they necessarily require warrantless searches that invade otherwise reasonable expectations of privacy. As the Supreme Court found in *Chapman*, “inconvenience to the officers and some slight delay necessary to prepare papers and present the evidence to a magistrate . . . are never very convincing reasons and, in these circumstances, certainly are not enough to by-pass the constitutional requirement.”<sup>369</sup>

### 3. *Voluntariness of Consent*

One of the arguments advanced in support of warrantless law enforcement searches of government e-mail accounts is user consent.<sup>370</sup> By clicking the logon banner, proponents argue that service members have consented to the terms of the logon banner and, by extension, the scope of any search defined in the logon banner.<sup>371</sup> Those who sign user agreements are similarly agreeing to the scope of any search defined in the user agreement. In the case of the new DoD Banner and User Agreement policy, the scope of permissive searches is “any U.S. Government-authorized purpose[.]” which includes both “personnel misconduct” and “law enforcement . . . .”<sup>372</sup> Given the breadth of such language, the government has a strong argument that service members have provided a priori consent to general or specific searches for criminal evidence, even assuming *arguendo*

---

368. See Edell, *supra* note 9, at 24; Mendelson, *supra* note 9, at 9.

369. *Chapman v. United States*, 365 U.S. 610, 615 (1961) (internal quotation marks omitted) (citation omitted).

370. The Military Rules of Evidence provide “[s]earches may be conducted of any person or property with lawful consent.” MIL. R. EVID. 314(e)(1).

371. See, e.g., Long, 64 M.J. at 67–68 (Crawford, J., dissenting); Coacher, *supra* note 152, at 176–77 (“Under most circumstances, use of a government computer constitutes actual consent to systems protection monitoring. . . . Even if the act of accepting the conditions in the monitoring banner does not constitute actual consent, sufficient circumstances should exist to find implied consent . . . .”); AFOSI/JA, *supra* note 321, at 1 (“This revised DoD Notice and Consent Banner/User Agreement provides the consent necessary to satisfy the Federal Wiretap Act . . . and the Stored Communications Act . . . .”); Memorandum from W. Kipling At Lee, Jr., *supra* note 257, at 1 (“At the time of login on each properly configured Air Force computer an electronic banner appears. . . . Courts have consistently held that consent to monitoring eliminates any expectation of privacy.”).

372. Memorandum from John Grimes, *supra* note 312, at Attachment 2.

they retain a reasonable expectations of privacy in their workplace communications.

This argument is not without considerable appeal. In her dissenting opinion in *United States v. Long*, Judge Crawford vigorously argued that even if Lance Corporal Long held a reasonable expectation of privacy, search authorization was unnecessary because she had consented to the logon banner, which, in Judge Crawford's view, contained language broad enough to encompass law enforcement searches:

Even when there is a reasonable expectation of privacy, one of the exceptions is consent to search. Consent is such that one would not rely upon an assumption of risk that the service provider would not reveal this information to law enforcement officials. . . . Certainly, a communicator's expectation of privacy is not reasonable once he or she has given consent to search. . . . Where consent is given to an administrator or someone with mutual use of the property, the originators of e-mail assume the risk that the administrator may give consent to law enforcement officials.<sup>373</sup>

Judge Crawford thus identified two types of consent which may be relevant in evaluating a claim of e-mail privacy: the consent of the user and the consent of the systems administrator.<sup>374</sup> By agreeing to the logon banner, the user essentially assumes the risk that the systems administrator will consent to searches by law enforcement officials.<sup>375</sup> This is especially true, one might argue, if the banner explicitly mentions law enforcement searches, as does the new Banner and User Agreement policy.

The merits of this argument rise or fall on the validity of the consent. Under MRE 314, "consent must be given voluntarily," a question derived from all the circumstances.<sup>376</sup> If the voluntariness of consent is questioned at trial, the government must "show by clear and convincing evidence that such consent was, in fact, freely and voluntarily given."<sup>377</sup> Importantly, "[m]ere submission to the color of authority of personnel performing law

---

373. 64 M.J. at 70 (Crawford, J., dissenting) (internal citations omitted).

374. *Id.* Even if service members do not possess a reasonable expectation of privacy in their government e-mail, the SCA still requires law enforcement agents to obtain "service provider" consent from appropriate military officials. This assumes, of course, the military is a non-public service provider. *See* 18 U.S.C. § 2701(a); *see also* Kerr, *supra* note 115, at 1220 n.82.

375. *Long*, 64 M.J. at 70.

376. MIL. R. EVID. 314(e)(4); *see also* *United States v. Middleton*, 10 M.J. 123, 132–33 (C.M.A. 1981).

377. *Middleton*, 10 M.J. at 132 (quoting *Schneckloth v. Bustamonte*, 412 U.S. 218, 222 (1973)) (internal quotation marks omitted); *see also* Coacher, *supra* note 152, at 176 (discussing actual and implied consent by use of government e-mail).

enforcement duties or acquiescence in an announced or indicated purpose to search is not voluntary consent.”<sup>378</sup>

Whether the government can prove by clear and convincing evidence that military service members have “freely and voluntarily” consented to the scope of searches in the new Banner and User Agreement policy remains to be seen. First, the government must overcome the argument that consent is hardly consensual when it is mandated as a term and condition of an employment relationship the employee is bound to continue.<sup>379</sup> Unlike their civilian counterparts, military members generally are obligated for a term of service and are not free to separate when they no longer approve of their employment conditions.<sup>380</sup> To leave is desertion, a criminal offense under Article 85, UCMJ, and to refuse to consent to the Banner and User Agreement policy (which, in turn, could prevent service members from accessing their computers and fulfilling their duties), could be construed as dereliction of duty in violation of Article 92, UCMJ. Second, to the extent service member consent is based on a contract theory of agreement, perhaps similar to the contracts at issue in *United States v. Hart*<sup>381</sup> and *United States v. Maxwell*,<sup>382</sup> the government must overcome the negative inference created by the inequality of bargaining power between the government and service members. As employees bound to their employer, service members have little choice in accepting the terms of the Banner and User Agreement policy.

378. MIL. R. EVID. 314(e)(4); *Middleton*, 10 M.J. at 133 (“An additional—and weighty—factor to be considered here is that for the Government to carry successfully its burden of proving that the ‘consent’ was, in fact, freely and voluntarily given, it must show that there was more than mere acquiescence to a claim of lawful authority.”) (citations omitted).

379. Although military workplace privacy law generally follows civilian workplace privacy law, one of the distinctions is that service members are not free to leave their employ if they no longer agree with the terms and conditions of employment. As a result, civilian cases involving employees who “consent” to logon banners are only partially persuasive. In those cases, actual or implied consent can be imputed into the employee’s decision to remain with the employer. In the military, it cannot. *See* AFOSI/JA, *supra* note 321, at 1 (citing only civilian cases to establish that employees may consent to searches by clicking logon banners).

380. *See In re Grimley*, 137 U.S. at 151–52. The Court held that

Enlistment is a contract; but it is one of those contracts which changes the status; and, where that is changed, no breach of the contract destroys the new status or relieves from the obligations which its existence imposes. . . . By enlistment the citizen becomes a soldier. His relations to the State and the public are changed. He acquires a new status, with correlative rights and duties; and although he may violate his contract obligations, his status as a soldier is unchanged. He cannot of his own volition throw off the garments he has once put on . . . .

*Id.*

381. 2009 U.S. Dist. LEXIS 72473 (W.D. Ky. Aug. 17, 2009).

382. 45 M.J. 406 (C.A.A.F. 1996).

As such, courts could construe the Banner and User Agreement policy as a non-binding contract of adhesion.<sup>383</sup>

C. THE NEED FOR A NORMATIVE APPROACH

Foundationally, Fourth Amendment privacy analysis requires a normative inquiry into what “society is prepared to recognize as ‘reasonable.’”<sup>384</sup> This is not a mechanical factual analysis into the limited privacy expectations a person has accepted as a cost of doing business as a citizen. Rather, as Justice Marshall observed in his dissent in *Smith v. Maryland*, it is an inquiry into “the risks [a person] *should* be forced to assume in a free and open society.”<sup>385</sup> Justice Blackmun, writing the majority opinion in *Smith*, acknowledged as much when he conceded the possibility of situations in which the *Katz* two-part inquiry “would provide an inadequate index of *Fourth Amendment* protection,” such as a general government announcement “that all homes henceforth would be subject to warrantless entry . . . .”<sup>386</sup> In that extreme situation, in which subjective privacy expectations had been obviated by “influences alien to well-recognized *Fourth Amendment* freedoms,” a person’s subjective expectation of privacy “could play no meaningful role in ascertaining what the scope of *Fourth Amendment* protection was” and a “normative inquiry would be proper.”<sup>387</sup>

To a significant degree, the DoD’s Banner and User Agreement policy amounts to a government announcement of the sort of warrantless search envisioned by Justice Blackmun in *Smith*. The Banner and User Agreement policy notifies DoD users that the government can search and seize otherwise protected communications for any and all purposes, including law enforcement, effectively quashing all subjective privacy expectations service members may have in their workplace communications. As a result, service members can never satisfy the *Katz* two-part inquiry. Absent the normative approach endorsed by Justice Marshall and notionally suggested by Justice Blackmun in *Smith*, service members’ Fourth Amendment protections in their workplace electronic communications would be utterly extinguished.

Critics may object to my comparing Justice Blackmun’s Orwellian vision to the Banner and User Agreement policy, arguing that the policy—and any searches conducted pursuant to it—is based on consent rather than executive

---

383. See BLACK’S LAW DICTIONARY 341(8th ed. 2004).

384. *Katz v. United States*, 389 U.S. 347, 361 (1967) (Harlan, J., concurring); see also *Minnesota v. Olson*, 495 U.S. 91, 97 (1990).

385. *Smith v. Maryland*, 442 U.S. 735, 750 (1979) (Marshall, J., dissenting) (emphasis added).

386. *Id.* at 741 n.5.

387. *Id.*

fiat.<sup>388</sup> As an “agreement” between parties, they might suggest, the policy is wholly unlike Justice Blackmun’s unilateral government announcement “that all homes henceforth would be subject to warrantless entry . . . .”<sup>389</sup> However, there are two similarities. In theory, citizens who disagreed with the government’s unilateral announcement about the warrantless search of their homes could, if they chose to do so, simply leave the country for a social contract elsewhere. Similarly, service members who disagree with the Banner and User Agreement policy could, if they choose to do so, refuse to sign the agreement and risk disciplinary action. In either case, however, the costs are so high as to render any implied consent (in the case of the citizen who chooses to remain in the country) or actual consent (in the case of the service member who accedes to the policy) anything but “freely and voluntarily” given.<sup>390</sup>

Even if the government *can* establish such an agreement, the normative question begging to be asked is *should* it? Applying Justice Marshall’s reasoning to service members, what risks *should* service members be forced to assume as they serve their country? Is it reasonable for these risks to include communications privacy expectations that are not simply diminished, like those in the civilian workplace, but fully extinguished? Does it matter that service members work in foreign and remote environments where the only communications access they have is provided by the government?

While not always explicit, the tenor of analysis in the *Maxwell*, *Monroe*, and *Long* decisions suggests these were precisely the types of questions the CAAF was asking as it reflected on the privacy interests inherent in electronic communications. In *Maxwell*, the CAAF presciently observed the vital nature of electronic communications in modern society:

---

388. For example, in her dissent in *Long*, Justice Crawford specifically focuses on Lance Corporal Long’s consent to the logon banner. *United States v. Long*, 64 M.J. 57, 70 (C.A.A.F. 2006) (Crawford, J., dissenting). The Air Force General Counsel likewise relied on the Air Force logon banner in its opinion that users had consented to any searches for law enforcement purposes. Memorandum from W. Kipling At Lee, Jr., *supra* note 257.

389. *Smith*, 442 U.S. at 741 n.5.

390. *United States v. Middleton*, 10 M.J. 123, 132 (C.A.A.F. 1981) (internal citation omitted). For additional discussion of consent, see *supra* text accompanying notes 370–75. Critics may also object that Justice Blackmun’s hypothetical is based on the warrantless search of homes, not workplace information systems. To be clear, I am not arguing that warrantless searches of government computers or workplace electronic communications are equivalent to warrantless searches of homes, which retain the highest Fourth Amendment protections. See *Silverman v. United States*, 365 U.S. 505, 511 (1961). I am simply offering the unsurprising suggestion that a normative inquiry is requisite when the government pierces established Fourth Amendment protections through an agreement to which the other party is not in a position to disagree.

New technologies create interesting challenges to long established legal concepts. Thus, just as when the telephone gained nationwide use and acceptance, when automobiles became the established mode of transportation, and when cellular telephones came into widespread use, now personal computers, hooked up to large networks, are so widely used that the scope of Fourth Amendment core concepts of “privacy” as applied to them must be reexamined. Consequently, this opinion and the ones surely to follow will affect each one of us who has logged onto the “information superhighway.”<sup>391</sup>

The court went on to discuss the risks a user accepts in sending e-mail messages over a network by comparing e-mail to first-class mail and telephone calls.<sup>392</sup> With first-class mail and telephone calls, the originator bears the risk that the recipient will “reveal what is said to others.”<sup>393</sup> The originator can reasonably expect, however, “the contents to remain private and free from the eyes of the police absent a search warrant founded upon probable cause.”<sup>394</sup> Similarly, the sender of an e-mail bears the risk of an e-mail being revealed by the recipient, but “enjoys a reasonable expectation that the initial transmission will not be intercepted by the police.”<sup>395</sup> In *Monroe*, the CAAF expanded the e-mail sender’s risk to include the risk of monitoring by systems administrators.<sup>396</sup> Even there, however, the CAAF noted, “[t]he transmitter of an e-mail message enjoys a reasonable expectation that police officials will not intercept the transmission without probable cause and a search warrant.”<sup>397</sup> This proposition was tested—and upheld—in *Long*, in which the CAAF found that the “sole purpose of seizing the e-mails was to search for evidence of misconduct.”<sup>398</sup>

The normative principles drawn from these cases are clear. First, despite distinctions, electronic communications merit privacy expectations similar to first-class mail and telephone. Second, users of electronic communications

---

391. *United States v. Maxwell*, 45 M.J. 406, 410 (C.A.A.F. 1996). As of 2008, nearly 75% of U.S. households had Internet access, approaching the ubiquity of telephone service. THE NIELSEN COMPANY, HOME TECHNOLOGY FACTBOOK, DEVICE AND SERVICE ADOPTION RATES (2008), [http://en-us.nielsen.com/forms/report\\_forms/Nielsen\\_Home\\_Technology\\_Factbook](http://en-us.nielsen.com/forms/report_forms/Nielsen_Home_Technology_Factbook). Congress has commented on the essential nature of information and communications services. *See* H.R. CONF. REP. NO. 109-702, at 840 (noting that “restoring basic information and communications capacity is a fundamental element of humanitarian and civic assistance” for civil military operations conducted in foreign locations).

392. *Maxwell*, 45 M.J. at 417–18.

393. *Id.* at 418 (citation omitted).

394. *Id.* at 417 (citation omitted).

395. *Id.* at 418.

396. *United States v. Monroe*, 52 M.J. 326, 330 (C.A.A.F. 2000).

397. *Id.*

398. *United States v. Long*, 64 M.J. 65 (C.A.A.F. 2006).

should bear some risk—like telephone or first-class mail users—that the recipient of the communication will disclose it to others. Third, users of electronic communications should bear the additional risk that systems administrators will monitor the content of their communications when they have been notified of such monitoring by contract, user agreement, or logon banner. Fourth, given the vital role of electronic communications in society, a normative distinction between search and seizure of e-mails for work-related and law enforcement purposes is wholly appropriate. Users who bear the risk of intrusion by systems administrators should not be made—absent exigent circumstances—to bear the additional risk of intrusion by law enforcement. Instead of simply asking what *is* protected, policy makers and courts should ask what *should be* protected.

Based on these normative principles, I propose the following: (a) the DoD revise its Banner and User Agreement policy, providing robust system protection through monitoring, but omitting the authority to conduct warrantless searches for general law enforcement purposes, and (b) the CAAF limit its holding in *Larson* to its facts and require the DoD, like other federal agencies, to seek appropriate search authorization whenever its primary purpose in searching service members' electronic communications is law enforcement.

For a number of reasons, recognizing a limited privacy interest from law enforcement searches in service members' electronic communications could serve the interests of both service members and the military. First, it could boost morale among service members, who could communicate with friends and family from remote, deployed locations without fear that “big brother” is perusing every communication for signs of misconduct. Second, it would provide service members with workplace privacy protections similar to those as their civilian counterparts, who work alongside service members in preserving national security but who enjoy certain workplace privacy expectations from law enforcement, as opposed to work-related, searches. Third, it could promote efficiency of resources by forcing federal investigators to focus on specific crimes, rather than trolling through vast storehouses of communications for general evidence of criminal activity. Fourth, it could increase the quality of criminal evidence by interposing a neutral and detached probable cause search authority between investigators and the evidence they hope to search and seize. Fifth, it would honor the implicit commitment to privacy the military services have made to military family and friends in providing e-mail and instant messaging capabilities to communicate with deployed loved ones. Sixth, it would bring military judge advocates into the investigative process sooner rather than later, helping

investigators focus on criminal evidence which is particular, relevant, and admissible. Finally, it would recognize that service members, who take an oath to uphold the Constitution, are protected by those same constitutional guarantees. As Blackstone observed, “he puts not off the citizen when he enters the camp; but it is because he is a citizen, and would wish to continue so, that he makes himself for a while a soldier.”<sup>399</sup>

## V. CONCLUSION

It remains to be seen whether the CAAF’s next communications privacy case will follow the reasoning in *Larson*, or return to the pre-*Larson* pattern of distinguishing work-related searches from those of law enforcement. It also remains to be seen whether the CAAF will entertain a constitutional challenge to the DoD’s new Banner and User Agreement policy and, if so, will find it an appropriate exercise of executive authority. I have argued that the policy unnecessarily violates constitutional protections by expanding the DoD’s search authority from mere systems monitoring, which is clearly work-related, to law enforcement, which the Supreme Court distinguished in *O’Connor*. In my view, there seems little to lose and much to preserve by adopting the view that service members, like their civilian counterparts, have a reasonable expectation of privacy from law enforcement searches in electronic communications sent over a government information system. Service members are deployed around the world in support of their country and often have no systems of communication to use other than those provided by the government. In that case, providing the limited workplace privacy protections recognized by the Supreme Court in *O’Connor* seems appropriate, especially given the normative importance of electronic communications in contemporary society. Doing so adheres to case law and ensures service members are protected by the very Constitution they are sworn to defend.

---

399. *United States v. Culp*, 14 U.S.C.M.A. 199, 206 (C.M.A. 1963) (quoting WILLIAM BLACKSTONE, COMMENTARIES \*408 (Wendell ed.)) (internal quotation marks omitted).



# THE IMPERFECT IS THE ENEMY OF THE GOOD: ANTICIRCUMVENTION VERSUS OPEN USER INNOVATION

Wendy Seltzer<sup>†</sup>

## TABLE OF CONTENTS

I.	<b>INTRODUCTION</b> .....	910
II.	<b>THE MECHANICS OF CODE AND LAW</b> .....	920
	A. BASIC TECHNOLOGY OF “DIGITAL RIGHTS MANAGEMENT”.....	920
	B. THE MECHANICS OF ANTICIRCUMVENTION LAW.....	923
	C. LICENSING FOR ROBUSTNESS: HOW CONTENT PRODUCERS EXERT INFLUENCE IN THE HARDWARE MARKET.....	927
	D. OPEN SOURCE SOFTWARE IS INCOMPATIBLE WITH ROBUSTNESS REQUIREMENTS.....	930
III.	<b>THE ACADEMIC DEBATE</b> .....	934
	A. ANTICIRCUMVENTION’S ADVOCATES.....	934
	B. ANTICIRCUMVENTION’S CRITICS.....	937
	1. <i>Anticircumvention Stops End-User Fair Use</i> .....	937
	2. <i>DRM Does Not Stop Copying</i> .....	940
	3. <i>Anticircumvention Hinders Technology Innovation</i> .....	941
IV.	<b>ANTICIRCUMVENTION’S APPLICATION</b> .....	943
	A. CD VERSUS DVD: THE EFFECT OF A LOCKED-DOWN MEDIA FORMAT.....	943
	B. SDMI AND THE FREEDOM TO TINKER.....	946
	C. AN ALTERNATIVE: NON-ROBUST ADVISORY MEASURES.....	950
V.	<b>NEW CRITIQUE: ANTICIRCUMVENTION TAXES OPEN DEVELOPMENT AND USER INNOVATION</b> .....	953
	A. THE HIDDEN COSTS OF DRM.....	958
	B. DRM LIMITS DISRUPTIVE INNOVATION.....	959

---

© 2010 Wendy Seltzer. Reproduction permitted under Creative Commons Attribution 3.0 License, <http://creativecommons.org/licenses/by/3.0>.

<sup>†</sup> Berkman Center for Internet & Society at Harvard University, and Silicon Flatirons Center at University of Colorado Law School, [wendy@seltzer.org](mailto:wendy@seltzer.org). The Author thanks many helpful discussants, including Yochai Benkler, Michael Carroll, Rashmi Dyal-Chand, Peter Jaszi, Fred von Lohmann, Benjamin Mako Hill, Paul Ohm, Betsy Rosenblatt, Seth David Schoen, Eric von Hippel, Jonathan Zittrain, and participants at the Telecommunications Policy Research Conference, Northeastern Law School’s Colloquium, and Berkman Center for Internet & Society luncheon.

C.	DRM LIMITS USER INNOVATION.....	964
D.	THE OVER-ARCHING COST: DRM CENTRALIZES INNOVATION, OPPOSING COPYRIGHT'S ORIGINAL GOALS.....	969
VI.	CONCLUSION .....	971

## I. INTRODUCTION

Imagine yourself a movie mogul wishing to use the latest and greatest in technology to protect your new release, “Pirates of the Caribbean, the Prequel.”

You can, of course, limit the film to display in high-tech theaters, with bonded security guards scanning the audience for “camers” and guarding the movie’s physical path to the projection booth.<sup>1</sup> Before it gets there, you’ll have to make sure everyone working on the film has the appropriate incentives to keep draft cuts from leaking pre-release.<sup>2</sup> Even then, the movie will likely leak to the streets through some chink in that armor (especially if it is as popular as you hope it will be). With luck, however, you have bought yourself some time and built enough buzz that people want to see the film in theaters even if they could get grainy copies to watch on small screens. You might even enhance the differential by showing in IMAX or 3D, creating an *experience* that is hard to replicate even as the bits are copied.<sup>3</sup>

Once its theatrical run winds down, though, you hope to exploit your investment further with a rental and sales run.<sup>4</sup> You could rent and sell unrestricted digital copies, relying on straight copyright law’s prohibitions on commercial-scale reproduction, distribution, and public performance, but you want technological backup. So, to “keep honest people honest,” you want to add “copy protection,” wrapping the movie in either encryption or contract, or both.

1. See, e.g., Dawn C. Chmielewski, *Secrecy cloaked ‘Dark Knight,’* L.A. TIMES, July 28, 2008, at C1. Compare these protections with the more limited release-window protection Disney had for “Snow White.” DAVID NIMMER, COPYRIGHT: SACRED TEXT, TECHNOLOGY, AND THE DMCA 17–18 (2003).

2. See Simon Byers et al., *Analysis of Security Vulnerabilities in the Movie Production and Distribution Process*, in DIGITAL RIGHTS MANAGEMENT WORKSHOP (Moti Yung ed., 2003) (finding that “insider attacks” accounted for more than three-quarters of leaks, many of those prior to the DVD release).

3. See Mark Milian, *Which ‘Avatar’ to see? A look at IMAX, Dolby 3-D, RealD (and, yeah, boring old 2-D)*, L.A. TIMES BLOG, Dec. 29, 2009, <http://latimesblogs.latimes.com/herocomplex/2009/12/which-avatar-to-see-a-look-at-imax-dolby-3d-reald-and-boring-old-2d.html>.

4. See David Waterman et al., *Enforcement and Control of Piracy, Copying, and Sharing in the Movie Industry*, 30 REV. INDUS. ORG. 255, 258 tbl.1 (2007) (discussing release windows).

Encryption could prevent an unwanted user—one without the key—from being able to play the bits as a movie, no matter how many copies of them he makes.<sup>5</sup> An encrypted movie looks like a random string of ones and zeros when read off its disc—to the buyer, as well as to everyone else. But, unless you give the buyer the key, he's left with just a coaster (which he can buy more cheaply elsewhere). So you must also give the purchaser the means of decryption. If you hand him the key straight out, however, he once again possesses all the information necessary to make copies.

Instead of giving the secret key to the user directly, then, you entrust it to a “black-box” emissary,<sup>6</sup> a software program or hardware unit that you restrict to playing the movie in the form you have deemed permissible: play but do not copy, for example. Now that you have shared the secret with software or hardware, however, you have to guard that software or hardware as zealously as you did the original movie, preventing the key or the decrypted movie from leaking or being hacked out.<sup>7</sup> First, you've increased the number of assets to protect: neither the work *nor its decryption key* must be allowed to leak. Second, you've just shifted the locus of trust, not removed it. You still need to let the user see the movie, and you've determined you don't trust him (hence the need for DRM), but in order to trust the software or hardware manufacturer, you must make the manufacturers obey you rather than their products' user/owner. Simultaneously, you must take measures to prevent the user either from copying the physical object<sup>8</sup> or from obtaining any hardware or software that will play a copied object.<sup>9</sup>

---

5. If you use a known strong algorithm, properly, you can be assured that no one without the key will be able to decrypt it, at least not with computing power available today. See BRUCE SCHNEIER, *APPLIED CRYPTOGRAPHY: PROTOCOLS, ALGORITHMS, AND SOURCE CODE IN C* 152 (1996).

6. “Black-box” refers to the opacity of the processing between encrypted input and movie output. While the user can see the encrypted blob go in and get a movie in viewable format on the other end, he can't see what goes on in between/during the time the movie is decrypted.

7. See, e.g., Ellen Messmer, *Black Hat: Researcher Claims Hack of Processor Used to Secure Xbox 360, Other Products*, NETWORK WORLD, Feb. 2, 2010, <https://www.networkworld.com/news/2010/020210-black-hat-processor-security.html>; Andy Patrizio, *Why the DVD Hack Was a Cinch*, WIRED, Nov. 2, 1999, available at <http://www.wired.com/science/discoveries/news/1999/11/32263> (describing reverse engineering of the Xing software DVD decoder implicated in the first break of the DVD Content Scramble System).

8. The DVD format is controlled through a combination of patents on the format, copyright with anticircumvention, and even trademark. *RealNetworks, Inc. v. DVD Copy Control Ass'n*, 641 F. Supp. 2d 913, 920 (N.D. Cal. 2009) (describing the DVD-CCA and its technology). See generally DEAN S. MARKS & BRUCE H. TURNBULL, *TECHNICAL PROTECTION MEASURES: THE INTERSECTION OF TECHNOLOGY, LAW AND COMMERCIAL LICENSES* (1999), [http://www.wipo.int/edocs/mdocs/copyright/en/wct\\_wppt\\_imp/wct\\_wppt\\_](http://www.wipo.int/edocs/mdocs/copyright/en/wct_wppt_imp/wct_wppt_)

You must, however, create this secure ecosystem with enough compatibility that users are attracted, *and* with enough control that content purveyors are assured their directives are being followed. After all, the users themselves are already in an environment before your movie comes along—they may have home computers, televisions, home theaters, home networks, video iPods, or iRivers. They are unlikely to want your single movie enough to retool all these systems around it or to buy a special-purpose viewing device just for it.<sup>10</sup> So your plan works best if you can take advantage of existing platforms. If those are insufficiently secure, you can try to engage a significant segment of the industry to work with you to move people simultaneously to a new standard (with the assistance of antitrust counsel to ensure that cross-industry collaboration is seen to expand the market, not control it).<sup>11</sup>

Thus you might, if you were a movie studio with enough market clout to be persuasive, propose to fellow studios, consumer electronics companies, and software developers that they jointly support a copy-protection scheme around a new digital format for video distribution, setting licensing terms for the use of and interoperability with that format.<sup>12</sup> Together, you could hope to achieve the market saturation that would make your format successful: only licensed “SuperDisc” players could show the latest and greatest Hollywood films in full digital glory. The virtual network thus created would make your usage rules seem to be natural complements to the new format rather than user-hostile restrictions of consumer rights. You would be able to hold companies to your licensing scheme with the threat of Digital Millennium Copyright Act (DMCA) liability if they defected; erstwhile

---

imp\_3.pdf.

9. See Wendy Seltzer, *The Broadcast Flag: It's Not Just TV*, 57 FED. COMM. L.J. 209, 210 (2005) (describing the Broadcast Flag Rule's combined mandate of watermarking detection and obedience to flagged commands).

10. In special cases, that condition may be malleable. In 2005, the Motion Picture Academy sent 6,000 special-purpose limited-function DVD players to the Academy Awards screeners—a limited number of people designated to receive movies before their general DVD release. Even then, the limitations annoyed viewers and copies leaked. See Gary Gentile, *Studios eye new anti-piracy technology*, USA TODAY, July 2, 2004, [http://www.usatoday.com/tech/news/2004-07-02-anti-piracy\\_x.htm](http://www.usatoday.com/tech/news/2004-07-02-anti-piracy_x.htm); Britt Leach, *Screeners for My Consideration*, Veritas (Nov. 30, 2007), <http://www.veritas-anydaynow.com/archives/reconsideringscr.html> (detailing one Academy member's complaints and how precautions did not prevent copies from leaking).

11. See, e.g., Letter from Joel I. Klein, Assistant Attorney Gen., Dep't of Justice, to Garrard R. Beeney, Sullivan & Cromwell (Dec. 16, 1998), available at <http://www.justice.gov/atr/public/busreview/2121.htm>.

12. See *id.*

competitors might even thank you (quietly) for helping to craft a set of legally sanctioned barriers to entry in the technology field.

While the copyright-backed licensing arrangement keeps the commercial players in check, you also want to limit upstart entrepreneurs and individual would-be copiers. It is not enough to demand that licensees limit the functionality of their devices if users can disable those limitations with a few keypresses of a remote control.<sup>13</sup> As your users get more sophisticated, you may also worry about their facility with screwdrivers, circuit boards, and software compilers. You therefore demand that your licensees impose their limitations in a manner “robust” against user modification.<sup>14</sup> The goal is that none but those who have bound themselves to your licensing terms must be able to access your movie.

Moreover, once you have succumbed to the technical-protection imperative, you are unlikely to stop at just one component. If you black-box the software or hardware decoder, but do not secure the digital outputs from that black-box, someone will be able to take the decrypted stream from there.<sup>15</sup> If you leave high-quality analog outputs, someone can re-digitize content obtained through the “analog hole.”<sup>16</sup> If you shut down those outputs, however, you face the protests of early adopters and audiophiles who do not expect their functional, capable electronics to be selectively disabled.<sup>17</sup>

---

13. For example, many early region-locked DVD players could be made to play discs from any region with a few extra key-presses on the remote control. *See* Post-Hearing Comments of The Electronic Frontier Foundation, *In re Exemption to Prohibition on Circumvention of Copyright Protection Systems for Access Control Technologies*, Docket No. RM 2002-4 (June 5, 2003), at 6, available at <http://www.copyright.gov/1201/2003/post-hearing/post10.pdf>.

14. For a DRM implementation to make any sense in a scenario of limited trust, its barriers against user modification of the rights management must be at least as strong as those against user access to its protected content. *See* DVD-CCA CSS Procedural Specification § I.6.2.4–2.5, available at <http://cyber.law.harvard.edu/seminar/internet-client/readings/week2/02-08CSS.pdf>.

15. Hence High-Bandwidth Digital Content Protection (HDCP) for High-Definition Multimedia Interface (HDMI) and Digital Visual Interface (DVI) that provides for link-level encryption of digital video outputs, relying on digital handshakes to verify the trustworthiness of components on the other end of the video cable. *See* Digital Content Protection, HIGH-BANDWIDTH DIGITAL CONTENT PROTECTION SYSTEM 8, July 8, 2009, [http://www.digital-cp.com/hdcp\\_technologies](http://www.digital-cp.com/hdcp_technologies) (click “HDCP Specification Rev. 1.4”).

16. *See* Copy Protection Technical Working Group, Charter of the Analog Reconversion Discussion Group, <http://www.cptwg.org/Assets/TEXT%20FILES/ARDG/analogcharterfinal11403.doc> (last visited Feb. 18, 2010) (addressing “[c]opyright owner concerns over the present and future security of commercial audiovisual content that has been converted from digital to analog format and reconverted to digital format”).

17. *See* Mark Hachman, *TV Digital Rights Management Surfaces Again*, PCMAG.COM, Nov.

Before you know it, if you take technological copy-restriction seriously, you're requiring that your viewers upgrade every device in their home networks to satisfy the digital handshakes and cryptographic demands of secure communications, or you are downgrading that high-definition video to low enough resolution that users wonder what is worth the fuss.<sup>18</sup> You are demanding that every "digital media device" manufactured incorporate anti-copying technology.<sup>19</sup> Now if only we could use a neuralyzer to erase viewers' memories of the movie after it finished,<sup>20</sup> we could even prevent them from making around-the-water-cooler "derivative works" by sharing detailed synopses with friends.

We haven't gotten the neuralyzer into mass production yet, but attempts to implement or mandate the other technological measures are well beyond science fiction. A host of acronym-laden associations and lobbying groups have procured or attempted to legislate various parts of this scenario.<sup>21</sup> Those who count on technology to solve the problems they believe technology has exacerbated are drawn inexorably toward stricter and stricter regulations of technology. As those legal and architectural regulations widen, so too does a serious unintended consequence: the limitation of independent development and user innovation.



So what if DRM freezes user innovation? Most people will never modify their own media players. In a world where VCRs (and their DVD successors) still flash "12:00," why should we care about facilitating the more difficult user innovation? User innovation indirectly benefits even the non-technical end-user. When tinkerers have access to modify and develop technology, they tend to share their improvements, making them easier for non-tinkerers

---

4, 2009, <http://www.pcmag.com/article2/0,2817,2355382,00.asp> (discussing MPAA petition to the FCC regarding Selectable Output Control, and Public Knowledge opposition).

18. See Eric A. Taub, *Encryption Schemes Aimed at Film Piracy*, N.Y. TIMES, Aug. 30, 2001, at G6.

19. See Consumer Broadband and Digital Television Promotion Act (CBDTPA), S. 2048, 107th Cong. (2d Sess. 2002). This bill was introduced by Sen. Fritz Hollings. See *id.* Princeton Computer Science Professor Ed Felten generated a "Fritz's Hit List" of technologies that would have been regulated had the bill passed—anything that digitized audio or video, including baby monitors and Big-Mouth Billy Bass, the talking fish. See Fritz's Hit List, Freedom to Tinker, <http://www.freedom-to-tinker.com/tags/fritzs-hit-list> (last visited Mar. 16, 2010).

20. In the movie *Men in Black*, agents equipped with neuralyzers selectively erase the memories of witnesses who have seen too much. MEN IN BLACK (Columbia Pictures, 1997).

21. The footnotes have pointed to the real-world scenarios on which this low-skimming flight of fancy is based.

to obtain. Thus even if you do not find yourself drawn into the do-it-yourself (DIY) culture of modifying your own gear, you might be able to buy an off-the-shelf product that suits your needs better because it has been developed by or with insights from other user-innovators with tastes like yours; or you might hire someone to add the features you would like to a purchased product, finding you have more, cheaper options because no manufacturer claims a monopoly on upgrades and improvements.

Contrast the ecosystems around pre-recorded music and movies. Since the early 1980s, recorded music has been available in unencrypted digital form, on compact discs, while movies basically jumped from analog (and Macrovision-protected) VHS to DRM-encumbered digital with the 1997 introduction of the DVD and 1998 DMCA.<sup>22</sup> This difference (which record labels endlessly bemoan) has meant that the pool of pre-recorded music may be lawfully manipulated much more readily than pre-recorded video. CDs provide music directly to end-users in high-quality, unencrypted, digital form. End-users could choose DRM-free digital music long before most publishers or music stores offered DRM-free tracks online. A complete environment of music is freely usable on open playback devices.

Innovators have taken that freedom and run. When the CD player was introduced in 1982 (at a retail cost of \$900),<sup>23</sup> it could play a disk, skip to an identified track, seek, or repeat. In the years since, the digital music experience has been enhanced by participants of all shapes and sizes: large manufacturers, small startups, and end-users. Disc-based players have added shuffle modes, digital outputs, menu-driven controls, multi-disc changers, and portable versions. Perhaps more importantly, music has not been confined to discs. The Diamond Rio, introduced in 1998, brought digital music to pockets too small for the “Discman.”<sup>24</sup> A decade later, though the Rio is no more, it ushered in hundreds of portable music players.<sup>25</sup> Some of these players are built open;<sup>26</sup> others have been opened.<sup>27</sup> Some features that

---

22. Laserdisc never reached significant market share. See Laura Landro, *Get Set for Laser Videodisks, Round Two*, WALL ST. J., Dec. 6, 1988, at B1.

23. John Marcom, Jr., *Sales of Consumer Electronics to Grow Modestly in 1985*, *Industry Group Says*, WALL ST. J., Jan. 7, 1985, at 7.

24. Ashley Dunn, *The Cutting Edge Gift Guide*, L.A. TIMES, Nov. 30, 1998, at C6 (noting that “the ultimate Christmas gift this year is Diamond Multimedia’s \$199 Rio PMP300 portable MP3 player—a cigarette-pack-size device”); *Sony Corp. Introduces New Compact-Disk Player*, WALL ST. J., Mar. 17, 1988, at 7 (describing a four-inch player in which five-inch disks “extend past the player’s edges”).

25. Mike Musgrove, *Everything Seems to Play MP3s Lately*, WASH. POST, Dec. 7, 2001, at E1.

26. See, e.g., Teuthis Open Source Kits, Daisy MP3 Project Page, <http://www.teuthis.com/daisy/index.html> (last visited Mar. 20, 2010).

first appeared in user-written code, such as audio recording for the iPod, have since been commercialized, bringing user-designed features to the masses.

The Squeezebox, for example, started out as a little Ethernet-connected music gadget from a small start-up, Slim Devices. The early Slimp3 could take music from the computer to the stereo system, via a digital-to-analog converter, a bit of computing power, and a bright display. It has since spawned a full line of music devices, from wireless consumer-grade products to high-end audiophile ones.<sup>28</sup> Connected to a computer running Squeezebox Server software, the Squeezebox liberates music from the hard drive for listening anywhere in the home, adding a “now playing” display, a remote, web-based selection menus, and all the flexibility of a random-access computer-based music library—long, uninterrupted playlists, easy access to all your music in one place, and no disks to change or scratch. Because of music’s open format and users’ ability to move it around without copy-controls, the Squeezebox could be developed with no support or permission required from any in the established music industry. Its developers could take the user-availability of digital music as given, and build to interoperate at that interface.

Moreover, those who purchase the Squeezebox are not limited to what comes in the box; they can customize the open-source Squeezebox Server software. Many have, writing and sharing plugins to set musical alarms, program radio stations, show the weather, and integrate with other applications.<sup>29</sup> Even those who do not write code have access to the community’s products, since many users share their additions.<sup>30</sup> Even the Chumby, an open-by-design hardware platform that looks like a beanbag

---

27. See, e.g., Rockbox Software Project, <http://www.rockbox.org/twiki/bin/view/Main/WhyRockbox> (last visited Mar. 20, 2010).

28. See Logitech, Logitech Squeezebox, <http://www.logitechsqueezebox.com/>. Slim Devices is now a unit of Logitech.

29. See SqueezeCenter Plugins, [http://wiki.slimdevices.com/index.php/SqueezeCenter\\_Plugins](http://wiki.slimdevices.com/index.php/SqueezeCenter_Plugins) (last visited May 15, 2010). Users can create playlists with third-party recommendation engine MusicIP, organize the music’s metadata against Gracenote or MusicBrainz indexes, or send listening habits as status listings to a blog.

30. The Squeezebox Server is released under the GNU General Public License, version 2. See Softpedia, Download SqueezeBox Server, <http://mac.softpedia.com/get/Audio/SqueezeCenter.shtml> (last visited May 15, 2010). The GNU GPL under which it was released does not *require* redistribution of the source, but says that one who distributes compiled binaries must also distribute their accompanying source. See Free Software Foundation, GNU General Public License, Version 2 §3 (1991), *available at* <http://www.gnu.org/licenses/gpl-2.0.html>. Many user-developers find it attractive to share their work with the community, inviting others’ improvements.

with a screen (and speaker), can be programmed to play music from a Squeezebox Server-managed library.<sup>31</sup>

The Squeezebox is but one example. Because of the open nature of digital music and broadcast television, users and independent developers can create and choose their preferred experience. We can fill portable devices with shuffled sets of music tracks; we can outfit our homes with networked audio–video systems that share content around the house or follow us as we move; we can time-shift television on our schedules; we can synchronize collections among devices and across formats. DJs—professional or home—can mesh beats seamlessly from track to track.<sup>32</sup>

Now contrast music’s vibrant development environment and the range of music-capable devices to the limits around recorded movies. The DVD has been one of the most successful consumer electronics products of all-time, its numbers mounting rapidly after its 1997 launch,<sup>33</sup> but the movie-watching experience has barely changed since then. HD-DVD and Blu-Ray put more higher-resolution images on the disc, but still let consumers do little more than watch the movie and extras.<sup>34</sup> For the most part, new movie-watching technologies offer only the same basic features that DVD players have had since their introduction a decade ago. No DVD jukebox,<sup>35</sup> no easy direct navigation, no option to select scenes from a few movies to show in sequence or in comparison. Moreover, it is only in the last year that end-users have gotten a studio-authorized opportunity to copy a movie to a portable

---

31. See Chumby: Squeezebox Server, [http://www.chumby.com/pages/cp\\_squeeze](http://www.chumby.com/pages/cp_squeeze) (last visited May 15, 2010).

32. True, copyright law constrains users when they make copies, but so long as they have purchased the music, they claim fair use rights to manipulate its listening experience even when that entails transitory “copies.” See JESSICA LITMAN, DIGITAL COPYRIGHT 26–28 (2001) (describing the consequences of treating everything digital as copying). Copyright law should not constrain the mere act of listening; cf. *The Cartoon Network, LP v. CSC Holdings, Inc.*, 536 F.3d 121, 139 (2d Cir. 2008) (finding that Cablevision’s “remote storage” digital video recorder did not violate plaintiffs’ copyrights).

33. See Ross Johnson, *Getting a Piece of a DVD Windfall; Sales Are Soaring And Hollywood Is Split Over Dividing Profits*, N.Y. TIMES, Dec. 13, 2004, at C1.

34. It is not even clear many consumers notice the difference. A number still have low-definition screens, and commercial counterfeiters have taken advantage of people’s lack of visual acuity to sell fake Blu-Ray discs compressed to lower resolution. Geoffrey A. Fowler, *Pirates Prey on Blu-Ray DVD Format*, WALL ST. J., Nov. 17, 2008, at B1.

35. Kaleidescape introduced a \$8,000 system; although it won the first round of a contract fight with DVD-CCA, the appellate court reversed and remanded for consideration of the CSS General Specification that Kaleidescape contended was inapplicable, *DVD Copy Control Ass’n v. Kaleidescape, Inc.*, 176 Cal. App. 4th 697 (Cal. App. 6th Dist. 2009). DVD-CCA successfully enjoined RealNetworks from building a cheaper competitor. See *RealNetworks, Inc. v. DVD Copy Control Ass’n*, 641 F. Supp. 2d 913 (N.D. Cal. 2009).

player, send it to a mobile phone, or put it onto the home network so it moves from kitchen to living room to bedroom. Whereas both commercial and home-brew MP3 players, spurred by amateur development, have been able to do most of this for music for years, movies lag far behind. DVD's DMCA-backed encryption locks out independent developers and much experimentation. Users have had to wait years for the "business models" to catch up with features such as digital downloads or an authorized "digital copy."<sup>36</sup>



An impressive body of scholarship has formed around digital rights management (DRM).<sup>37</sup> Most legal academics criticize DRM for its effects on fair use: in a DRM-encumbered world, a media educator cannot cue movie clips for classroom commentary without special exemption; a literary critic is blocked from extracting e-book pages (or has the e-book deleted out from under her)<sup>38</sup>; and a mashup artist is restricted in sampling scope. These restrictions are direct consequences of DRM, problematic for copyright and culture.<sup>39</sup> Most scholars have thus characterized the "DRM problem" as that of accommodating fair use. Some argue that the loss of some marginal fair uses is an appropriate tradeoff for greater security of copyright protection.<sup>40</sup> Others argue that fair use may be approximated by user permissions, overrides, or appeals to a third party.<sup>41</sup> Still others contend that fair use of the

36. See, e.g., Disney File Digital Copy, <http://disney.go.com/disneyvideos/disneyfile/> (last visited Mar. 25, 2010). Rentals are still limited in variety, and portability often restricted to a small and inconsistent set of compatible devices.

37. See *infra* Part III.

38. See Geoffrey A. Fowler, *An Orwellian Moment for Amazon's Kindle*, WALL ST. J. DIGITS BLOG, Jul. 17, 2009, <http://blogs.wsj.com/digits/2009/07/17/an-orwellian-moment-for-amazons-kindle/> (describing Amazon's erasure of the e-book "Nineteen Eighty-Four" from users' Kindles because the supplier lacked the rights to sell it).

39. See generally LITMAN, *supra* note 32; Julie Cohen, *Lochner in Cyberspace: The New Economic Orthodoxy of 'Rights Management,'* 97 MICH. L. REV. 462 (1998); Pamela Samuelson, *Intellectual Property and the Digital Economy: Why the Anti-Circumvention Regulations Need to be Revised*, 14 BERKELEY TECH. L.J. 519 (1999).

40. E.g., Randal Picker, *Copyright and the DMCA: Market Locks and Technological Contracts*, in ANTITRUST, PATENTS AND COPYRIGHTS: EU AND US PERSPECTIVES 180 (Francois Leveque & Howard Shelanski eds., 2005); Jane Ginsburg, *From Having Copies to Experiencing Works: the Development of an Access Right in U.S. Copyright Law*, 50 J. COPYRIGHT SOC'Y U.S.A. 113 (2003).

41. E.g., Barbara L. Fox & Brian A. LaMacchia, *Encouraging Recognition of Fair Uses in DRM Systems*, 46 COMM. ASSOC. COMPUTING MACH. 61-63 (2003). *Contra* Dan L. Burk & Julie E. Cohen, *Fair Use Infrastructure for Rights Management Systems*, 15 HARV. J.L. & TECH. 41 (2001) (considering and ultimately rejecting this path); John S. Erickson & Deirdre K. Mulligan, *The Technical and Legal Dangers of Code-based Fair Use Enforcement*, 92 PROC. INST.

digital media is unnecessary, because fair use can be made from a work's other formats.<sup>42</sup> Yet others argue that because the heart of fair use is use without permission in unanticipated manner; technological controls and exceptions can never match the range of considerations a judge would be able to address if the matter came to litigation,<sup>43</sup> nor the spontaneity of "use first, ask permission later."

The fair use debate is important, but it is not the only problem with DRM. Equally important, but thus far largely overlooked, is the impact on user innovation and on the permitted development of media technology. Because DRM systems, by design and contract, must be hardened against user-modification, they foreclose a whole class of technology and an entire mode of development. Moreover, this problem is distinct from that of fair use. Even if we could wave a magic wand and fully accommodate fair use in DRM, the incompatibility with user innovation would persist, because it stems from a different and deeper aspect of the DRM system. Even the "fairest" DRM systems on the market today are unfair to the developers of new technology.

Anticircumvention law, backing technological protection measures (TPMs) and robustness rules, is fundamentally incompatible with deep-level user innovation. In a utilitarian copyright regime, where, as Thomas Macaulay put it, copyright is accepted as "a tax on readers for the purpose of giving a bounty to writers,"<sup>44</sup> the law must account for *all* the costs of foreclosing open modes of development. The "mode-of-development tax" is a significant unrecognized burden on the cultural, creative, and technology-based economy.<sup>45</sup>

---

ELECTRICAL & ELECTRONICS ENGINEERS 985, 995 (2004) ("In this sense the system would be broken from a copyright perspective: the system may protect the creator's copyright while upsetting the balance of copyright law by taking away users rights and the ability of new "rights" to emerge through the organic legal process.").

42. The court in *Universal City Studios, Inc. v. Corley*, 273 F.3d 429 (2d Cir. 2001), adopted the plaintiff's arguments to this effect.

[T]he DMCA does not impose even an arguable limitation on the opportunity to make a variety of traditional fair uses of DVD movies, such as commenting on their content, quoting excerpts from their screenplays, and even recording portions of the video images and sounds on film or tape by pointing a camera, a camcorder, or a microphone at a monitor as it displays the DVD movie.

*Id.* at 459.

43. See Timothy K. Armstrong, *Digital Rights Management and the Process of Fair Use*, 20 HARV. J.L. & TECH. 49, 57–59, 85–87 (2006).

44. Thomas Macaulay, Speech Delivered in the House of Commons (Feb. 5, 1841), in PROSE AND POETRY 731, 734–35 (G.M. Young ed., 1952).

45. For an example of this technological tax, one need only look to the development of

Part II of this Article examines the law and technology of digital rights management, particularly the interaction of statutory law, technological measures, and the contractual “robustness” conditions generally attached to them. Part III briefly reviews the history and existing academic debates around DRM to consider why they have overlooked the user-innovation impacts. Part IV develops examples of the DRM conflict with open development, contrasting more flexible “advisory” anti-copying features. Part V then introduces the rich economics and business literature on disruptive technology and user innovation, to argue that DRM’s copyright-driven constraints substantially harm cultural and technological development as well as user autonomy. Part VI concludes that the mode-of-development tax is too high a price to pay for imperfect copyright protection.

## II. THE MECHANICS OF CODE AND LAW

### A. BASIC TECHNOLOGY OF “DIGITAL RIGHTS MANAGEMENT”

The technology of digital rights management aims to “give” users digital works while managing their uses or copies: a DRM-protected track from the iTunes music store can be transferred to only five devices; a DVD can be played only on authorized players, coded to the region for which it was sold; a pay-per-view movie “expires” twenty-four hours after ordering. DRM’s fundamental challenge is to provide the desired uses but not more: give users enough control to enjoy the work and not enough to allow them (or systems under their control) to copy the works.<sup>46</sup> At the extreme, of course, one could develop a completely secure system by denying access to everyone, but that would find few buyers in the marketplace.<sup>47</sup>

---

music players versus DVD players, described *supra*.

46. See Jason F. Reid & William J. Caelli, *DRM, Trusted Computing and Operating System Architecture*, CONFS. RES. & PRAC. INFO. TECH., Jan. 2005, at 127, available at <http://crpit.com/confpapers/CRPITV44Reid.pdf>.

The essential premise of DRM is that a rights owner wishes to license digital content (which is represented as binary digits or bits) to a licensee or customer who agrees to be bound by the terms of the license. Note that the customer is not buying the bits themselves. Rather, they are buying the right to use the bits in a defined and restricted manner, as authorized in the terms of the license. Hence the license defines a type of usage policy.

*Id.*

47. Short of that extreme, media-producers dream of a price-segmented market, where each use can be priced according to its users’ willingness to pay. See Michael J. Meurer, *Price Discrimination, Personal Use and Piracy: Copyright Protection of Digital Works*, 45 *BUFF. L. REV.* 845, 877 (1997). See generally William W. Fisher III, *When Should We Permit Differential Pricing of Information?*, 55 *UCLA L. REV.* 1 (2007).

A digital media file is a series of bits—ones and zeros—written in a format that can be read by a hardware or software player and output as music, text, video, or a multi-media combination. Digital files are inherently copyable; there is usually no scarcity of bits or storage media to hold them. As cryptographer Bruce Schneier says, “trying to make digital files uncopyable is like trying to make water not wet.”<sup>48</sup> To manage the bits, therefore, providers try to control access—restricting listening to authenticated, paying subscribers or hindering copying—by enclosing the bits in a container of sorts, either physical or digital, that resists access by a would-be copyist.<sup>49</sup>

Since bits are readily copied, copy controls depend on the cooperation of their access and playback devices to function. Publishers try to embed their works in an ecosystem where copies are unplayable. Videocassettes, an analog recording medium, use Macrovision’s embedded noise as a containment strategy.<sup>50</sup> While this protection could be defeated if either the first VCR were told to suppress the Macrovision signal or the second to ignore it, it was effective when used with a pair of compliant devices.<sup>51</sup> Effective VHS copy-control, therefore, depended on both manipulating the format of the signal and constraining the design of playback and recording devices. Since what technology could set, technology could alter; technology is necessary but not sufficient to protect digital content. To control copying, anti-copying schemes control environments (and their inhabitants).

---

48. See Bruce Schneier, *Quickest Patch Ever*, WIRED NEWS, Sept. 7, 2006, <http://www.wired.com/politics/security/commentary/securitymatters/2006/09/71738> (describing how quickly Microsoft patched its Media Player application to disable the newly-released FairUse4WM software, which stripped the copy protection from Windows Media DRM 10 and 11 files).

49. More precisely, we can distinguish *access controls*, *copy controls*, and *watermarks*: access controls aim to stop unauthorized users from accessing a resource; copy controls to prevent its reproduction; and watermarks to track the usage or copying of a resource, without necessarily preventing any action.

50. See How Stuff Works, *How Does Copy Protection On a Video Tape Work?*, <http://electronics.howstuffworks.com/question313.htm> (last visited Mar. 20, 2010) A signal embedded in the vertical blanking interval of the video data is not displayed on television playback, but instead interferes with the automatic gain control component of other videocassette recorders, hindering VCR-to-VCR recording. *Id.*

51. Macrovision initially took advantage of accidental properties of the VCR technology. Once they were aware of its use as copy-control, however, VCR makers could design their devices not to be fooled by Macrovision’s spurious signals. Therefore, to make this copy-control more robust, Congress added a legal mandate in the DMCA. 17 U.S.C. § 1201(k)(1)(A)(i) (2006) (“[N]o person shall manufacture, import, offer to the public, provide or otherwise traffic in any . . . VHS format analog video cassette recorder unless such recorder conforms to the automatic gain control copy control technology. . .”).

Pure technology content control is a perpetual arms race. Protected environments persist for a time, then fall to stronger attacks through code analysis, hardware manipulation, or signal capture from the device.<sup>52</sup> Because there are many more people trying to break protection systems than to strengthen them, the attackers have the long-run advantage.<sup>53</sup> Thus,

---

52. If playback is in software, users might try to get the software to dump its unencrypted data, e.g., by copying it from memory, emulating a sound or video card, or emulating an entire environment. See SETH DAVID SCHOEN, TRUSTED COMPUTING, PROMISE AND RISK (2003), [http://www.eff.org/files/20031001\\_tc.pdf](http://www.eff.org/files/20031001_tc.pdf); see also ANDREW HUANG, HACKING THE XBOX 119–37 (2003) (hacking hardware); Messmer, *supra* note 7 (same).

53. Readers may ask how DRM differs from strong encryption, which can be implemented in open code and yet withstand breaks against the significant state and non-state actors who would like to break it. Encryption to protect content against eavesdropping by a third-party adversary is a hard but well-understood problem. We now have algorithms implementable (and implemented) on personal desktop computers that are believed to be impervious to attack by all the computing power in the world. Only a brute force attack, trying every possible key, could decrypt, and even with a mere 64 bit key, that leaves  $1.8 \times 10^{19}$  possibilities.

DRM's problem is different, though. As Cory Doctorow puts it, "In DRM, the attacker is \*also the recipient\*." Cory Doctorow, Address to the Microsoft Research Group, June 17, 2004, *available at* <http://craphound.com/msftdrm.txt>. The viewer of DRM-protected media is also the one against whose eavesdropping the system is trying to protect. The speedbump must block the user from doing unwanted things with the file, while permitting him or her to do the things for which he or she paid. While modern cryptography has solved many hard problems, it is helpless against the challenge of showing you something and simultaneously preventing you from seeing it.

Open source works beautifully for encryption because modern cryptosystems are built, following Kerckhoffs' Principle, on the maxim of least secrecy: disclose your algorithms, and secure your keys. Anyone can implement encryption compatible with Pretty Good Privacy (PGP), and decrypt a PGP-signed message in the open-source GNU Privacy Guard (GPG) so long as he has the private key to which it was encrypted. Users can independently verify (or have third parties verify for them) the security of their applications—and yet keep particular communications secured by the algorithms secret from anyone who does not know the private keys to that particular exchange. The threat model, as security researchers describe it, is the third party eavesdropper. Alice and Bob may communicate securely without Eve listening in. Even if Eve captures the communications stream, without the key, she sees only a stream of gibberish.

Asymmetric, or public key, encryption lets senders and recipients exchange encrypted messages without ever exchanging prior secrets. The recipient publishes a public key, half of a public-private key-pair, and guards the private key. The sender encrypts to the public key using public algorithms, and only the recipient in possession of the private key can decrypt the message. Even the eavesdropper, with all the other information about the message (algorithm and public key), can do nothing but try brute force attacks, which will fail if the parties have used a sufficiently long key-length.

This method works fine as an initial *access* control: only those who have the private key can read the messages sent to it, but it fails to assert any *use* controls after decryption, as DRM attempts. Yet when DRM systems use encryption for *use*-control, they are trying to secure communications against the same users to whom they're trying to sell media, all the

anticircumvention law tries to prevent this inevitability by bringing the heavy artillery of civil sanctions and criminal punishment to the battle between DRM-makers and DRM-breakers. It supports the DRM manipulation of the environment in which digital media can be played, constraining devices so that they can “contain” protected media. Copy protection can never regulate just the object itself—it must regulate the entire ecosystem to protect a work effectively. Thus, DRM technology entails a whole collection of subsidiary regulations to enforce it.<sup>54</sup>

#### B. THE MECHANICS OF ANTICIRCUMVENTION LAW.

Anticircumvention law extends the control of copyright, providing a legal hook from which to hang additional *contractual* restrictions. U.S. entertainment industries pushed the World Intellectual Property Organization (WIPO) to mandate legal protection for technical protection measures in the WIPO Copyright Treaty, Article 11 Obligations Concerning Technological Measures.<sup>55</sup> To comply, Congress added Chapter 12 to the Copyright Act through the DMCA.<sup>56</sup>

Section 1201, the core anticircumvention provision, provides that copyright holders who put technological locks on their works can use the law’s civil and criminal penalties<sup>57</sup> to block others from “circumventing”

---

while needing to give the user the use for which she has paid. It is as though the same person is both Bob, the intended recipient, and Eve the eavesdropper. DRM’s solution is to hand the keys to Bob for viewing without giving them to Eve, his alter ego. We could forbid Bob from doing bad things with the keys, but that is what copyright law already does in forbidding infringement. So now there are two things the system must hide while making them usable: its key and the plaintext.

54. See Susan Crawford, *The Biology of the Broadcast Flag*, 25 HASTINGS COMM. & ENT. L.J. 603, 629 (2003) (providing an extreme example of the environmental impact: once bitten by the DRM vampire, every other device connected to a broadcast-flagged DTV system would have been subject to regulatory control).

55. World Intellectual Property Organization Copyright Treaty, art. 11, Dec. 20, 1996, 112 Stat. 2860, 2186 U.N.T.S. 152.

Contracting Parties shall provide adequate legal protection and effective legal remedies against the circumvention of effective technological measures that are used by authors in connection with the exercise of their rights under this Treaty or the Berne Convention and that restrict acts, in respect of their works, which are not authorized by the authors concerned or permitted by law.

*Id.*; see also LITMAN, *supra* note 32, at 134–45.

56. Digital Millennium Copyright Act, Pub. L. No. 105-304, 112 Stat. 2860, 2863 (1998). For more discussion of the DMCA’s genesis, see *infra* Section III.A.

57. 17 U.S.C. §§ 1203–1204 (2006) set out the civil and criminal enforcement provisions.

those protections.<sup>58</sup> The law protects DRM technology with three prohibitions, forbidding anyone to “circumvent a technological measure that effectively controls access” to a copyright-protected work,<sup>59</sup> or to

manufacture, import, offer to the public, provide, or otherwise traffic in any technology, product, service, device, component, or part thereof, that

(A) is primarily designed or produced for the purpose of circumventing a technological measure that effectively controls access to a work protected under this title;

(B) has only limited commercially significant purpose or use other than to circumvent a technological measure that effectively controls access to a work protected under this title; or

(C) is marketed by that person or another acting in concert with that person with that person’s knowledge for use in circumventing a technological measure that effectively controls access to a work protected under this title.<sup>60</sup>

A parallel anti-trafficking provision prohibits tools for circumvention of *copy* controls, while the act of circumventing those controls—copying—is left to the prohibitions on ordinary infringement.<sup>61</sup>

The technological prerequisites for legal protection are minimal. “[A] technological measure ‘effectively controls access to a work’ if the measure, in the ordinary course of its operation, requires the application of information, or a process or a treatment, with the authority of the copyright owner, to gain access to the work.”<sup>62</sup>

Critical to the functioning of the anticircumvention hook, “authority of the copyright holder” can be granted conditionally.<sup>63</sup> While early critics argued that access control should be binary—that once “access” had been authorized, further uses were no longer within the realm of the DMCA but subject only to ordinary tests of infringement<sup>64</sup>—the courts have not agreed. Rather than determining that authority to access was granted

58. 17 U.S.C. § 1201(a)(3)(A) (“[T]o ‘circumvent a technological measure’ means to descramble a scrambled work, to decrypt an encrypted work, or otherwise to avoid, bypass, remove, deactivate, or impair a technological measure, without the authority of the copyright owner . . .”).

59. 17 U.S.C. § 1201(a)(1)(A).

60. 17 U.S.C. § 1201(a)(2).

61. 17 U.S.C. §§ 501, 1201(b)(1) (2006).

62. 17 U.S.C. § 1201(a)(3)(B).

63. See MARKS & TURNBULL, *supra* note 8, at 10.

64. See, e.g., Letter from Copyright’s Commons to David O. Carson, Esq., General Counsel (Mar. 31, 2000), available at [http://www.copyright.gov/1201/comments/reply/109selzer\\_bcis.pdf](http://www.copyright.gov/1201/comments/reply/109selzer_bcis.pdf) (giving reply comments in anticircumvention rulemaking).

straightforwardly through purchase of a DVD, the Second Circuit found it instead to be acquired only with a licensed DVD player, and only for licensor-approved uses.<sup>65</sup> With such an option, the copyright holder can then condition access on adherence to terms that are well beyond copyright, such as the region coding requirements in the DVD-CCA license, limitations on features or interconnections, and robustness rules.<sup>66</sup> Anticircumvention transforms weak technical measures into strong use controls, limiting technological possibility.

A copyright holder's adoption of a technological measure fortifies these works against access without the "authority" of the copyright holder; it also protects officially sanctioned playback devices against unauthorized competition or tinkering.<sup>67</sup> Even a weak scrambling scheme imports the full panoply of anticircumvention rights. Interoperation with a scrambled work against the "authority" of the copyright holder becomes a violation of the law, even if none of the aims of interoperation or intended uses of the product is an infringement of traditional copyright.

In short, before § 1201, someone who wanted to build a multimedia player for a newly acquired work would be legally free to do so,<sup>68</sup> perhaps improving the player options along the way. Under an anticircumvention regime, however, if any "technological measure" has been applied to the works, developers must seek permission to lawfully build a player or modify an existing one.<sup>69</sup> Section 1201(f), which permits some acts of circumvention for "reverse engineering,"<sup>70</sup> has not been useful as a shield for independent development of media technology.<sup>71</sup>

As described above, the DMCA protection on DVDs helps explain the lag in video playback options compared to their music-player counterparts. Other legal cases illustrate the comparative leeway pre-DMCA copyright gave

---

65. *Universal City Studios, Inc. v. Reimerdes*, 111 F. Supp. 2d 294, 317 n.137 (S.D.N.Y. 2000), *aff'd sub nom. Universal City Studios, Inc. v. Corley*, 73 F.3d 429, 443 n.13 (2d Cir. 2001).

66. See Prepared Testimony of Gwen Hinze, Staff Attorney, Elec. Frontier Found., May 15, 2003, [http://w2.eff.org/IP/DMCA/copyrightoffice/20030515\\_region\\_dvd.php](http://w2.eff.org/IP/DMCA/copyrightoffice/20030515_region_dvd.php).

67. See 17 U.S.C. §§ 1201(a)(1), 1201(b).

68. This assumes the only IP rights available are copyright restrictions. Patents, such as those claimed on MP3 encoding, may serve as a separate impediment.

69. See 17 U.S.C. § 1201.

70. See § 1201(f).

71. See *Universal v. Reimerdes*, 82 F. Supp. 2d 211, 218 (S.D.N.Y. 2000) ("[T]he legislative history makes it abundantly clear that Section 1201(f) permits reverse engineering of copyrighted computer programs only and does not authorize circumvention of technological systems that control access to other copyrighted works, such as movies."); see also *Davidson & Assocs. v. Jung*, 422 F.3d 630, 642 (8th Cir. 2005).

to reverse engineering and investigation. For example, when Sony tried to use copyright to monopolize its PlayStation platform, for which it manufactured the game consoles and licensed games, the Ninth Circuit rejected Sony's claims against the interoperable "Virtual Game Station."<sup>72</sup>

The district court found that "[t]o the extent that such a substitution [of Connectix's Virtual Game Station for Sony PlayStation console] occurs, Sony will lose console sales and profits." We recognize that this may be so. But because the Virtual Game Station is transformative, and does not merely supplant the PlayStation console, the Virtual Game Station is a legitimate competitor in the market for platforms on which Sony and Sony-licensed games can be played. For this reason, some economic loss by Sony as a result of this competition does not compel a finding of no fair use. Sony understandably seeks control over the market for devices that play games Sony produces or licenses. The copyright law, however, does not confer such a monopoly.<sup>73</sup>

The Ninth Circuit similarly rejected Sega's copyright attempt to lock the converse market, for games to run on its proprietary consoles: "[A]n attempt to monopolize the market by making it impossible for others to compete runs counter to the statutory purpose of promoting creative expression and cannot constitute a strong equitable basis for resisting the invocation of the fair use doctrine."<sup>74</sup>

Anticircumvention gives Sony and Sega a power copyright alone did not. As Dean Marks and Bruce Turnbull describe, anticircumvention laws supporting technical protection measures serve as the binding agent between technological controls and multi-party licensing agreements governing the

---

72. Sony Computer Entm't v. Connectix Corp., 203 F.3d 596, 606–07 (9th Cir. 2000). We find that Connectix's Virtual Game Station is modestly transformative. The product creates a new platform, the personal computer, on which consumers can play games designed for the Sony PlayStation. This innovation affords opportunities for game play in new environments, specifically anywhere a Sony PlayStation console and television are not available, but a computer with a CD-ROM drive is. More important, the Virtual Game Station itself is a wholly new product, notwithstanding the similarity of uses and functions between the Sony PlayStation and the Virtual Game Station. . . . Connectix reverse-engineered the Sony BIOS to produce a product that would be compatible with games designed for the Sony PlayStation. We have recognized this purpose as a legitimate one.

*Id.*

73. *Id.*

74. Sega Enters. Ltd. v. Accolade, Inc., 977 F.2d 1510, 1523–24 (9th Cir. 1992).

use and limitations of media subject to those controls.<sup>75</sup> Only those who promise to obey various non-copyright conditions may be granted authority.<sup>76</sup>

Though nothing in the text of the law speaks specifically to modes of development, that does not mean that the law has no hand in it. Under majority interpretation, anticircumvention prohibits user modification of players (hardware or software) used to play copyrighted works that have had any technical protection.<sup>77</sup> Tinkering with a device would seemingly void the “authorization” conveyed by the player license. The prohibition on open development more generally comes from a common feature of the license agreements through which DRM platforms are created: “robustness rules” and their implementation.

### C. LICENSING FOR ROBUSTNESS: HOW CONTENT PRODUCERS EXERT INFLUENCE IN THE HARDWARE MARKET

If one can access DRM-encumbered works only with the “authority of the copyright owner,”<sup>78</sup> then the licenses on which that authority is conditioned become the public law for those works. Those licenses enforce terms on the playback-device maker, and through them, upon the end-user for the copyrighted works.

While their particular usage terms may vary, DRM system licenses follow a common structural pattern. They require protection of the content with “usage rules” to be passed through to the end-user, and protection of the DRM system itself, with internal “compliance” and “robustness” rules.<sup>79</sup> If you are going to impose technical protection measures, it is because you distrust your users and want to stop them, through technology, from doing things that would otherwise be possible. As your understanding of users’ abilities and interests improves, you try to fill the cracks in your technical protections. In the logic of DRM designers, this condition makes sense: if

---

75. MARKS & TURNBULL, *supra* note 8, at 10–15.

76. *Id.*

77. *See, e.g.*, *Universal City Studios, Inc. v. Corley*, 273 F.3d 429 (2d Cir. 2001); *RealNetworks, Inc. v. DVD Copy Control Ass’n*, 641 F. Supp. 2d 913 (N.D. Cal. 2009); *321 Studios v. MGM Studios, Inc.*, 307 F. Supp. 2d 1085 (N.D. Cal. 2004).

78. 17 U.S.C. § 1201(a)(3)(A) (2006).

79. “Compliance” rules regulate the licensee devices’ adherence to the usage rules, while “robustness” requires efforts to withstand modifications. *See, e.g.*, *Advanced Access Content System Adopter Agreement F-1* (June 9, 2009), [http://www.aacsla.com/license/AACS\\_Adopter\\_Agrmt\\_090619.pdf](http://www.aacsla.com/license/AACS_Adopter_Agrmt_090619.pdf); *DVD-CCA CSS Procedural Specification ¶ 6.2.6*, <http://cyber.law.harvard.edu/seminar/internet-client/readings/week2/02-08CSS.pdf>; *Microsoft Corp., Compliance and Robustness Rules for Windows DRM*, <http://wmlicense.smdisp.net/wmdrmcompliance/> (last visited Mar. 25, 2010).

anticircumvention is to stop copying, the anti-copying system must be as resistant to hacking as the encryption itself. After all, a chain is only as strong as its weakest link, and a speedbump will not slow traffic if it is easy to avoid at full speed. And so § 1201 entails hardening of playback technology even if the law itself does not directly require it.

A review of license agreements for various content protection systems finds nearly identical robustness rules across the board.<sup>80</sup> The hardware or software implementations of media playback or transport must be designed to “effectively frustrate”<sup>81</sup> or “resist attempts to modify such products so as to defeat”<sup>82</sup> the content-protections: they must not include in the protected path any user-modifiable components such as switches, buttons, jumpers, or traces that may be cut; they must not be accessible to a debugger; and they must “keep secrets.” In short, to be authorized to access a work, an implementation must be hardened against tinkering and user exploration.<sup>83</sup>

The robustness rules are design rules.<sup>84</sup> They shape the architecture of the systems licensees are permitted to make available to end users. As Tarleton Gillespie describes it, these rules restructure the relationship of the user not only with the media, but also with the technology itself. They set the user up as a passive consumer, rather than an active participant in creating both

---

80. See sources cited *supra* note 79; see also Tarleton L. Gillespie, *Designed to 'Effectively Frustrate': Copyright, Technology, and the Agency of Users*, 8 NEW MEDIA & SOC'Y 651, 651–69 (2006).

81. Advanced Access Content System Interim Content Participant Agreement, Exhibit C, Part 2, § 3.2, available at [http://www.aacsla.com/license/AACS\\_Interim\\_Content\\_Participant\\_Agrmt\\_090605.pdf](http://www.aacsla.com/license/AACS_Interim_Content_Participant_Agrmt_090605.pdf) (“Licensed Products shall be manufactured in a manner clearly designed to effectively frustrate attempts to modify such Licensed Products or the performance of such Licensed Products to defeat the Content Protection Requirements.”). Even Sun’s DrEAM, the purportedly open DRM specification, mandates robustness on its clients: “Client security: the implementation of a robust client that will be required for a viable solution. The implementation of the robust client will depend on the hardware and software support available.” DReaM-CAS Client Specification Version 1.0 Rev A, Technical Specification, § 1.1 (2007) (on file with author).

82. Microsoft, Microsoft Windows Media 10 SDK Robustness Rules, <http://wmlicense.smdisp.net/wmdrmcompliance/> (click on “Robustness Rules for WMDRM10 Devices”) (last visited Mar. 20, 2010) (“Licensed Products as shipped . . . must be designed and manufactured so as to resist attempts to modify such products so as to defeat the functions of the Microsoft Implementation.”).

83. See TARLETON GILLESPIE, WIRED SHUT: COPYRIGHT AND THE SHAPE OF DIGITAL CULTURE 225–29 (2007).

84. See CARLISS Y. BALDWIN & KIM B. CLARK, DESIGN RULES, VOL. 1: THE POWER OF MODULARITY 80 (2000) (describing design rules as a set of constraints on manufacture).

culture and technology.<sup>85</sup> As Lawrence Lessig puts it, the rules move us from a “Read/Write” to a “Read/Only” culture.<sup>86</sup>

Section 1201’s authority requirement imports the terms imposed by contractual licenses. When all the licenses for major DRM systems require robustness as a condition, robustness becomes the equivalent to a direct legal requirement. Private law is transmuted into public. As scholars such as Lessig have pointed out, however, this private law can be both equally constraining and more opaque for its indirection.<sup>87</sup>

Indeed, the robustness mechanism could have been written into public law. In the Broadcast Flag Rules, regulations adopted by the FCC for the protection of digital television transmissions, robustness was mandated in the Rule.<sup>88</sup> So written, it could be challenged in court. The American Library Association and other public interest groups did just that, and successfully argued that the Broadcast Flag Rule exceeded the FCC’s authority.<sup>89</sup> Challenges to DMCA-backed DRM should get the same hearing.

---

85. GILLESPIE, *supra* note 83, at 226–27.

86. LAWRENCE LESSIG, REMIX 28 (2008).

87. *See* LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE 95–98, 223–25 (2000); *see also* GILLESPIE, *supra* note 83, at 219 (describing “technically imposed copyright protections that depend on encrypted content, technologies that abide by a set of rules applied to that content, and laws making it illegal to tamper with or produce alternatives to those technologies”).

88. *See* Robustness Requirements for Covered Demodulator Products, 47 C.F.R. § 73.9007 (2005) (“The content protection requirements set forth in the demodulator compliance requirements shall be implemented in a reasonable method so that they cannot be defeated or circumvented merely by an ordinary user using generally-available tools or equipment.”) Even “generally available” is defined in the regulation:

Generally-available tools or equipment means tools or equipment that are widely available at a reasonable price, including but not limited to, screwdrivers, jumpers, clips and soldering irons. Generally-available tools or equipment also means specialized electronic tools or software tools that are widely available at a reasonable price, other than devices or technologies that are designed and made available for the specific purpose of bypassing or circumventing the protection technologies used to meet the requirements set forth in this subpart. Such specialized electronic tools or software tools includes, but is not limited to, EEPROM readers and writers, debuggers or decompilers.

*Id.* at note.

89. *See* Am. Library Ass’n v. Fed. Comm’n Comm’n, 401 F.3d 489 (D.C. Cir. 2005).

D. OPEN SOURCE SOFTWARE IS INCOMPATIBLE WITH ROBUSTNESS  
REQUIREMENTS

In sharp contrast to the constrained environment of the DMCA, free and open source software invites end-user modification.<sup>90</sup> By disclosing its details and granting users permission to modify (in copyright terms, to create derivative works), free and open source software both opens the car's hood and provides a schematic diagram (more or less well labeled) of the mechanical components. Even relative novices can learn their way around by tweaking a few lines and recompiling to see the effect—much as one learning web design can “view source” on a webpage and learn by imitation and adaptation. Experts can refine the software to their needs, both fixing bugs and adding features. User communities formed around free and open source software have developed complex applications, operating systems, and environments.<sup>91</sup> Free and open source software powers much of the Internet's infrastructure. More than half of Internet web servers run the open-source Apache web server,<sup>92</sup> often on the free GNU/Linux operating system, and many of their visitors now use open-source browsers such as Mozilla Firefox or Google Chrome. Even someone who never reads a line of source code benefits from openness: from the competition of independent programmers available to service the software, from the pressure it puts on proprietary vendors, and from the ease of developing complementary applications.

Free and open source software development depends critically on openness. While the terms “free software” and “open source” reflect different emphases and motivations of their participants,<sup>93</sup> at heart, both refer to software whose source code (the human-readable version of the computer

---

90. “Free software” tends to be the label of choice for those who, following the Free Software Foundation, give an explicitly political dimension to the sharing of source code and the freedom to modify software; “open source” is often used pragmatically to emphasize the economic and efficiency benefits of disclosed source. Whether by philosophic or economic temperament, their mode of development has the effect of making software features much more accessible to user innovation.

91. Karim R. Lakhani & Eric von Hippel, *How Open Source Software Works: “Free” User-to-user Assistance*, 32 RES. POL'Y 923, 924 (2003).

92. See Netcraft, Web Server Survey Archives, [http://news.netcraft.com/archives/web\\_server\\_survey.html](http://news.netcraft.com/archives/web_server_survey.html) (last visited Mar. 20, 2010).

93. See Wendy Seltzer, *Why Open Source Needs Copyright Politics*, in OPEN SOURCES 2.0: THE CONTINUING EVOLUTION 149, 149 (DiBona et al. eds. 2005); Yochai Benkler, *The Battle over the Institutional Ecosystem in the Digital Environment*, 44 COMM. ASS'N COMPUTING MACHINERY 84, 84 (2001).

instructions) is made available to the programs' users for use and modification.<sup>94</sup>

The Free Software Foundation expresses the core principles of Free Software as “four essential freedoms”:

- The freedom to run the program, for any purpose (freedom 0).
- The freedom to study how the program works, and change it to make it do what you wish (freedom 1). Access to the source code is a precondition for this.
- The freedom to redistribute copies so you can help your neighbor (freedom 2).
- The freedom to improve the program, and release your improvements (and modified versions in general) to the public, so that the whole community benefits (freedom 3). Access to the source code is a precondition for this.<sup>95</sup>

All four of these components are necessary to give users full autonomy in their software environment; to use and learn from the program and to modify it to suit their needs. They guard against lock-in to an uncooperative vendor or defunct system, and assure that users will be able to reuse their individual investments in the program. Moreover, the Free Software Foundation asserts, “Freedom 1 [the freedom to modify] must be practical, not just theoretical; i.e., no tivoization.”<sup>96</sup> Version 3 of the GPL, issued in 2007,<sup>97</sup> requires licensees to provide with a program “installation information” sufficient to allow the use of modified code in the same manner as the originally installed program.<sup>98</sup> “Look but don’t touch” is not freedom.

---

94. See Free Software Foundation, The Free Software Definition, <http://www.gnu.org/philosophy/free-sw.html> (last visited Mar. 5, 2010); Open Source Initiative, The Open Source Definition (Annotated) <http://www.opensource.org/docs/definition.php> (last visited Mar. 5, 2010).

95. Free Software Foundation, *supra* note 94.

96. *Id.* (referring to the TiVo digital video recorder, which runs a GNU/Linux operating system and makes source available, but does not permit the user to install modifications on the TiVo box).

97. Developers of code can choose which license to apply, subject to any requirements they inherit from upstream code they wish to use under license. The Linux kernel remains under GPLv2, while new versions of the FSF’s GNU utilities are released under GPLv3. Those who wish to distribute the newest GNU utilities, therefore, are required to make both source and installation information available.

98. Free Software Foundation, GNU General Public License, Version 3 § 6 (2007),

The GNU General Public License (GPL) maintains these freedoms by a “copyleft” provision: anyone is free to reuse GPL-licensed code, so long as those who do release their derivative works on the same terms, under the GPL.<sup>99</sup>

The Open Source Initiative (OSI) is more explicitly oriented toward the “economic and strategic advantage” to be gained from openness.<sup>100</sup> It takes openness as a foundation for diversity and productive innovation. “We require access to un-obfuscated source code because you can’t evolve programs without modifying them. Since our purpose is to make evolution easy, we require that modification be made easy.”<sup>101</sup> The OSI aims to leverage the community of open source developers: “In order to get the maximum benefit from the process, the maximum diversity of persons and groups should be equally eligible to contribute to open sources. Therefore we

---

*available at* <http://www.gnu.org/licenses/gpl.txt>. The GPL states:

If you convey an object code work under this section in, or with, or specifically for use in, a User Product, and the conveying occurs as part of a transaction in which the right of possession and use of the User Product is transferred to the recipient in perpetuity or for a fixed term (regardless of how the transaction is characterized), the Corresponding Source conveyed under this section must be accompanied by the Installation Information.

“Installation Information” for a User Product means any methods, procedures, authorization keys, or other information required to install and execute modified versions of a covered work in that User Product from a modified version of its Corresponding Source. The information must suffice to ensure that the continued functioning of the modified object code is in no case prevented or interfered with solely because modification has been made.

*Id.*

99.

[5] You may convey a work based on the Program, or the modifications to produce it from the Program, in the form of source code . . . provided that you also meet all of these conditions:

- a) The work must carry prominent notices stating that you modified it, and giving a relevant date.
- b) The work must carry prominent notices stating that it is released under this License . . . .
- c) You must license the entire work, as a whole, under this License to anyone who comes into possession of a copy.

[6] You may convey a covered work in object code form under the terms of sections 4 and 5, provided that you also convey the machine-readable Corresponding Source under the terms of this License. . . .

*Id.* §§ 5, 6.

100. Open Source Initiative, About the Open Source Initiative, <http://www.opensource.org/about> (last visited Mar. 5, 2010).

101. Open Source Initiative, *supra* note 94.

forbid any open-source license from locking anybody out of the process.”<sup>102</sup> OSI identifies a number of licenses as “Open Source” compliant.<sup>103</sup> Like the GPL, which is among them, all Open Source-licensed distributions include the source code and ability to modify the software.<sup>104</sup>

It would be impossible to build a player for DRM-encumbered media that complied with either the Free Software or the Open Source definition. DRM is incompatible with both the letter and the spirit of open source and free software licenses.<sup>105</sup> Anticircumvention forbids users from exploring the possibilities of their media, and it forecloses developers from offering media players that can be user-modified.

The Internet multiplies the opportunities for open development, connecting the forces of plentiful open source software, a critical mass of networked potential contributors, and cheaper communications among them on a neutral platform. As Federal Communication Commission Chairman Julius Genachowski asked in the September 2009 speech launching the Commission’s “open Internet” discussion: “Why has the Internet proved to be such a powerful engine for creativity, innovation, and economic growth? A big part of the answer traces back to one key decision by the Internet’s original architects: to make the Internet an open system.”<sup>106</sup>

---

102. *Id.*

103. See Open Source Initiative, Open Source Licenses by Category, <http://www.opensource.org/licenses/category> (last visited Mar. 5 2010).

104. The Open Source Definition does not require a copyleft-style provision mandating openness on downstream redistributors, although it is compatible with the GPL’s copyleft requirement. See sources cited, *supra* note 94.

105. See *infra* Section V. In the 2007 revision of the GPL, the Free Software Foundation added a clause explicitly forbidding the use of GPL in technological protection measures. See GNU General Public License, *supra* note 98, § 3 (“No covered work shall be deemed part of an effective technological measure under any applicable law fulfilling obligations under article 11 [*sic*] of the WIPO copyright treaty adopted on 20 December 1996, or similar laws prohibiting or restricting circumvention of such measures.”). This specific prohibition is distinct from the general incompatibility of DRM with the free and open source mode of development.

106. Julius Genachowski, Chairman, FCC, Remarks at The Brookings Institution: Preserving a Free and Open Internet: A Platform for Innovation, Opportunity, and Prosperity (Sept. 21, 2009), available at <http://openinternet.gov/read-speech.html>. Genachowski stated:

The Internet’s creators didn’t want the network architecture—or any single entity—to pick winners and losers. Because it might pick the wrong ones. Instead, the Internet’s open architecture pushes decision-making and intelligence to the edge of the network—to end users, to the cloud, to businesses of every size and in every sector of the economy, to creators and speakers across the country and around the globe. In the words of

### III. THE ACADEMIC DEBATE

#### A. ANTICIRCUMVENTION'S ADVOCATES

Attempts to “copy protect” media have been around for a long time, but the rise of digital storage accelerated their move from technology to law. Digital media, copyright holders argued, allowed users to make perfect copies, while high-speed communications networks would allow them to share those copies easily.<sup>107</sup> Intellectual property industries proclaimed a “digital dilemma”: there would be no cars to transit the digital information superhighway unless copyright law guaranteed protection for their copyrighted contents.<sup>108</sup> They sought this protection in both technology and law.

Publishing industries laid out a syllogism: “content” was key to the growth of the nascent Internet—then known as the National Information Infrastructure (NII); content production would halt if its protection could not be assured; therefore, content protection must be made a core part of the Internet.<sup>109</sup> The NII taskforce did not overestimate technology: “it is clear

---

Tim Berners-Lee, the Internet is a “blank canvas”—allowing anyone to contribute and to innovate without permission.

*Id.*

107. See, e.g., *NII Copyright Protection Act of 1995: Hearing on H.R. 2441 Before the Subcomm. on Courts and Intellectual Property of the H. Comm. on the Judiciary*, 104th Cong. (1996) (statement of Barbara A. Munder, Information Industry Association); see also *NII Copyright Protection Act of 1995: Joint Hearing on H.R. 2441 and S. 1284 Before the Subcomm. on Courts and Intellectual Property of the House Judiciary Comm. and the Senate Judiciary Comm.*, 104th Cong. (1995).

108. INFORMATION INFRASTRUCTURE TASKFORCE, INTELLECTUAL PROPERTY AND THE NATIONAL INFORMATION INFRASTRUCTURE: THE REPORT OF THE WORKING GROUP ON INTELLECTUAL PROPERTY RIGHTS 10–17 (1995), available at <http://www.uspto.gov/web/offices/com/doc/ipnii/ipnii.txt> [hereinafter NII REPORT].

109. *Id.* at 10–11.

[T]he full potential of the NII will not be realized if the education, information and entertainment products protected by intellectual property laws are not protected effectively when disseminated via the NII. Creators and other owners of intellectual property rights will not be willing to put their interests at risk if appropriate systems—both in the U.S. and internationally—are not in place to permit them to set and enforce the terms and conditions under which their works are made available in the NII environment. Likewise, the public will not use the services available on the NII and generate the market necessary for its success unless a wide variety of works are available under equitable and reasonable terms and conditions, and the integrity of those works is assured. All the computers, telephones, fax machines, scanners, cameras, keyboards, televisions, monitors, printers, switches, routers, wires, cables, networks and satellites in the world will not create a successful NII, if there is no content. What will drive the NII is the content moving through it.

that technology can be used to defeat any protection that technology may provide,”<sup>110</sup> but it drew from that the conclusion that law must be added to the mix, in support of technology-based restriction. The NII White Paper proposed the hybrid: legal prohibition on circumvention of technical measures.

The Working Group finds that legal protection alone will not be adequate to provide incentive to authors to create and to disseminate works to the public. Similarly, technological protection likely will not be effective unless the law also provides some protection for the technological processes and systems used to prevent or restrict unauthorized uses of copyrighted works.

The Working Group finds that prohibition of devices, products, components and services that defeat technological methods of preventing unauthorized use is in the public interest and furthers the Constitutional purpose of copyright laws. Consumers of copyrighted works pay for the acts of infringers; copyright owners have suggested that the price of legitimate copies of copyrighted works may be higher due to infringement losses suffered by copyright owners. The public will also have access to more copyrighted works via the NII if they are not vulnerable to the defeat of protection systems.

Therefore, the Working Group recommends that the Copyright Act be amended to include a new Chapter 12, which would include a provision to prohibit the importation, manufacture or distribution of any device, product or component incorporated into a device or product, or the provision of any service, the primary purpose or effect of which is to avoid, bypass, remove, deactivate, or otherwise circumvent, without authority of the copyright owner or the law, any process, treatment, mechanism or system which prevents or inhibits the violation of any of the exclusive rights under Section 106. The provision will not eliminate the risk that protection systems will be defeated, but it will reduce it.<sup>111</sup>

From the White Paper, through the “policy laundering”<sup>112</sup> of treaty-making at WIPO, Chapter 12 was added to the Copyright Act in 1998.<sup>113</sup>

---

*Id.* That the NII taskforce could not have envisioned or explained Wikipedia, whose authors and editors contribute knowing that their works are shared freely—even so that others may profit—or the Creative Commons licenses many use to share their works is perhaps the first hint that theirs is not the only path to the Progress of Science.

110. *Id.* at 136.

111. *Id.* at 139–40.

112. Policy laundering takes an unpalatable policy argument from the domestic realm and “launders” it through an international treaty organization, WIPO, before bringing it back to the national legislature as “treaty obligation.” *See* Ian Hossein, Paper Presented at the

Section 1201 of the DMCA prohibits “circumvention” of technological access-control measures, and prohibits trafficking in circumvention tools for access or copy controls.<sup>114</sup> It gives legal force to technological barriers, however weak or strong they may be, and forbids distribution of the tools of circumvention even when their intended aim is not copyright infringement.<sup>115</sup>

The anticircumvention provision has been controversial from its inception. Early critics questioned its constitutionality<sup>116</sup> and blamed it for expanding the rights of copyright holders at the expense of the public<sup>117</sup>; while proponents argued that it was necessary to keep the traditional rights of copyright holders viable in new markets.<sup>118</sup> By and large, these arguments have taken place *within copyright*, sharing copyright’s focus on the production, use, and (perhaps) marketing of creative expression, which explains why they tend to miss anticircumvention’s *outside-copyright* effects on technology development.

Technologists and academics supporting anticircumvention law said it would sustain copyright and an ecology of new business models around copyrighted works.<sup>119</sup> Mark Stefik first described a “trusted system” to envelop copyrighted works and control their transfer:

The term *trusted system* refers to computers that can be relied on to do certain things. For example, suppose that a creator or publisher forbids all copying of a particular digital work. A trusted system in this context would reliably and infallibly carry out that stipulation; no amount of shouting or coaxing would coerce it to copy the work.<sup>120</sup>

Similarly, Jane Ginsburg acknowledges that anticircumvention provides new rights, but argues that the technological shift from possession of hard copies to “experiencing works” required more extensive grants.<sup>121</sup> To

International Studies Association, Montreal, Quebec, Canada: International Relations Theories and the Regulation of International Dataflows: Policy Laundering and other International Policy Dynamics 136 (Mar. 17, 2004).

113. WIPO Copyright to Performances and Phonograms Treaties Implementation Act of 1998, 17 Pub. L. No. 105-304, 112 Stat. 2860, 2863.

114. Digital Millennium Copyright Act, 17 U.S.C. § 1201 (2006).

115. *Id.*; see also *supra* Section II.B.

116. See, e.g., Neil W. Netanel, *Locating Copyright Within the Fair Use in the First Amendment Skein*, 54 STAN. L. REV. 1, 78–80 (2001); see also *infra* Section III.B.

117. See *infra* Section III.B.

118. *Id.*

119. See Mark Stefik, *Trusted Systems*, SCI. AM., Mar. 1997, at 78–81.

120. Mark Stefik, *Letting Loose the Light: Igniting Commerce in Electronic Publication*, in INTERNET DREAMS: MYTHS, AND METAPHORS 249, 257 (Mark Stefik ed., 1996).

121. Jane Ginsburg, *From Having Copies to Experiencing Works: the Development of an Access*

Ginsburg, anticircumvention law flows from the spirit of the Copyright Clause of the Constitution<sup>122</sup> as a natural response to changes in technology and the markets it supports, suggesting that “in the digital environment, the ‘exclusive Right’ that the Constitution authorizes Congress to secure to authors is not only a ‘copy’-right, but an access right.”<sup>123</sup> As copyright holders lost control in dissemination, they should get back a different lever of control.

## B. ANTICIRCUMVENTION’S CRITICS

### 1. *Anticircumvention Stops End-User Fair Use*

Much of the previous anticircumvention scholarship has focused on the direct constraints DRM places on media usage and the constrictions that places on fair use.<sup>124</sup> When media is available only through DRM-respecting applications, users are forced to accept the usage limitations even if those limitations are more restrictive than those of copyright. A user buying a song tethered to the computer on which she downloaded it can never resell that purchase.<sup>125</sup> A film critic seeking to display a clip from DVD cannot easily take an excerpt for this purpose.<sup>126</sup>

For some, this tradeoff is the cost of access to digital copies. With finer-grained permissions comes a more tailored set of costs—price discrimination that gives users access to lower cost and more abundant copies.<sup>127</sup> For others, the tradeoff is too great, damaging the important public benefits of fair use: privacy in reading,<sup>128</sup> freedom from asking permission in advance, and freedom to criticize.<sup>129</sup> The positive externalities of fairly-used media, for

---

*Right in U.S. Copyright Law*, 50 J. COPYRIGHT SOC. 113, 115 (2003).

122. U.S. CONST. art. 1, § 8, cl. 8.

123. Ginsburg, *supra* note 121, at 115.

124. *See, e.g.*, sources cited *supra* note 39.

125. *See* LITMAN, *supra* note 32, at 83 (“Augmenting copyright law with legally enforceable access control could completely annul the first sale doctrine.”); *see also* 17 U.S.C. § 109 (2006) (first sale doctrine).

126. *See* the petitions for exemptions in the Copyright Office’s triennial rulemaking under 17 U.S.C. § 1201(a)(1)(C), U.S. Copyright Office, Rulemaking on Exemptions from Prohibition on Circumvention of Technological Measures that Control Access to Copyrighted Works, available at <http://www.copyright.gov/1201/>.

127. *See* Tom Bell, *Fair Use Vs. Fared Use: The Impact of Automated Rights Management on Copyright’s Fair Use Doctrine*, 76 N.C. L. REV. 557, 588(1998); Randal C. Picker, *From Edison to the Broadcast Flag: Mechanisms of Consent and Refusal and the Propertization of Copyright*, 70 U. CHI. L. REV. 281, 296 (2003).

128. Julie Cohen, *A Right to Read Anonymously: A Closer Look at “Copyright Management” in Cyberspace*, 28 CONN. L. REV. 981, 1007–10 (1995).

129. *See* Yochai Benkler, *Free as the Air to Common Use: First Amendment Constraints on Enclosure of the Public Domain*, 74 N.Y.U. L. REV. 354, 410 (1999); Netanel, *supra* note 116, at

example, mean the law should be willing to subsidize it rather than restrict it. Cutting off fair use diminishes the commons on which future creativity depends.<sup>130</sup>

Academic critics attacked anticircumvention first from within copyright: technological protections interfere with the fair use limitations built into copyright law.<sup>131</sup> While some unauthorized use of copyrighted works is legally non-infringing, DRM has no way to recognize the difference between a fair reproduction for classroom use, commentary, or parody, and an infringing reproduction. The statute prohibits DRM's circumvention even when the protection blocks access for non-infringing uses.<sup>132</sup>

Pamela Samuelson, Jessica Litman, Julie Cohen, and Yochai Benkler have all highlighted the inconsistency between technological enforcement of absolutes and the nuanced case-by-case and use-by-use exceptions fair use grants to copyright's "exclusive" rights.<sup>133</sup> Some, including the author of this Article, have argued that this incompatibility renders the DMCA unconstitutional, because its making absolute of copyright-style rights, "paracopyright," in David Nimmer's terms, violates the First Amendment.<sup>134</sup>

A number of these arguments have been raised in constitutional challenges to the DMCA,<sup>135</sup> but as these challenges appeared to defend conduct that could also enable mass reproduction, assertions of fair use harms fell on deaf ears. For example, the Second Circuit told the *Universal v.*

26.

130. LAWRENCE LESSIG, FREE CULTURE: THE NATURE AND FEATURE OF CREATIVITY 97–99 (2005); PATRICIA AUFTERHEIDE & PETER JASZI, UNTOLD STORIES: CREATIVE CONSEQUENCES OF THE RIGHTS CLEARANCE CULTURE FOR DOCUMENTARY FILMMAKERS (2004), available at [http://www.centerforsocialmedia.org/files/pdf/UNTOLDSTORIES\\_Report.pdf](http://www.centerforsocialmedia.org/files/pdf/UNTOLDSTORIES_Report.pdf).

131. Copyright does not prohibit *all* reproductions, only those that interfere with the copyright holder's rights. Reproductions "for purposes such as criticism, comment, news reporting, [and] teaching" are permitted as fair use. 17 U.S.C. § 107 (2006).

132. See Samuelson, *supra* note 39, at 524 ("[T]here are far more legitimate reasons to circumvent a technical protection system than the DMCA's act-of-circumvention provision expressly recognizes.").

133. See generally Benkler, *supra* note 129; Julie E. Cohen, *Copyright and the Jurisprudence of Self-Help*, 13 BERKELEY TECH. L.J. 1089 (1998); Cohen, *supra* note 128; LITMAN, *supra* note 32; Samuelson, *supra* note 39.

134. Brief for Electronic Frontier Foundation et al. as Amici Curiae Supporting Plaintiff's Opposition to Defendant's Motion for Partial Summary Judgment at 321 Studios v. Metro Goldwyn Mayer Studios Inc., 307 F.Supp.2d 1085 (N.D. Cal. 2004) (No. C 02-1955 SI); DAVID NIMMER, 3 NIMMER ON COPYRIGHT § 12A.15[C] (1999 supp.); Netanel, *supra* note 116, at 78.

135. E.g., *Universal City Studios, Inc. v. Corley*, 273 F.3d 429 (2d Cir. 2001); *321 Studios*, 307 F. Supp. 2d 1085.

*Corley* defendants that would-be fair users of DVD video could point camcorders at their television screens.<sup>136</sup>

A second round of fair use-inspired critics has asked whether technology can help solve the problems technology has created: can fair use be accommodated within “hybrid” DRM systems? While we cannot put a judge on a chip, some have proposed that we can approach the pre-DMCA regime (and public good) more effectively by pairing more generous defaults with systems built to accommodate a call-out to an external authority or trusted third party. Julie Cohen and Dan Burk note the law’s shortcomings in the cultural realm, but wonder whether that is a mere function of implementation.<sup>137</sup> Could a “fair use infrastructure” for rights management systems, a trusted third party who could adjudicate requests for access to make fair use, capture the nuance and spontaneity of fair use abilities? Burk and Cohen endeavor to design a “second-best solution designed to make the best of a bad situation,” but ultimately reject even their modified DRM.<sup>138</sup> Tim Armstrong proposes a system that sets the defaults toward use, while keeping an audit trail of asserted fair uses.<sup>139</sup> Deirdre Mulligan and John Erickson describe the possible use of rights expression languages, rather than automated restrictions.<sup>140</sup>

The fair use strain of scholarship often comes down to a cost–benefit analysis within copyright: does DRM increase expression and artistic creation, by providing greater security in the chance to profit from that work, and does that benefit outweigh the cost in the expressive opportunities of audiences and follow-on creators? In the decade since the DMCA’s enactment, evidence on the harm side of the balance sheet has mounted.<sup>141</sup> Moreover, while proposals for technical recognition of fair use go some distance toward mitigating one of the problems with DRM and anticircumvention, they create a new set of problems. By mandating that developers of technology harden their devices or software, they force the deployment of user-resistant technology and methods capable of being hardened before distribution to end-users.

---

136. *Corley*, 273 F.3d at 459.

137. Burk & Cohen, *supra* note 41, at 50–51.

138. *Id.* at 80.

139. Armstrong, *supra* note 43, at 99–108.

140. Erickson & Mulligan, *supra* note 41, at 994.

141. See generally Deirdre K. Mulligan, John Han & Aaron J. Burstein, *How DRM-based Content Delivery Systems Disrupt Expectations of “Personal Use,”* in DIGITAL RIGHTS MANAGEMENT WORKSHOP 77 (Moti Yung ed., 2003); FRED VON LOHMANN, UNINTENDED CONSEQUENCES: TEN YEARS UNDER THE DMCA (2010), <http://www.eff.org/wp/unintended-consequences-ten-years-under-dmca/>.

## 2. *DRM Does Not Stop Copying*

Along with the fair use criticisms,<sup>142</sup> analysts have added another concern in the copyright sphere: DRM does not in fact stop copying. Even if the technical restrictions are defended as “speedbumps,” to “keep honest people honest,”<sup>143</sup> even honest drivers can take simple detours to avoid the bump. As Peter Biddle and colleagues explained in the “Darknet” hypothesis, it takes only one user to break copy protection and make content available on peer-to-peer networks; all subsequent users need merely to find that copy.<sup>144</sup> Ironically, finding illegal copies on the internet remains simpler than programming a VCR, despite entertainment industry attempts to crack down on the practice. DMCA notwithstanding, popular new movies and music tracks are available DRM-free from file sharing networks or download sites almost as soon as they are released in DRM-encumbered form.<sup>145</sup> Reviewing this evidence of *increased* copying, Fred von Lohmann concludes that “the DMCA has thus far failed to deliver under its policy rationale.”<sup>146</sup> Von Lohmann suggests that on its own terms, as a measure to stop mass digital distribution of copyrighted works, anticircumvention-backed DRM has failed, instead causing end-users to seek out the unauthorized copies that are more functional than the DRM-saddled licensed versions.<sup>147</sup>

The upshot is that copyright holders are getting little of the copyright benefit they claimed—DRM is not reducing infringing reproduction—while adding hurdles to lawful but perhaps unwanted or unanticipated uses of their works.<sup>148</sup> A few industry players have managed to turn the momentum toward more open offerings, but most continue to use the weaknesses of

---

142. *See supra* Section III.B.1.

143. *See Piracy Prevention and the Broadcast Flag: Hearing Before the Subcommittee on Courts, the Internet, and Intellectual Property of the H. Comm. on the Judiciary*, 108th Cong. 45–49(2003) (statement of Fritz Attaway, Exec. Vice President, Gov’t Relations & Wash. Gen. Counsel, Motion Picture Ass’n Am.); BBC News, Digital Film: Industry answers, <http://news.bbc.co.uk/2/hi/entertainment/4691232.stm#7/> (last visited Mar. 20, 2010).

144. Peter Biddle et al., *The Darknet and the Future of Content Distribution*, in DIGITAL RIGHTS MANAGEMENT WORKSHOP 155, 156 (Joan Feigenbaum ed., 2002).

145. *See* Douglas Wolk, *Days of the Leak*, SPIN MAGAZINE, Aug. 2007, at 86–88. (describing how albums are often leaked to the public on file sharing services before official release).

146. Fred von Lohmann, *Measuring the Digital Millennium Copyright Act Against the Darknet: Implications for the Regulation of Technological Protection Measures*, 24 LOY. L.A. ENT. L. REV. 635, 640 (2004).

147. *Id.* at 642–43.

148. This Article does not defend copyright infringement. Rather, it concludes that the cost of technological measures against infringement is too high, and that there is no way of drawing a less costly technical barrier.

existing DRM systems as an excuse to ramp up the “protections” even further.<sup>149</sup>

3. *Anticircumvention Hinders Technology Innovation*

If DRM does not stop copying, then, what does it do? Both critics and proponents have recognized it as a method of technological control that continues to function, through the mechanism of anticircumvention, despite its weakness against piracy.

Fair use and use without permission are not the only casualties of DRM. A second branch of scholarship has focused on anticircumvention’s impact on scientific inquiry and innovation in product design. Pamela Samuelson decried the impact on science, since the opacity of the DMCA’s research exemptions, the difficulty of obtaining permission to research, and the need for clarification of the right to do so without permission chill investigation into computer security.<sup>150</sup> Ed Felten became an active critic of anticircumvention after a computer science research paper garnered DMCA threats.<sup>151</sup> Fred von Lohmann chronicled the “unintended consequences” of anticircumvention law, particularly in dampening the opportunities to innovate in the complementary markets *around* copyrighted works.<sup>152</sup>

These criticisms broaden the inquiry beyond fair use. Even if DRM furthered copyright’s purpose, promoting creative authorship, extra-copyright effects add to the cost side of the equation. These cross-domain comparisons add the challenge of even greater incommensurability, forcing policymakers (if they want to make a fully informed decision) to compare the value of a new playback technology to that of a new creative work.

Anticircumvention changes the market structure around copyrightable expression, giving the creator of a *work of authorship*—or, more often, a group of copyright holders—the right and ability to control the market for *playback technologies*. It lets copyright holders leverage their statutory monopoly on

---

149. In some realms, DRM use is now diminishing, most notably through Apple’s recent change of track to offer DRM-free sales from the iTunes Music Store. This shift came from Apple, not the music labels, once Apple had achieved sufficient dominance in the music market to maintain its technological hold without the DRM lock. *See* Steve Jobs, Thoughts on Music, <http://www.apple.com/hotnews/thoughtsonmusic/> (last visited Feb. 22, 2010).

150. Pamela Samuelson, *Anticircumvention Rules: Threat to Science*, 293 SCIENCE 2028, 2028–29 (2001).

151. *See* Edward W. Felten, *DRM and Public Policy*, 48 COMM. ASS’N COMPUTER MACHINERY 112 (2005); *see infra* Section IV.B.

152. *See* VON LOHMANN, *supra* note 141, at 1; Fred von Lohmann, *Fair Use as Innovation Policy*, 23 BERKLEY TECH. L.J. 829, 851–53 (2008).

expression into technology.<sup>153</sup> Tim Wu finds that DRM's market structure holds back innovation. Giving copyright holders too much control over dissemination of their works and communications denies opportunities to a more widely distributed pool of potential innovators.<sup>154</sup> On a probabilistic analysis, having fewer potential innovation opportunities lowers the chances of successful innovation.<sup>155</sup>

Not all those who look at copyright and innovation oppose this extension of control. Randall Picker argues that technological tying can add market opportunities. By giving creators a broader scope in which to exploit their monopolies, he suggests tying can facilitate price and product differentiation.<sup>156</sup> Picker argues that between the increased incentives for the creators of copyrighted works and the decreased opportunities for "distributional" entry, on balance, DRM does more good than harm.<sup>157</sup>

More scholars, however, question the DMCA's impact on market structures. Anticircumvention's control presumes either that the successful creator is in the best position to design or recognize playback technologies or that these technology markets matter less than the creator's incentive.<sup>158</sup> This technological copyright control exacts a high price, however, given the multi-purpose nature of many playback technologies, and the "long tail" and communicative aspects of media,<sup>159</sup> many of whose most important applications relate to personal uses and freedoms, not mass market content.

---

153. See WILLIAM PATRY, *MORAL PANICS AND THE COPYRIGHT WARS* 165 (2009) ("One of the new rights in the DMCA granted the motion picture studios the power to dictate the functional design of consumer electronic devices.").

154. See Timothy Wu, *Copyright's Communications Policy*, 103 MICH. L. REV. 278, 331–32 (2004); Tim Wu, *Intellectual Property, Innovation, and Decentralized Decisions*, 92 VA. L. REV. 123, 141–46 (2006).

155. See ARNOLD KLING & NICK SCHULZ, *FROM POVERTY TO PROSPERITY* 8 (2009) ("Often, innovation is the result of the unplanned trial-and-error learning that takes place among new enterprises, rather than the organized research and development efforts of large organizations.").

156. Picker, *supra* note 40, at 181.

157. *Id.*

158. Cf. DAN L. BURK & MARK A. LEMLEY, *THE PATENT CRISIS AND HOW THE COURTS CAN SOLVE IT* 73–74 (2009) (discussing cumulative innovation in patent improvements).

159. See generally CHRIS ANDERSON, *THE LONG TAIL* (2006); Andrew M. Odlyzko, *Content Is Not King*, FIRST MONDAY (Feb. 5, 2001), <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/833/742/>.

#### IV. ANTICIRCUMVENTION'S APPLICATION

Law-backed technical protection measures throw a wrench into the mix of ever more user-accessible technologies and greater “generative” power in the hands of individuals and small-scale entrepreneurs.<sup>160</sup> By confining tinkering, experimentation, and exploration to those authorized by copyright holders, the law chokes off much of the potential of user-created technological improvements.<sup>161</sup> The examples of music and movie copy controls demonstrate the impact of anticircumvention on innovation and research, while that of Adobe PDF suggests that similar anti-infringement goals could be served less intrusively with advisory rather than mandatory features.

##### A. CD VERSUS DVD: THE EFFECT OF A LOCKED-DOWN MEDIA FORMAT

This Article contrasted music's vibrant development environment and the range of music-capable devices against the limits around recorded movies, *supra* Part I. The DMCA explains this comparative poverty: in contrast to the CD, which, to preserve compatibility with legacy equipment has remained a vector for unencrypted, unprotected content,<sup>162</sup> the DVD was born encrypted. Since that encryption triggers the anticircumvention protections of the DMCA, no one may decrypt “without authorization” or build players to do so.<sup>163</sup> In order to play back a commercially recorded DVD movie, the player requires multiple keys: player, disc, and title.<sup>164</sup> Player keys

---

160. *See supra* Section II.D.

161. *See infra* Part V (discussing the benefits of user-created technological improvements).

162. Vendors have tried to deploy “copy-protected” CDs, by pushing the CD format's specification in attempts to foil copying but not playback as Macrovision did for VHS, e.g. by introducing bogus tracks that a player will pass over but a copier will hang trying to correct. *See Low-Tech Pen Foils CD Copy-Protection Device*, L.A. TIMES, May 21, 2002, at C1. The multitude of unprotected CDs keeps player manufacturers on the right side of the anti-trafficking provisions of 17 U.S.C. § 1201(a)(2)(B) (2006), however, if they develop better anti-skip mechanisms for copying.

163. *Universal City Studios, Inc. v. Corley*, 273 F.3d 429, 444 (2d Cir. 2001); *see also infra* Section IV.A.

164. *Universal City Studios, Inc. v. Reimerdes*, 111 F. Supp. 2d 294, 317–18 (S.D.N.Y. 2000), *aff'd sub nom. Universal City Studios, Inc. v. Corley*, 273 F.3d 429 (2001).

One cannot gain access to a CSS-protected work on a DVD without application of the three keys that are required by the software. One cannot lawfully gain access to the keys except by entering into a license with the DVD CCA under authority granted by the copyright owners or by purchasing a DVD player or drive containing the keys pursuant to such a license.

(which give access to the other keys on the disc) are distributed only in players licensed by the DVD consortium (DVD CCA). Thus, the terms on which the consortium is willing to license players—including required limitations on outputs, restrictions on copying, and geographically-limited playback enforced by region coding—set a ceiling on *all* players' capabilities, while the requirement of authorization prevents independent development without permission.<sup>165</sup> Hence the public's options for movie playback and manipulation are significantly poorer than those for music.<sup>166</sup>

Under the threat that movie studios would withhold their content from insufficiently secure devices, consumer electronics companies and software makers negotiated protections through a cross-industry consortium with movie studios. They had competing interests in places: the studios wanting strong protection for their movies, the consumer electronics and software wanting something achievable in hardware and software, respectively, at commercially feasible cost under the manufacturing constraints of the early '90s. Their discussions produced the Content Scramble System (CSS)—a combination of disc and player keys and a scrambling system that relied upon both to decrypt the contents of the movie on disc. "What CSS does," explains Tarleton Gillespie, "is prevent consumers from watching the DVD using the wrong device—that is, one that hasn't been certified by the movie studios."<sup>167</sup>

As important as the technological scrambling were the conditions of authorization in the licenses required to de-scramble. The scrambling, no matter how weak, serves as the hook for a collection of usage rules and limitations: stay within the bounds of the license agreement and you could be authorized to descramble; or proceed without a license or exceed its authorization and it will be deemed a circumvention.<sup>168</sup> Only those licensed

---

*Id.*

165. *See infra* Part V.

166. The difference between options available for pre-recorded music and movies is not merely one of consumer preference. While we may enjoy video differently from music, the market reflects significant innovation around modes of video delivery other than pre-recorded movies, both the creation and hosting of short video, and the manipulation and time- and space-shifting of television broadcasts, which are sent unencrypted—from the early disrupter Betamax, through TiVo and Slingbox. Ultimately, non-movie video may grow large enough to fuel its own ecosystem, putting pressure on that of pre-recorded movies, but not yet.

167. *See* GILLESPIE, *supra* note 83, at 171 (2007) ("Control of copying and redistribution is imposed by ensuring that authorized DVD players themselves do not allow copying; CSS ensures that consumers will only use these authorized machines.")

168. MARKS & TURNBULL, *supra* note 8. One might ask whether patent already does this work, making anticircumvention superfluous. Namely, DVD playback also implicates

by the consortium can decrypt; and once hooked, only to do a limited set of activities: playback, on a region-matched playback device, without permitting copying or the skipping of promotional content. And do so “robustly,” with prohibitions on end-user reverse engineering.<sup>169</sup>

These conditions were ineffective to prevent CSS from being broken. In late 1999, programmers analyzed the CSS scheme and produced DeCSS, a computer program capable of decrypting the scrambled contents of a DVD.<sup>170</sup> Jon Johansen, a fifteen-year-old Norwegian on the team who posted the code to his website, stated that he did so to enable people to play DVDs on Linux, as no available player at the time ran on that platform.<sup>171</sup> *2600 Magazine* took the DeCSS code and posted it to their website.<sup>172</sup> They were forced to remove and then link to the code after receiving a takedown notice. Despite arguments about the communicative power of code, the courts found this posting to be a provision of circumvention tools, in violation of §§ 1201(a)(2) and (b)(1).<sup>173</sup>

The court never distinguished carefully between access and copying, finding that DeCSS circumvented both types of controls. Moreover the court only cursorily analyzed the question of authorization. Since descrambling or decrypting amounts to circumvention only when conducted “without the authority of the copyright owner,”<sup>174</sup> we need an account of “authority” to distinguish between the legitimacy of the DVD player that decrypts CSS and the circumvention of DeCSS through substantially the same mathematical

---

numerous patents, licensed by the DVD-CCA pool. *See* Letter, *supra* note 11. Patent seems to exert less of a chill on independent research and innovation. *See* Mark A. Lemley, *Ignoring Patents*, 2008 MICH. ST. L. REV. 19 (2008). MP3 is claimed by Fraunhofer yet widely implemented without license. Because it is not legally bound to the usage rules, the patent hook sinks less deep.

169. *See Reimerdes*, 111 F. Supp. 2d at 310.

170. *Id.* at 311.

171. While the *Reimerdes* court finds that “[a]t the time of trial [early 2000], licenses had been issued to numerous hardware and software manufacturers, including two companies that plan to release DVD players for computers running the Linux operating system,” *id.* at 310, no commercially available DVD player was available for Linux at that time. *See id.* at 337 n.243. As late as 2004, companies were still heralding the “First Linux DVD Player,” years after DVD playback was possible in Windows or Mac (the Xing player implicated in the CSS break was released in 1998). *Xing’s Premier Software-Only DVD Player Provides Most Complete, Highest-Quality Solution for Multimedia PCs*, BUS. WIRE, Sep. 8, 1998, available at [http://findarticles.com/p/articles/mi\\_m0EIN/is\\_1998\\_Sept\\_8/ai\\_50290471/](http://findarticles.com/p/articles/mi_m0EIN/is_1998_Sept_8/ai_50290471/). Of course the offering of any Linux player already depends on a different threat model, since in an open source environment, the player can even less effectively secure against capture of the decrypted data between the player and the content’s human viewer.

172. *Reimerdes*, 111 F. Supp. 2d at 309.

173. *Universal City Studios, Inc. v. Corley*, 273 F.3d 429, 455 (2d Cir. 2001).

174. 17 U.S.C. §§ 1201(a)(3)(A), (b)(2)(A) (2006).

operations. The customer, after all, is not party to license agreements among the consortium, nor asked to sign a license on purchase of a DVD. It is unclear therefore how the DVD or its player conveys to the viewer her authorization to watch the DVD.

The CD/DVD fork in media paths stems from the interaction of law and technology, in which law supports privately created systems of copyright protection through technological restriction. With access to open music formats, developers can invent first, negotiate market position later (if at all), and users can become developers. By contrast, faced with encrypted movies, only those developers who can pre-negotiate access can offer improvements, and only in locked-down ways approved and licensed by the movie producers. Thus the law-backed encryption on DVDs (and their high-definition successors) locks out interoperation and modification from those who hope to be public about their work. While those who care little about the legal consequences have already broken the encryption and built work-arounds, anticircumvention law largely deters mainstream developers from building without permission—and thereby from building much of the innovative use that video might support. The markets for both copyrighted works and their complementary tools and players are stunted by these conditions.

#### B. SDMI AND THE FREEDOM TO TINKER

As music companies saw the other side to the DVD's cautionary tale, the Secure Digital Music Initiative (SDMI) group proposed “to develop the voluntary, open framework for playing, storing, and distributing digital music necessary to enable a new market to emerge.”<sup>175</sup> Responding to “analog hole” and redigitization concerns, SDMI proposed an ecosystem of devices that would recognize an embedded watermark and refuse to play watermarked content if it appeared outside of a licensed context. Within this ecosystem, watermarks would prevent copying or uses without permission.<sup>176</sup>

---

175. SDMI Fact Sheet, [http://web.archive.org/web/20040213143259/www.sdmi.org/who\\_we\\_are.htm](http://web.archive.org/web/20040213143259/www.sdmi.org/who_we_are.htm) (Jan. 27, 2004) (accessed by searching for [sdmi.org](http://sdmi.org) in the Internet Archive index).

176. Watermarks are one response to the “analog hole” redigitization problem. By marking content as protected-origin, they can indicate that the content was originally restricted. If the initial restrictions never allow the content to be exchanged in unencrypted form, then compliant devices might be programmed to refuse to play unencrypted watermarked content. Of course non-compliant devices might simply ignore the watermark. Other proposals from SDMI suggested that they intended to use these watermarks to indicate that an audio track had not been subjected to compression since application of the watermark. See CRAVER ET AL., *READING BETWEEN THE LINES: LESSONS FROM THE SDMI CHALLENGE* (2001), available at <http://www.usenix.org/events/sec01/craver.pdf>. SDMI

Audio watermarks must serve two potentially conflicting goals: they must not interfere perceptibly with the sound of the music in which they are embedded, yet they must be detectable mechanically.<sup>177</sup> Thus someone determined to thwart the watermark would aim to remove it without changing the track's listening quality, so that both devices and listeners were satisfied with it.

On September 6, 2000, SDMI identified four watermarking technologies as possible elements of its music-protection strategy and issued an "Open Letter to the Digital Community," inviting people to "attack" the proposed watermarks.<sup>178</sup> A team of computer scientists and electrical engineers took up the challenge, downloading the provided samples and analyzing them using signal processing methods to determine how the watermarks had been applied and how they might be removed imperceptibly. This attack analysis

---

invited participation by "companies that have significant direct activity in digital music or digital music technology. These companies must express their commitment to SDMI by agreeing to abide by its Terms of Participation and paying a \$20,000 annual membership fee." SDMI.org, Frequently Asked Questions, <http://web.archive.org/web/20020924131640/www.sdmi.org/FAQ.htm> (originally available at <http://www.sdmi.org/FAQ.htm>).

177. See CRAVER ET AL., *supra* note 176, at 3.

[W]atermarking technologies [are those] in which subtle modifications are made to an audio file to encode information without perceptible change in how the file sounds. Watermarks can be either *robust* or *fragile*: robust watermarks are designed to survive common transformations like digital-to-audio conversion, compression and decompression, and the addition of small amounts of noise to the file; whereas fragile watermarks do not survive such transformations, and are used to indicate modification of the file.

*Id.*

178. Leonardo Chiariglione, An Open Letter to the Digital Community, [http://web.archive.org/web/20040216013811/http://www.sdmi.org/pr/OL\\_Sept\\_6\\_2000.htm](http://web.archive.org/web/20040216013811/http://www.sdmi.org/pr/OL_Sept_6_2000.htm) (Sept. 6, 2000) (accessed by searching for sdmi.org in the Internet Archive index).

Here's an invitation to show off your skills, make some money, and help shape the future of the online digital music economy.

The Secure Digital Music Initiative is a multi-industry initiative working to develop a secure framework for the digital distribution of music. SDMI protected content will be embedded with an inaudible, robust watermark or use other technology that is designed to prevent the unauthorized copying, sharing, and use of digital music.

We are now in the process of testing the technologies that will allow these protections. The proposed technologies must pass several stringent tests: they must be inaudible, robust, and run efficiently on various platforms, including PCs. They should also be tested by *you*.

So here's the invitation: Attack the proposed technologies. Crack them.

By successfully breaking the SDMI protected content, you will play a role in determining what technology SDMI will adopt.

*Id.*

was not merely invited by SDMI, it is a standard part of computer security research. The peer review of technologies is critical to assessing their strength and improving their security.<sup>179</sup> Before copyright holders, device manufacturers, and public purchasers bought into the SDMI scheme, they might want to know how effective it would be at meeting its stated aims.

Edward Felten's team successfully broke all four watermark technologies, creating new samples from watermarked tests that were audibly indistinguishable from unwatermarked originals and bore no detectable trace of the watermark.<sup>180</sup> The team chose to present their work as an academic paper, securing its acceptance to the peer-reviewed Fourth International Information Hiding Workshop.<sup>181</sup> Before they could present the paper, however, Felten received a letter from the Recording Industry Association of America (RIAA), threatening suit under the DMCA.

[A]ny disclosure of information gained from participating in the Public Challenge would be outside the scope of activities permitted by the Agreement and could subject you and your research team to actions under the Digital Millennium Copyright Act ("DMCA").

Unfortunately, the disclosure that you are contemplating could result in significantly broader consequences and could directly lead to the illegal distribution of copyrighted material. Such disclosure is not authorized in the Agreement, would constitute a violation of the Agreement and would subject your research team to enforcement actions under the DMCA and possibly other federal laws. . . .

In addition, because public disclosure of your research would be outside the limited authorization of the Agreement, you could be subject to enforcement actions under federal law, including the DMCA. The Agreement specifically reserves any rights that proponents of the technology being attacked may have "under any applicable law, including, without limitation, the U.S. Digital Millennium Copyright Act, for any acts not expressly authorized by their Agreement." The Agreement simply does not "expressly authorize" participants to disclose information and research

---

179. See Declaration of Ed Lazowska, *Felten v. Recording Indus. Ass'n Am.*, No. CV-01-2669 (D.N.J. Aug. 13, 2001), available at [http://w2.eff.org/IP/DMCA/Felten\\_v\\_RIAA/20010813\\_cra\\_decl.html](http://w2.eff.org/IP/DMCA/Felten_v_RIAA/20010813_cra_decl.html). See generally SCHNEIER, *supra* note 5 (discussing computer security).

180. CRAVER ET AL., *supra* note 176, at 12.

181. See Janelle Brown, *Is the RIAA running scared?*, SALON, Apr. 26, 2001, <http://www.salon1999.com/technology/log/2001/04/26/felten/index.html>; Reading Between the Lines: Lessons from the SDMI Challenge, <http://www.cs.princeton.edu/sip/sdmi/> (last visited Mar. 30, 2010).

developed through participating in the Public Challenge and thus such disclosure could be the subject of a DMCA action.<sup>182</sup>

The researchers and conference organizers were concerned enough by the legal threats that Felten and his team withdrew the paper from the April 2001 conference. The RIAA promptly issued a press release claiming that the organization had never intended to sue. Nonetheless, the researchers felt severe enough chill as they prepared to present their work in future papers and conference presentations that they filed suit seeking a declaratory judgment that their work did not violate the DMCA.<sup>183</sup>

The incident demonstrated the chills of the DMCA's broad prohibition on dissemination of technology and "components." While the researchers did ultimately publish and present their work, the RIAA and SDMI were able to use DMCA claims to delay it by half a year, and might well have scared off entirely researchers not backed by a university professor and pro bono legal assistance. Anticircumvention law, thus, blocks the scientific and educational examination of technology, including interoperation anticircumvention.

And to what end? The research that demonstrated flaws in these watermarks points to gaps in the ultimate strategy. If even the best watermarks SDMI could design were vulnerable to analysis and removal, it is unlikely that these disclosure-oriented researchers were the only ones who could do so. Among the music sharers targeted by the technological restrictions would be others able to skirt their controls, and who having done so, could share the resulting cleared files with others.

Since the breaking of its watermark technologies, the SDMI initiative has faded into insignificance as a technological force. A note on its now-defunct website indicated that the "SDMI Forum is on hiatus as of June 2001, and is not accepting new members."<sup>184</sup> Meanwhile, Felten's research has been cited by others in both computer security and copyright protection research.<sup>185</sup>

---

182. Letter from Matthew Oppenheim, Secretary, SDMI Foundation, to Edward Felten, April 9, 2001, *available at* <http://cryptome.org/sdmi-attack.htm> (last visited Mar. 25, 2010).

183. *See* First Amended Complaint, Felten v. Recording Indus. Ass'n Am., No. CV-01-2669 (D.N.J. June 26, 2001), *available at* [http://w2.eff.org/IP/DMCA/Felten\\_v\\_RIAA/20010626\\_eff\\_felten\\_amended\\_complaint.html](http://w2.eff.org/IP/DMCA/Felten_v_RIAA/20010626_eff_felten_amended_complaint.html). The lawsuit was dismissed for lack of standing. Felten v. Recording Indus. Ass'n Am., No. 01-CV-2669 (D.N.J. Nov. 30, 2001) (docket report on file).

184. SDMI.org, Frequently Asked Questions, <http://web.archive.org/web/20040213073219/www.sdmi.org/FAQ.htm> (Jan. 27, 2004) (accessed by searching for [sdmi.org](http://sdmi.org) in the Internet Archive index).

185. *E.g.*, Alin C. Popescu & Hany Farid, *Statistical Tools for Digital Forensics*, in INFORMATION HIDING 128, 128 (Jessica J. Freidrich, ed., 2004) (security research); J. ALEX HALDERMANN & EDWARD W. FELTEN, LESSONS FROM THE SONY CD DRM EPISODE

SDMI-like expansion is not the only alternative. Technological responses to law need not be pushed to the illogical extreme, but can leave room for independent development if they forego the legal backing for their tamper-proofing. Adobe's deployment of a limited technological control in its PDF (portable document format) authoring software illustrates that option and pinpoints the spots where legally-enforced technology could block open development, as well as the advantages of calling off the DMCA's hounds.

### C. AN ALTERNATIVE: NON-ROBUST ADVISORY MEASURES

Document authors using high-end versions of Acrobat, Adobe's PDF-authoring application, can choose to control initial *access*, encrypting and password-protecting documents; and to control *use*, by restricting printing, text selection, and even usage of screen reader applications.<sup>186</sup> The differing implementations of the two branches—and their interactions with open source and with anticircumvention law—give a glimpse of the way the law affects independent development.

PDF's access control is provided by encryption using modern, public algorithms. All can be given an encrypted blob, which only those who have been given the decryption key can decrypt. But 'access' is binary: on or off. Once the reader has decrypted the document, he or she has the decrypted document in its full digital glory, with the potential to print, save, and resend.

This encryption is robust: even implemented in open source, fully modifiable software, it gives access to the document only to those who enter the correct password or certificate. The strength of the encryption is independent of the implementation's publicity—indeed, public algorithms and implementations that have been subject to testing are likely to be stronger than privately developed alternatives.<sup>187</sup> With a key space sufficiently large to stymie brute-force attacks, an author can be relatively confident that her documents will be accessible only to those with the password. She can create a separate pass-key for each user and document, or use public-key infrastructure to encrypt to a recipient's existing private key.

As she wants to share a document more widely, however, the author may worry whether one authorized recipient will share it with another, make extra

---

(2006), <http://itpolicy.princeton.edu/pub/sonydrm-ext.pdf> (copyright protections).

186. See Adobe, Adobe Acrobat Pro Extended: Features, <http://www.adobe.com/products/acrobatproextended/features/> (last visited Mar. 20, 2010).

187. See Bruce Schneier, *The Ethics of Vulnerability Research*, INFO, SECURITY MAG., May 2008, <http://www.schneier.com/essay-211.html> ("Anyone can design a security system that he cannot break.").

copies, or leave printouts lying around. By itself, encryption does not address those concerns.

In Adobe's system, some of these *use* controls are provided by flags marking restrictions on what the Adobe software will do with document. Adobe's reader software enforces the restrictions, so the recipient who uses Acrobat to open a document flagged "Printing: Not Allowed, Selection: Not Allowed" will find the usual print and text selection options greyed-out and unavailable. In Adobe's reader, such a document can be viewed on-screen but not printed, excerpted, or converted to other formats. The file itself can be copied infinitely many times, but each copy will have these same flags set.<sup>188</sup>

The PDF specification is publicly disclosed<sup>189</sup> and has been implemented in other applications including Apple's Preview and the GPL-licensed xpdf.<sup>190</sup> Preview, the Apple Macintosh's default PDF reader, responds to a "do not print" flag by disallowing printing and prompting for a document-authoring password. It likewise disallows selection of text in a document whose author has set that flag.

As distributed, xpdf behaves similarly, complying with the flags as well. Attempts to print flagged documents from an unmodified copy of xpdf are met with the error message: "Printing this document is not allowed." But the xpdf implementation is not 'robust.' Since xpdf's source is available to those who want to modify it, users frustrated by the flagged recommendations of a PDF document can compile their own versions. Among other customizations, they can tell modified-xpdf to ignore the flags of the program—by removing a check in five simple lines of code.<sup>191</sup>

188. The document may be encrypted, but the user who gets view-only privileges does not need to enter a key, it must therefore be one that is held by the program itself—and shared by every copy of the program. So the effect is only that of a flag.

189. *See* Adobe, PDF Reference, [http://www.adobe.com/devnet/pdf/pdf\\_reference.html](http://www.adobe.com/devnet/pdf/pdf_reference.html) (last visited Mar. 20, 2010).

190. *See* Apple, What is Mac OS X—Graphics and Media, <http://www.apple.com/macosx/what-is-macosx/graphics-media.html> (last visited Mar. 20, 2010); Xpdf, Home, <http://www.foolabs.com/xpdf/home.html> (last visited Mar. 20, 2010).

191. Even to those without knowledge of C, this source code is fairly straightforward:

```
if (! doc->okToPrint()) {
    error(-1, "Printing this document is not allowed.");
    exitCode = 3;
    goto err1;
}
```

In English: If the document has no `okToPrint` indication, send an error message and abort, otherwise, continue with the user's request. A user who wanted to print a "printing disallowed" document could simply remove this conditional block;

```
// if (! doc->okToPrint()) {
```

Xpdf's code could have been obfuscated to make the check harder to find, but that would just mean a bit more work for the would-be-printer. Whether or not printing is allowed, the software must have access to the document to render it for on-screen display. At that point, the format relies on the software to enforce its restrictions, and software can be changed to ignore a simple flag and redirect the plaintext output to a printer as well as to a screen.<sup>192</sup>

The PDF specification's openness contributes to its popularity as a standard display and exchange format. Even before Adobe provided applications for most platforms, users could read and create PDFs on Unix and GNU/Linux systems as well as on Macintosh and Windows. Apple could decide at low overhead to build PDF support into its OS X operating system. Users of GNU/Linux operating systems such as Ubuntu can choose among Adobe's reader, xpdf, and ghostview, among others.<sup>193</sup> Independent developers can add features to the open-source xpdf or incorporate its functionality into new programs such as the pdf2html text–webpage generator. Users can create full-text indices of PDF files, facilitating better search, or independent programs to annotate PDF documents. Screen-reader applications can read the text aloud. Google and other search engines can parse the PDF files to include their text in search. Meanwhile, Adobe benefits from this ecosystem through royalties from additional sales of its enhanced PDF-reader and PDF-writer applications. Document authors benefit from the wide availability of tools, lowering the barriers to reading the works they make available. Adobe's open DRM format is successful

---

```
// error(-1, "Printing this document is not allowed.");  
// exitCode = 3;  
// goto err1;  
// }
```

If you don't check for the presence or value of `okToPrint`, a flagged document prints as easily as it displays.

This Author had to compile a print-friendly version of xpdf when, as a former editor of the Harvard Journal on Law & Technology (JOLT), she was contacted by one of that Journal's authors. That author had created a PDF version of his own article with the "Printing Not Allowed" and "Text Selection Not Allowed" flags set. After JOLT had published the article on its website, the JOLT author lost the original and wanted to recover the text. A modified xpdf enabled him to do this. This Author understands that similar functionality is now available in commercial programs.

192. It is likely that even if the reader were available only in binary form, it could still be reverse engineered, decompiled and edited to remove the flag check. Binary or obfuscated code would serve as just a minor obstacle to the determined flag-ignorner.

193. *See, e.g.*, Ubuntu Linux, Details of Package Pdf-viewer in Karmic, <http://packages.ubuntu.com/karmic/virtual/pdf-viewer/> (last visited Mar. 20, 2010) (listing seven open source pdf viewers).

because Adobe does not insist on making it robust. Contrast this minimal advisory DRM format with other more robust DRM strategies, including Adobe's own eBook format. These tie the two branches of the strategy together, using encryption and § 1201, to force the authorized reader to use a particular application and obey the strictures of that application.

Advisory anti-copying features may have a place, especially if they do not invoke the DMCA. But Adobe's attitude toward breaks of the PDF anti-printing feature contrasts with DMCA prosecution it instigated against Elcomsoft after Dmitry Sklyarov broke the encryption on its eBook format. Sklyarov, a Russian Ph.D. student, was in the country for a computer security conference at which he gave a presentation on the insecurity of Adobe's eBook encryption technology. Following the presentation, he was arrested and charged with trafficking in a product designed to circumvent copyright protection, in criminal violation of the DMCA, based on his employer's sale of a software program to read encrypted eBooks.<sup>194</sup> Adobe had encouraged the prosecution, but then encouraged the government to drop charges against Sklyarov individually after widespread public protest. A federal jury ultimately rejected DMCA charges against the company.<sup>195</sup>

For many companies, the DMCA inclines them toward behavior like Adobe's close guarding of its eBook format rather than its openness around the PDF specification, to the detriment of open-source and user-accessible technology. With these examples in mind, this Article considers the implications of anticircumvention's foreclosure of open development more deeply through the lenses of economic and legal research on distributed innovation.

## V. NEW CRITIQUE: ANTICIRCUMVENTION TAXES OPEN DEVELOPMENT AND USER INNOVATION

Beyond fair use, the DMCA extracts costs by foreclosing an entire mode of development. The DMCA, and the contractual ties built around its anticircumvention regime, foreclose open source development of media software and open hardware because it is impossible to build the secrecy DRM requires into a product designed for user-modification and collaborative development. This foreclosure adds yet another set of weights

---

194. Criminal Complaint, U.S. v. Sklyarov, No. 5-01-257 (N.D. Cal. July 17, 2001); *see also* Press Release, Dep't of Justice, First Indictment Under Digital Millennium Copyright Act Returned Against Russian National, Company, in San Jose, California (Aug. 28, 2001), available at <http://www.cybercrime.gov/Sklyarovindictment.htm>.

195. Lisa M. Bowman, *ElcomSoft Verdict: Not Guilty*, CNET NEWS.COM, (Dec. 17, 2002), [http://news.cnet.com/2100-1023\\_3-978176.html](http://news.cnet.com/2100-1023_3-978176.html).

to the structural balance described by Tim Wu and Yochai Benkler<sup>196</sup>: the costs not only of centralizing creation and innovation, but of locking users—those most familiar with their technology needs and wants—out of the design process. Perversely, anticircumvention exerts this development-closing effect just as legal scholars, as well as economists and management scholars, are recognizing the significant contributions from open source development, user innovation, and peer production.<sup>197</sup>

The particular impact of anticircumvention-induced closure may have been overlooked in a tallying of the costs of DRM because it has not been recognized, or because the mode-of-development tax has not been distinguished from the overall impact of DRM on innovation. While this problem shares many of the elements with generalized harm to innovation, it is distinctly deeper-rooted; it is not one that can be designed around within the framework of anticircumvention. Even were DRM designers to follow the urging of academics and try to make room for innovation, they would ultimately face an irreconcilable gap between DRM and openness.

This Article thus adds to the literature an analysis of the mode-of-development tax. Earlier Sections described the progression by which digital rights management, supported by anticircumvention law, is driven towards ever higher degrees of lock-out. Each reaction to a perceived threat (whether to media control or profitability) drives the locking attempts deeper into the core of product design: from simplistic devices to hinder copying, to more complicated designs, to “robust” mechanisms to protect those devices and assure they operate as intended, and finally to architectural mandates and forced incompatibility. This Part will now explore the cost of such a lock-out. Changing not only *who* can innovate but also *how* they may do so severely limits the scope of development and its zone of potential. Characterizing anticircumvention as a mode-of-development barrier explains why the law is problematic even when the DRM it supports is chosen by private inter-industry standard-setting rather than by government mandate. Anticircumvention provides the hook by which to demand *some* DRM, and no matter how open the process by which the DRM standard was developed, devices implementing it will have to be closed.<sup>198</sup>

---

196. See generally Yochai Benkler, *Freedom in the Commons, Towards a Political Economy of Information*, 52 DUKE L.J. 1245 (2003); Wu, *Intellectual Property, Innovation, and Decentralized Decisions*, *supra* note 154.

197. See generally YOCHAI BENKLER, *THE WEALTH OF NETWORKS* (2007); ERIC VON HIPPEL, *DEMOCRATIZING INNOVATION* (2005); Karim R. Lakhani & Eric von Hippel, *How Open Source Software Works: “Free” User-to-User Assistance*, 32 RES. POL’Y 923 (2003); Josh Lerner & Jean Tirole, *Some Simple Economics of Open Source*, 50 J. INDUS. ECON. 197 (2002).

198. This is the flaw in the purportedly “open source” model behind Sun’s DREAM

While DRM is ineffective against mass redistribution,<sup>199</sup> it imposes several costs on every would-be developer of interoperable devices: either licensing costs or the costs of circumventing the requested license, including the costs of breaking DRM (or finding an existing break); costs of assuring that other users will also be able to use the fix if developing for more than personal use; and costs of legal uncertainty. Law-backed DRM limits the potential upside to innovation because if a developer hopes to commercialize

---

platform. Even if anyone can build an implementation of the specification, it would win “authorization” to play protected content only after proving its un-modifiability by others as a prerequisite to obtaining permission. Developers writing such code would be unable to comply with the downstream “freedom to modify” condition of the Free Software Foundation GPL. *Cf.* Gerard Fernando, Tom Jacobs & Vishy Swaminathan, PROJECT DREAM, AN ARCHITECTURAL OVERVIEW (2005), <http://www.openmediacommons.org/collateral/DReaM-Overview.pdf>.

199. So long as DRM is trying to protect mass-distributed content, it faces an asymmetric challenge of adversaries as widely distributed as the interest in what it protects. *See* Stephen Lewis, *How Much Is Stronger DRM Worth?*, in *ECONOMICS OF INFORMATION SECURITY* 53, 54 (L. Jean Camp & Stephen Lewis eds., 2004).

Even with the strongest DRM mechanisms we have today, the BORA (break once run anywhere) principle still holds. Once content is retrieved from a DRM system, and re-encoded in a non-DRM protected form, the duplication of that content is as easy as moving the bits around. This means that the cost of breaking the DRM on a particular piece of content need only be borne once. The marginal costs of the duplication to the consumer who can obtain the content are near-zero, and furthermore the consumer need not expend any resources in breaking the DRM.

*Id.*; *see also* Stuart E. Schechter et al., *Trusted Computing, Peer-To-Peer Distribution, and the Economics of Pirated Entertainment* (May 2003), <http://www.eecs.harvard.edu/~stuart/papers/eis03.pdf> (describing the cost of pirated goods as a function of one-time extraction costs (or first-copy costs) and per-copy distribution costs).

So why is DRM not equally ineffectual against innovation? A different kind of hacking is needed for distribution of the extracted original versus innovation around uses of that original. To stop Darknet-enabled mass copying and redistribution, you need to stop every would-be copyist, because once one copy is out, it can be shared with others at much lower cost. Whereas to stop user innovation, you can throw up barriers that stop a group from reaching critical mass—such that each person needs to overcome the barrier. Of course, some breaks, in software that is not individualized, can be shared rather than having to be recreated. *See, e.g.*, Adam Pash, Jailbreak Your iPhone or iPod Touch with One Click, LIFEHACKER (Oct. 29, 2007), <http://lifehacker.com/316287/jailbreak-your-iphone-or-ipod-touch-with-one-click/>. It may also be easier to break than to interoperate—finding a buffer overflow versus actually understanding the obfuscated code.

While Paul Ohm’s *Myth of the Superuser* makes a persuasive argument against the focus of computer crime laws on the “superuser,” in the DRM context it takes only one moderately-super user to make unencrypted content available to all. Paul Ohm, *The Myth of the Superuser: Fear, Risk, and Harm Online*, 41 U.C. DAVIS L. REV. 1327, 1348 (2008). Fred von Lohmann refers to this as Mike Godwin’s “smart cow” problem. Fred von Lohmann, *Licensing in the Digital Age: The Future of Digital Rights Management*, 15 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 1009, 1042 (2005).

a product, she can predict having to seek a license, and expect that even if successful in obtaining that license, she will have to share the rents from innovation.<sup>200</sup>

Even if the usage rules posed no *practical* use restriction, such as that a device “must permit fewer than six billion copies,” the robustness rules, mandating their implementation by technological fiat, would still bind the hands of developers. To build a system capable of *assuring* that it makes no more than 5,999,999,999 reproductions, the builder must lock it down, forbidding the end-user from tampering with the system and voiding those guarantees. For a “trusted system” to be reliable, it must be hardened against its users.<sup>201</sup>

The same challenge meets those who would design systems to adjudicate fair use or to check with a third-party arbiter before permitting new uses.<sup>202</sup> Even if the arbitration is perfect, the judge attuned to all the nuances of free expression, the *system* must be hardened to be capable of enforcing those determinations robustly. Technology’s users must still be forbidden from modifying the core of their playback devices. Likewise, standardizing the DRM infrastructure might do away with custom DRM, making interoperation *among licensees* and marketing of works for the platform easier, but the standard would still bar user modification.

It is not possible to harden technical protection measures without closing devices to user development.<sup>203</sup> Trusted computing proposes a closed core

---

200. The potential for holdup is similar to that of blocking patents supported by injunctions. *Cf.* Mark A Lemley & Carl Shapiro, *Patent Holdup and Royalty Stacking*, 85 TEX. L. REV. 1991, 2010 (2006) (arguing that patent injunctions “encourage rent seeking by patent trolls and discourage innovation by firms that design and manufacture complex products”).

201. *See* SETH DAVID SCHOEN, TRUSTED COMPUTING, PROMISE AND RISK 10–11 (2005), [http://www.eff.org/files/20031001\\_tc.pdf](http://www.eff.org/files/20031001_tc.pdf).

202. *See supra* Section III.B.

203. *See* Reid & Caelli, *supra* note 46, at 1. Reid and Caelli state:

Once dissociated from its protection, the content can be freely copied, played, modified and redistributed, albeit in violation of the license terms. Consequently, to reliably enforce typical DRM policies, it must not be possible for the platform user to access the plaintext bits that represent the content, despite the practical reality that the platform is under the user’s direct control. This is an access control problem that cannot be solved purely by cryptography. On open computing platforms that can run arbitrary software, it is a difficult problem to which there is currently no practical, deployed solution, particularly in terms of ‘software-only’ techniques. Recent trusted computing initiatives, namely Microsoft’s Next Generation Secure Computing Base (NGSCB) and the Trusted Computing Group (TCG) specification, formerly known as TCPA aim in part, to address this issue through both hardware and software based methods.

with open interfaces. A system that offered users access to various features, while keeping the media decoding stream hidden, might be better than nothing, but it is not open source. It essentially lets users change the faceplates on their car radios, but not fix the tuners. In that mode, where users cannot manipulate the media stream directly, they are barred from implementing deeper features, such as modulating the audio to match sped-up video or adding echo cancellation.<sup>204</sup>

The copyright holders want a setup that cannot be made to leak the work, except through authorized outputs. They would prefer to have assurances up the viewer's eyeballs and eardrums, but if implants are impossible, they demand at least encrypted and signed pathways up to the analog output point of screen and speakers. While closed software and hardware can engage in the arms race, *appearing* secure until a smarter hacker comes along, open source software and user-modifiable hardware cannot even play the game. Their secrets are *intentionally* disclosed. Trusted computing, offered as the next generation solution for media on general-purpose hardware, does not allow for greater openness; it just pushes the closure into "trusted platform modules" and "remote attestation," and still forecloses open development of media software and hardware.<sup>205</sup> Moreover, trusted computing would not permit open source hardware, as Peter Gutmann notes in his catalog of the costs of Windows Vista content

---

The goal of trusted computing is to deliver systems that are highly resistant to subversion by malicious adversaries, allowing them to operate reliably and predictably in almost any circumstance. Trusted computing is an important ingredient in DRM because it provides a sound basis for license enforcement. Given the way the NGSCB and TCG initiatives have been promoted, one could be forgiven for thinking that trusted computing is an entirely new concept. As we discuss in Section 3.1, trusted computing actually has a long history but the lessons this history can teach have been largely ignored over the last 20 years, particularly in the design of mainstream PC operating systems. As a consequence, such systems are fundamentally ill equipped to provide the level of protection that a robust DRM system demands.

*Id.* at 1–2 (internal citations omitted).

204. See generally Seltzer, *supra* note 9; PETER GUTMANN, A COST ANALYSIS OF WINDOWS VISTA CONTENT PROTECTION (2007), [http://www.cs.auckland.ac.nz/~pgut001/pubs/vista\\_cost.html](http://www.cs.auckland.ac.nz/~pgut001/pubs/vista_cost.html).

205. Trusted computing provides a trusted platform, authenticated to be in a secure state with secure inputs and output paths, and audited source code that could "attest" to its integrity, checking in with a remote service for real-time verification before running with media input. See SCHOEN, *supra* note 201, at 4–5. In this case, the source code for this application could be delivered, but since a modified version would not pass the check-in, and thus would not actually be capable of functioning. Such tethered code would not pass the Open Source or Free Software definitions. See *id.*

protection.<sup>206</sup> Neither software nor hardware in a trusted computing platform is actually open and modifiable, even if its parts are visible under glass. This kind of one-way translucency would not support the vibrant innovation environment we see around genuinely open software and hardware. Openness is foreclosed by the design rules of DRM.

#### A. THE HIDDEN COSTS OF DRM

Beyond impeding fair use, anticircumvention's restrictions are troublesome because they explicitly bar a particular mode of development—the public mode of free and open source software development—that has been increasingly successful in both commercial and non-commercial production.<sup>207</sup> Anticircumvention forecloses end-user tinkering and innovation and cements a centralized industrial structure, just at the time when technology offers us the means and networked opportunity to do more from the distributed edges of the Internet.<sup>208</sup>

While large incumbent firms have many economic advantages over smaller competitors, including operating at large scales and minimizing transaction costs, recent research suggests that a diversity of sources can more effectively foster innovation.<sup>209</sup> Non-incumbent firms and individual user-innovators may be better positioned to stretch limits, pursuing avenues that might not have been explored by the dominant industry players. Thus companies without a base of existing customers may be more likely than incumbents to produce disruptive innovations that find support outside an established customer base; though the disruptive innovations fail at first to serve existing customers, they may ultimately improve to meet the demands of both old and new customers better.<sup>210</sup> Meanwhile, users themselves may adapt or innovate upon products, using information and motivation more readily available to them than to corporate manufacturers.<sup>211</sup> Under an anticircumvention regime, however, any innovator who would challenge the incumbent media industry is precluded by law that centralizes the “permission to innovate” with the copyright owner.

---

206. See GUTMANN, *supra* note 204.

207. See generally Lerner & Tirole, *supra* note 197.

208. See generally BENKLER, *supra* note 197; CLAY SHIRKY, *HERE COMES EVERYBODY: THE POWER OF ORGANIZING WITHOUT ORGANIZATIONS* (2008); Jonathan L. Zittrain, *The Generative Internet*, 119 HARV. L. REV. 1974 (2006).

209. See *infra* Section V.B.

210. For the leading exponent of disruptive innovation, see CLAYTON M. CHRISTENSEN, *THE INNOVATOR'S DILEMMA: WHEN NEW TECHNOLOGIES CAUSE GREAT FIRMS TO FAIL* 79–81, 132–34 (1997).

211. The leader in user innovation research is Eric von Hippel. See VON HIPPEL, *supra* note 197, at 19–22.

Anticircumvention affects different kinds of innovators differentially. Disruptive innovation by commercial-scale firms may still be possible in limited circumstances—if the would-be disrupter can obtain permission in advance. A newcomer to digital media who is an established player in another field might have the clout to negotiate authorization with sufficient relevant copyright holders, and by such authorization convert its actions to non-circumvention; individual users are unlikely to be able to do so. Even in an un-concentrated, uncoordinated media market, neither companies nor users will be able to obtain “permission to tinker” or to offer general-use media-access products modifiable by other users, because such access to the underlying media stream is precisely what DRM is designed to prevent. In a concentrated, coordinated market, a few privileged participants may get access to the bulk of media, but even then, they will be unable to offer open user-modifiable products.

#### B. DRM LIMITS DISRUPTIVE INNOVATION

Numerous industries have been dramatically reconfigured by “disruptive innovations,” non-linear developments that benefit the public through greater productivity, efficiency, or choice. As Clayton Christensen describes, disruptive technologies are often initially dismissed as inferior by those in the industry, but they soon catch up to and outpace the old ones.<sup>212</sup> These technologies have two key characteristics: “[f]irst, they typically present a different package of performance attributes—ones that, at least at the outset, are not valued by existing customers. Second, the performance attributes that existing customers do value improve at such a rapid rate that the new technology can later invade those established markets.”<sup>213</sup> Typically, Christensen finds, disruptive innovations are overlooked or cast aside by existing producers, even in “good” companies.<sup>214</sup> A company’s strengths at listening to existing customer demands (“sustaining” innovations) can deafen it to the appeals of a new product with different characteristics and initial targets.<sup>215</sup> Hence in industry after industry, competitors have usurped the spots of prior giants.<sup>216</sup>

Disruptive products or services appear at first to be inferior alternatives for mainstream customers. The market applications these products might fit

---

212. See CHRISTENSEN, *supra* note 210, at 79–86.

213. Joseph L. Bower et al., *Disruptive Technologies: Catching the Wave*, HARV. BUS. REV., Jan.–Feb. 1995, at 44.

214. CHRISTENSEN, *supra* note 210, at 207.

215. *Id.* at 20.

216. See CLAYTON M. CHRISTENSEN & MICHAEL E. RAYNOR, *THE INNOVATOR’S SOLUTION: CREATING AND SUSTAINING SUCCESSFUL GROWTH* at 34–43 (2003).

offer lower profit margins, making them less attractive diversions of attention for an established company. A newcomer facing a different set of opportunity costs may be willing to experiment, however, selling to the lower-margin customer segments.<sup>217</sup> As it does so, it improves its competing product. The “disruption” occurs when some of these improved competitors attract mainstream attention, migrating from the sidelines to overtake the prior standard.<sup>218</sup> The erstwhile market leader is left to regret not focusing more attention on smaller, once-inferior disk drives or minicomputers.<sup>219</sup>

Christensen suggests that incumbents can learn to innovate disruptively,<sup>220</sup> but the sheer numbers and risk appetite of competitors make it likely more radical change will come from outside. An incumbent who recognizes this pattern of disruptive innovation but is unable to identify the potentially successful disruptive technology (and grab the profits for itself) may prefer to block new technology altogether. If the incumbent can do so using intellectual property, it can preserve its own position for a bit longer at the expense of a public denied the opportunities of technological improvement. It takes less foresight to seek stability by blocking others from innovating than to innovate for oneself.

The VCR and MP3 could both be considered disruptive technologies. The initial recording capacity and quality of the Betamax made it poorly suited for television broadcasters or movie studios to use in-house, but to television viewers with no other option for recording shows, poor quality was better than nothing.<sup>221</sup> As the quality was shown to be good enough, and the capacity improved, a virtual network of VCR-owners developed, providing opportunities for the rental of pre-recorded movies as well.<sup>222</sup> Similarly, the MP3's compressed audio was dismissed by audiophiles as inferior for music-listening, but its smaller file size made it popular with fans, who could “rip” MP3s to music players and trade them over even slow Internet connections.<sup>223</sup> In time, of course, both found sizeable audiences.

---

217. CHRISTENSEN, *supra* note 210, at 26.

218. *Id.* at 77–95.

219. *Id.* at 134–38.

220. *See* CHRISTENSEN & RAYNOR, *supra* note 216, at 33–35.

221. *See* JAMES LARDNER, FAST FORWARD: HOLLYWOOD, THE JAPANESE, AND THE ONSLAUGHT OF THE VCR 95–97 (1987). This argument, which considers Betamax and VHS together, is distinct from the ongoing debate about whether the triumph of VHS over Beta illustrates path dependence. *See generally* S.J. Liebowitz, & Stephen E. Margolis, *Path Dependence, Lock-in, and History*, 11 J.L. ECON. & ORG. 205 (1995).

222. *See* LARDNER, *supra* note 221, at 312–20. On network effects of technology, see CARL SHAPIRO & HAL R. VARIAN, INFORMATION RULES: A STRATEGIC GUIDE TO THE NETWORK ECONOMY 183 (1998).

223. *See* STEVEN LEVY, THE PERFECT THING: HOW THE IPOD SHUFFLES COMMERCE,

The entertainment companies' response to both VCR and MP3 as they became popular was similar to their reaction to newer technologies under the DMCA: attempt to sue them away.<sup>224</sup> None of the existing copyright laws gave them traction, however. The Supreme Court ruled that Sony was not liable for enabling a combination of non-infringing, fair, and possibly infringing uses of its Betamax. Instead, the Court held that a device "capable of substantial non-infringing use," including that of "time-shifting" broadcast programming, was lawful to sell.<sup>225</sup>

The Diamond Rio, introduced in 1998, offered users portable storage for an hour's worth of MP3-encoded music, imported from a computer.<sup>226</sup> The Recording Industry Association of America sued, claiming that the device violated the Audio Home Recording Act (AHRA).<sup>227</sup> The Ninth Circuit rejected the argument on statutory grounds: The AHRA requires that "digital audio recording devices" contain serial copyright management features to bar second-generation copying, but because the Rio obtained its music through a personal computer (not primarily a music copying device) it was out of scope of the AHRA's mandate.<sup>228</sup> The decision properly found the AHRA's scope limited to special-purpose digital audio recorders (effectively rendering it irrelevant), rather than requiring every general-purpose computer to be restricted to watch for and respond to anti-copying watermarks. The *Diamond* court found that the Rio's space-shifting was likely non-infringing: "In fact, the Rio's operation is entirely consistent with the Act's main purpose—the facilitation of personal use."<sup>229</sup> Pre-DMCA copyright holders could not claim a right to control or bill for personal uses, even if those uses incidentally required copying.

These technologies disrupted existing entertainment production-and-distribution business models, while providing the public more access to art and entertainment. Under prior copyright law, companies in the entertainment business had to adapt or fail. But what pre-existing copyright

---

CULTURE, AND COOLNESS 144–55 (2006); Robert Levine, *The Death of High Fidelity*, ROLLINGSTONE.COM, Dec. 27, 2007, [http://www.rollingstone.com/news/story/17777619/the\\_death\\_of\\_high\\_fidelity/print/](http://www.rollingstone.com/news/story/17777619/the_death_of_high_fidelity/print/).

224. *See, e.g.*, Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984); Recording Industry Ass'n of Am. v. Diamond Multimedia Sys., Inc. 180 F. 3d 1072 (9th Cir., 1999).

225. *Sony*, 464 U.S. at 456.

226. LEVY, *supra* note 223, at 27.

227. *Diamond*, 180 F. 3d at 1075.

228. *Id.* at 1078.

229. *Id.* at 1079.

law failed to control, the anticircumvention laws give new ammunition against.

Although music can still be accessed from unencrypted CDs, unburdened by technical protection measures, innovation upon and around other kinds of content has become harder. The difference the DMCA made was apparent from the first case decided under the new law in 2000: *Real Networks v. Streambox*.<sup>230</sup> Streambox developed the “Streambox VCR,” a bit of software to play Real video streams, the then-dominant format of streaming video, and save them. Real sued for circumvention and won. To receive RealNetworks video, the Streambox VCR had to mimic the “Secret Handshake” of the RealPlayer client, a move the district court found to be circumvention.

[A]t least a part of the Streambox VCR circumvents the technological measures RealNetworks affords to copyright owners. Where a RealMedia file is stored on a RealServer, the VCR “bypasses” the Secret Handshake to gain access to the file. The VCR then circumvents the Copy Switch, enabling a user to make a copy of a file that the copyright owner has sought to protect.<sup>231</sup>

By contrast to the pre-DMCA world, when movie studios lost their bid to stop the Betamax,<sup>232</sup> now it sufficed for Real to say that its Secret Handshake (not so secret, since every streaming client knew it) and Copy Switch were “technological measures” for it to win an injunction against the Streambox VCR’s further production or distribution.<sup>233</sup> Users who wanted to time-shift online video—and the companies who wanted to offer them that and other options beyond those of Real’s product—were out of luck.<sup>234</sup>

Other disruptive innovators have managed to negotiate with those who control entertainment copyrights—Apple’s iTunes music store became the first commercially successful music store after various music companies had tried and failed.<sup>235</sup> Apple could secure copyright licenses and then, using the

---

230. *RealNetworks, Inc. v. Streambox, Inc.*, No. C99-2070P, 2000 U.S. Dist. LEXIS 1889 (W.D. Wash. Jan. 18, 2000).

231. *Id.* at \*20.

232. *See Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417 (1984).

233. *RealNetworks*, 2000 U.S. Dist. LEXIS 1889, at \*18–\*19.

234. The *Streambox* court held that the DMCA eliminated *Sony*’s “substantial non-infringing use” defense. “Streambox’s primary defense to Plaintiff’s DMCA claims is that the VCR has legitimate uses . . . [but] ‘equipment manufacturers in the twenty-first century will need to vet their products for compliance with Section 1201 in order to avoid a circumvention claim, rather than under *Sony* to negate a copyright claim.’” *Id.* at \*21–\*23 (quoting DAVID NIMMER, NIMMER ON COPYRIGHT § 12A.18[B] (1999 Supp.)).

235. LEVY, *supra* note 223, at 168–70.

DMCA, create a format with which others were forbidden from interoperating.<sup>236</sup> With that combination, Apple could use the music to sell its iPod music players, the only ones with “authorization” to decode the encrypted AAC tracks, and then lock out other innovators from that piece of the market.<sup>237</sup>

While disruption is painful to those whose businesses are leapfrogged, it generally benefits end-users. Through competition, they get access to a wider range of products, better tailored to their needs in feature selection or price.<sup>238</sup> Customers who cannot attract the attention of a major producer, on whose scale they would be just a speck, may be able to find a supplier elsewhere who sees them as opportunities to break in to a new market. Some would-be disruptors fail, of course, but the larger number of innovators who can try when barriers to entry are lower gives more opportunities for unexpected successes.<sup>239</sup>

Armed with anticircumvention law, however, media companies can barricade themselves against disruption by such upstarts. By denying the critical “authorization” to anyone who would operate differently, the incumbents can lock out challengers. Fewer authorized innovators means fewer entrepreneurs free to pursue their own experiments around what the markets and public might support. The requirement of advance permission—and the possibility that permission will be denied or conditioned—imposes a hurdle that will stop many. Particularly around media where many are frustrated by the slow pace of entertainment companies’ technology adoption, the public loses when competitors are forbidden from experimentation.

---

236. See Peter Burrows, *DoubleTwist is Dancing in Dangerous Legal Territory*, BUSINESS WEEK, Feb. 19, 2008, [http://www.businessweek.com/technology/ByteOfTheApple/blog/archives/2008/02/doubletwist\\_is.html/](http://www.businessweek.com/technology/ByteOfTheApple/blog/archives/2008/02/doubletwist_is.html/).

237. Apple’s recent decision to drop the DRM from its iTunes store does not contradict this story. DRM served to lock users into the Apple ecosystem early, as the iPod was taking off, but now annoys those same users as they try to move tracks among a growing number of Apple products. Meanwhile, compatibility with custom-designed software and peripherals can now serve to perpetuate the lock-in. See Brad Stone, *Want to Copy iTunes Music? Go Ahead, Apple Says*, N.Y. TIMES, Jan. 7, 2009, at B1.

238. See generally W. KIP VISCUSI, JOSEPH E. HARRINGTON & JOHN M. VERNON, *ECONOMICS OF REGULATION AND ANTITRUST* (4th ed. 2005).

239. Distributed innovation thus functions like a diversified portfolio of options, giving more chances to succeed in the face of uncertainty. Some might argue that the incentives to compete are lower when the potential gain is a smaller, more rapidly disputed monopoly. However, as we see, *infra* Part V, the motivations to innovate are more varied than just the size of the profit pot, suggesting that innovation and entrepreneurship will continue to flourish even if not assured a winner-take-all outcome.

Yet if the circumstances are bad for would-be disruptive companies, they are worse for individuals or those seeking to enable end-user innovation. While innovation is still possible by those who can convince incumbents that they will share the profits, *end-user* innovation is precluded entirely in the domains where TPMs are in force. Robustness rules lock out open-source or user-accessible design.

### C. DRM LIMITS USER INNOVATION

Eric von Hippel has led the field in describing the ways in which users—and not just manufacturers—improve upon and innovate the products they use.<sup>240</sup> Because they are closer to their problems and benefit directly from their solutions, end-users often have better information and greater motivation to improve products than manufacturers. As von Hippel finds, from ten to forty percent of users across a range of fields engage in developing or modifying products, rather than merely “consuming” them.<sup>241</sup>

By innovating for themselves, end-users can obtain products not available commercially, products that may be unavailable because manufacturers have not gotten there yet or because it is not economical for manufacturers to offer the variety of items that would include all of their users’ distinct individual preferences. Once lead-users have demonstrated the value of innovations and their commercial potential, however, manufacturers or a user community may develop it on a larger scale. Von Hippel and colleagues have found this user innovation pattern in business and consumer products alike, in fields from scientific and medical instruments to mountain bikes and rodeo kayaks.<sup>242</sup>

The current technology environment is particularly fertile for user innovation around consumer products, particularly in digital media. The Internet reduces costs of communication, enabling user communities to share information and development more cheaply and rapidly.<sup>243</sup> Where the product is bits, the Internet also reduces to near-zero the cost of distribution. End-users can get involved with little start-up capital. We have seen “DIY culture” reinvigorated as people learn to manage the complexity of computer technology and leverage it to their own ends: user innovation is enabled and

---

240. See generally VON HIPPEL, *supra* note 197; ERIC VON HIPPEL, SOURCES OF INNOVATION (1988).

241. VON HIPPEL, *supra* note 197, at 20 (2005).

242. See VON HIPPEL, SOURCES OF INNOVATION, *supra* note 240, at 11; Christoph Hienerth, *The Commercialization of User Innovations: The Development of the Rodeo Kayak Industry*, 36 RES. & DEV. MGMT. 273 (2006); Christian Lüthje, Cornelius Herstatt & Eric von Hippel, *User-innovators and “Local” Information: The Case of Mountain Biking*, 34 RES. POL’Y 951 (2005).

243. See Zittrain, *supra* note 208, at 1988.

manifested in do-it-yourself electronics and mechanical projects, from Make Magazine and its Maker Faires to Lego Mindstorms.<sup>244</sup> Online, the World Wide Web has accelerated the development of expression, software, and mashups (blogging, scripting, photography, videography, and mapping, to name a few).<sup>245</sup> Yochai Benkler suggests that these conditions for decentralized decision-making enable commons-based peer-production to improve upon market-based or hierarchical organization.<sup>246</sup>

The rising popularity of free and open source software both reflects the interest in user-accessible products and provides toolkits for it. With the source code available and freely modifiable, users can reconfigure software products or hire independent consultants to do so for them, even when their modification demands do not rise to the scale or expected profitability sufficient to interest a commercial supplier. Both hobbyists and commercial vendors have been willing to share their software's source code, often collaborating on the same project. While their motives may differ, user-producers of both types recognize value from enabling end-user investigation and modification.<sup>247</sup>

User innovation is not merely an interesting alternate source of products, it enhances social value. Joachim Henkel and Eric von Hippel “conclude that an innovation system where user innovation is present is welfare superior to one where it is not.”<sup>248</sup> Even assessing only aggregate value—a society's total income and not its distribution—user-innovated products tend to suit their users more precisely, leaving less deadweight loss in the usual mismatch between product and function.<sup>249</sup> User innovation is less subject to the “consumer surplus effects” that can limit manufacturers' incentives to innovate when they feel they will be unable to capture the total value of new

244. See Tim O'Reilly, *Where Real Innovation Happens*, FORBES.COM, Feb. 2, 2009, [http://www.forbes.com/2009/02/03/innovation-tim-oreilly-technology-breakthroughs\\_0203oreilly.html](http://www.forbes.com/2009/02/03/innovation-tim-oreilly-technology-breakthroughs_0203oreilly.html) (“[A]lpha geeks exercise an idea or a gadget, push[] it past its current limits, reinvent[] it and eventually pav[e] the way for entrepreneurs who figure out how to create mainstream versions of their novel ideas.”); Makezine.com, About MAKE, <http://makezine.com/about/> (last visited Feb. 22, 2010).

245. See Zittrain, *supra* note 208, at 1994.

246. See Yochai Benkler, *Coase's Penguin, or Linux and The Nature of the Firm*, 112 YALE L.J. 369, 375 (2002).

247. See Karim Lakhani & Robert Wolf, *Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects*, in PERSPECTIVES ON FREE AND OPEN SOURCE SOFTWARE 12, 12–15 (Joseph Feller et al. eds., 2005); Josh Lerner & Jean Tirole, *Some Simple Economics of Open Source*, 50 J. INDUS. ECON. 197, 212–15 (2002).

248. Joachim Henkel & Eric von Hippel, *Welfare Implications of User Innovation*, 30 J. TECH. TRANSFER 73, 74 (2004).

249. *Id.* at 78.

products; as simultaneous producers and consumers, user innovators do not feel loss from this effect. User innovation thus increases social welfare through several vectors, including better products, better R&D, and, not to be underestimated, the enjoyment of innovation itself.<sup>250</sup>

First, user innovation may produce more valuable products directly. That is, user-innovators may become sources of product, for themselves or for others through non-commercial sharing or commercial entrepreneurship. They contribute new products and better customized versions of existing products, often, products that would not have been created—at least not so early or with the same attributes—if all development were commercial supplier-driven.<sup>251</sup> Users have advantages in development, such as the ability to iterate quickly to improve and respond to changing needs. They have better information about what they need and do not filter those needs through the barriers of cross-discipline communication, nor support the costs of disaggregating “sticky” information.<sup>252</sup>

Eric Raymond, computer programmer and open source advocate, explains why he began developing the fetchmail program:

I needed a POP3 client. So I went out on the Internet and found one. Actually, I found three or four. I used one of them for a while, but it was missing what seemed an obvious feature, the ability to hack the addresses on fetched mail so replies would work properly. . . . This was clearly something the computer ought to be doing for me. But none of the existing POP clients knew how!<sup>253</sup>

Working to “scratch[] [the] developer’s personal itch,” and soliciting the bug reports and code contributions of other user-developers, Raymond turned his stub of code into a robust mail-delivery program, tailored to the features he and the community needed in practice.<sup>254</sup>

Eric von Hippel tells similar innovation stories, from clinical chemists designing their own assays for research use, to mountain bikers building or modifying their equipment.<sup>255</sup> More than half of experimental results reported in the chemical literature derived from user-designed and -adapted

---

250. *Id.* at 81–82.

251. *See* VON HIPPEL, SOURCES OF INNOVATION, *supra* note 240, at 14–15.

252. Henkel & von Hippel, *supra* note 248, at 75; Eric von Hippel, “Sticky Information” and the Locus of Problem Solving: Implications for Innovation, 40 MGMT. SCI. 429, 430 (1994).

253. ERIC S. RAYMOND, THE CATHEDRAL AND THE BAZAAR 23 (2001).

254. *Id.* at 23–27.

255. VON HIPPEL, SOURCES OF INNOVATION, *supra* note 251, at 11; *see also* Lüthje et al., *supra* note 242, at 951.

tests.<sup>256</sup> Nineteen percent of mountain bike enthusiasts reported developing and building components of their equipment, often using skills from hobby or professional background.<sup>257</sup>

Thus users—whether in their roles as amateur hobbyists, professional scientists, or programmers—often build or adapt tools for their own use, to serve previously unmet needs. Some users become developer–distributors, as Raymond did maintaining the fetchmail client for a growing user base. Their innovations may later be followed by and shared with other similarly situated users.

Second, end-users can contribute to commercial innovation indirectly, serving as research labs for commercial suppliers. Even when the end-users do not themselves produce in large scale, their innovations may be adopted by established firms. As von Hippel describes, this diffusion is facilitated because user-innovators often “freely reveal” their improvements, either because it is more convenient for them to speak openly than to be secretive, or because they gain more from the ease with which other users can contribute than they would from a competitive head start.<sup>258</sup> Especially if they do not themselves plan to produce the product commercially or use it as a key component of a business process, end-users may see no disadvantages to revealing—no advantages to trade secrecy that justify the costs of keeping secrets—and may see benefits to having their small-scale efforts picked up for commercial use.<sup>259</sup> Thus users of open source software often contribute their modifications back as patches to the main source tree.<sup>260</sup> Along with recognition and reputational advantages, contributors can save themselves work, gaining better assurance of continued compatibility with a wider range of possible complementary applications,<sup>261</sup> bringing “more eyes” to their bugs, and, perhaps, improving the fixes further. The clinical chemists’ independently developed tests are now available for purchase, pre-packaged, for their developers and the broader research community. Free revealing enables users to get both non-commercial and commercial support and commercial firms to add their expertise in larger-scale manufacture or distribution, using the community of lead users as a test bed or field research lab.

---

256. VON HIPPEL, *SOURCES OF INNOVATION*, *supra* note 251, at 11.

257. Lüthje et al., *supra* note 242, at 957, 961–62.

258. VON HIPPEL, *supra* note 197, at 77–91.

259. Lüthje et al., *supra* note 242, at 954.

260. Lakhani & von Hippel, *supra* note 197, at 926.

261. In software, contributing patches to the main codebase saves developers the work of reapplying the patches with each new release.

Further, the *process* of innovation may itself be rewarding to the user–innovator—it offers community, intellectual stimulation and development of new skills, and engagement with technology.<sup>262</sup> Engagement in development may make users happier with results by changing baseline expectations for products: a user may be more satisfied with a self-made product or improvement and more forgiving of its rough edges than if he had obtained the same product commercially.<sup>263</sup>

In addition to differences in the quality and variety of products and services developed, user innovation produces distributional benefits. The distribution of wealth and access may be fairer in a field open to user innovation than in one closed to it. Access may be more democratic, open to those who are time-rich and money-poor (and offering new fields of entrepreneurship by which people may make time into money). Particularly users with niche needs will be better served by self-innovating. Moreover, the user-innovator is empowered to think of him or herself as more than a mere consumer, and perhaps, like the Jeffersonian yeoman farmer, to become more involved in governance of the information environment.<sup>264</sup>

That society as a whole is better off when fields are open to user innovation does not, however, ensure that enough participants have the incentive to provide for it. While some companies recognize the opportunities user innovation provides (IBM, famously, contributes to Linux because better software, widely distributed and open, complements its proprietary hardware),<sup>265</sup> others feel threatened by the potential competition. They may fear losing the first-mover advantage of trade secrecy or prefer licensing revenue to its absence (e.g. licensing a smaller pool of technology rather than manufacturing a part of a big pool or making ancillary revenues).<sup>266</sup> Jonathan Zittrain suggests that even users will react against generatively open platforms if those platforms are more easily overrun by bad actors.<sup>267</sup>

---

262. See BENKLER, *supra* note 197, at 122–27; CHRIS DIBONA, SAM OCKMAN & MARK STONE, OPEN SOURCES: VOICES FROM THE OPEN SOURCE REVOLUTION 13 (1999). See generally Ed Felten et al., Freedom to Tinker, <http://freedom-to-tinker.com/> (last visited Mar. 7, 2010).

263. VON HIPPEL, *supra* note 197, at 33–43.

264. This Author uses governance by reference to ELINOR OSTROM, GOVERNING THE COMMONS (1990), to mean voluntary self-organization to overcome collective action problems and manage common pool resources.

265. See David Berlind, *Open Source: IBM's Deadly Weapon*, ZDNET, Apr. 8, 2002, [http://news.zdnet.com/2100-10532\\_22-296366.html](http://news.zdnet.com/2100-10532_22-296366.html).

266. See Thomas R. Eisenmann, Geoffrey Parker & Marshall Van Alstyne, *Opening Platforms: How, When and Why?* (Harvard Bus. Sch., Working Paper No. 09-030, 2008), at 5.

267. JONATHAN L. ZITTRAIN, THE FUTURE OF THE INTERNET AND HOW TO STOP IT 8

For those reluctant to be innovated against, anticircumvention has proven a powerful means of blocking user innovation, whether deliberately or incidentally. User innovation depends upon openness and accessibility of the underlying product.<sup>268</sup> Christina Raasch describes the impact of technical accessibility in the International Moth racing sailboat. This class is marked by a high degree of user-development, but as the technological complexity of the hull materials increased (in a shift from plywood to fiber-reinforced plastic and then carbon fiber), most users stopped innovating on hull design, although they continued to innovate elsewhere in the boats.<sup>269</sup> Robustness rules similarly increase the barriers to user innovation in media technologies.

D. THE OVER-ARCHING COST: DRM CENTRALIZES INNOVATION,  
OPPOSING COPYRIGHT'S ORIGINAL GOALS

Contrary to the main trunk of copyright, the anticircumvention branch centralizes decision-making authority. Copyright as a whole decentralizes choices about the kinds and quantities of creative works that should be produced. By offering property rights of exclusion, copyright enables markets in creative works despite the basically non-excludable nature of creative expression.<sup>270</sup> Markets decentralize decision making<sup>271</sup>: instead of waiting for the state cultural minister or a wealthy patron to finance a new work of cinematic art, Walt Disney Co. can listen to the demands of millions of princess-favoring kids and their indulgent parents. Traditional copyright thus lets producers of creative works of art, literature, and music self-organize to meet what they perceive to be the interests of their audiences.<sup>272</sup> In a world of imperfect information, decentralization gives us more chances to match consumer interests and a more democratic opportunity to serve

---

(2008). Thus we cannot invoke the invisible hand to ask, “if it’s so good, why aren’t we there yet?”

268. The characteristics that make a product conducive to user innovation are similar to those Zittrain identifies as key to a technology’s “generativity”: capacity for leverage, adaptability, ease of mastery, and accessibility. Zittrain, *supra* note 208, at 1981. Where some of the characteristics of generativity, particularly “capacity for leverage,” are directed to the creation of something else using the technology, user innovation focuses on changing the technological product directly. *Id.*

269. Christina Raasch, Cornelius Herstatt & Phillip Lock, *The Dynamics of User Innovation*, 12 INT’L J. INNOVATION MGMT. 377, 390 (2008).

270. See Wendy Gordon, *Fair Use as Market Failure*, 82 COLUM. L. REV. 1600, 1610–12 (1982).

271. F.A. Hayek, *The Use of Knowledge in Society*, 35 AM. ECON. REV. 519, 526 (1945).

272. See Wu, *Intellectual Property, Innovation, and Decentralized Decisions*, *supra* note 154, 146–47.

them.<sup>273</sup> Tim Wu suggests that intellectual property regimes should be assessed by their effects on the structure of decision making.<sup>274</sup>

Along with the economic benefits of letting markets organize production and enabling disruptive innovation, decentralization has social and cultural value. It helps to produce a democratic information environment, where everyone is a potential creator and consumer,<sup>275</sup> a read–write culture.<sup>276</sup> It lets users self-organize, in the mode Yochai Benkler terms “commons-based peer-production.”<sup>277</sup>

While copyright decentralizes independent production, some have also noted that it centralizes control over follow-on expression. Through the exclusive rights over reproduction and derivative works, the initial creator can control use of his work as input if the use exceeds fair use.<sup>278</sup> There, they conclude that the problems get more serious (costs rise against benefits) as copyright’s restrictions become more severe, apply to more conduct, and last longer.<sup>279</sup>

Yet if basic copyright, by creating markets, decentralizes at least the independent production of expressive works, the DMCA’s anticircumvention provisions centralize control of the technologies that work with them, swapping one set of coordination problems for another. Where we celebrate copyright’s “romantic author” and resisted centralized decision making for works of creative expression,<sup>280</sup> we have centralized innovation around the technological means to enjoy the created texts, sounds, and movies.

273. Even here, copyright is not costless. Economists discuss the tradeoffs between static costs and dynamic benefits. We accept the inefficiency of monopolies in individual works in exchange for the innovation derived from competition among them over shares of the broader “literature” or “entertainment” markets.

274. Wu, *Intellectual Property, Innovation, and Decentralized Decisions*, *supra* note 154, at 123–24.

275. WILLIAM FISHER, PROMISES TO KEEP 20 (2004).

276. LESSIG, *supra* note 86, at 28.

277. See Benkler, *supra* note 246, at 376.

278. See JAMES BOYLE, THE PUBLIC DOMAIN: ENCLOSING THE COMMONS OF THE MIND (2008); Derek Bambauer, *Faulty Math: The Economics of Legalizing ‘The Grey Album,’* 59 ALA. L. REV. 345 (2007); Benkler, *supra* note 129; Mark Nadel, Questioning the Economic Justification for (and thus Constitutionality of) Copyright Law’s Prohibition Against Unauthorized Copying: § 106 (2003), AEI-BROOKINGS JOINT CTR., RELATED PUBL’N 3-1 (Jan. 2003), <http://www.reg-markets.org/admin/pdffiles/Nadel.pdf>.

279. For a fictional treatment of the expanding-derivatives problem, see SPIDER ROBINSON, MELANCHOLY ELEPHANTS (1983), available at <http://www.spiderrobinson.com/melancholyelephants.html/> (describing a future in which nothing can be composed because everything imagined is too similar to works already under perpetual copyright).

280. See ROSEMARY COOMBE, THE CULTURAL LIFE OF INTELLECTUAL PROPERTIES:

Wu makes a similar critique of copyright's "communications policy," interpreting the current Copyright Act as a product of incumbent disseminators' rent-seeking activities to keep upstart challengers out of the market.<sup>281</sup> The pattern fits anticircumvention, helping to explain the expansion of the zone of authorization and control from authors of copyrighted works to developers of technologies of copy-protection. Not only do authors and copyright holders seek to capture all possible rents around their copyrighted works, non-authors try to use others' copyrights as a guarantor of profits. In the centralized marketplace, a few stagnating incumbent media companies<sup>282</sup> can use the continuing market power of back-catalog to prevent technologists or competing independents from inventing new models of media use.

## VI. CONCLUSION

When we grant anticircumvention control to copyright holders, we foreclose a set of possibilities and a solution space for the challenges of cultural exchange and technical productivity.<sup>283</sup> It is hard to quantify what does not yet exist,<sup>284</sup> but comparisons from other, less encumbered fields, and the not-too-distant past of media have suggested several reasons to prefer openness here.

The DMCA's anticircumvention provisions centralize the production of media playback technology, giving copyright holders the right to authorize—or forbid—the interoperation with their copyrighted works. This is a new phenomenon. In the "old world" of copyright, once the copyright holder had exercised his first sale rights, the public gained the ability and opportunity to use the work in a variety of ways that did not implicate copyright.

Anticircumvention closes the frontiers of media innovation. The open frontier is not just specific possibility, but an inspiration, an invitation to explore.<sup>285</sup> Not all the prospectors searching California for gold found that, but their risk-fueled exploration set the stage for commercial development of

---

AUTHORSHIP, APPROPRIATION, AND THE LAW 219 (1998).

281. Wu, *Copyright's Communications Policy*, *supra* note 154, at 325–28.

282. See WILLIAM PATRY, *COPYRIGHT WARS AND MORAL PANICS* 173 (2009).

283. See Timothy F. Bresnahan, & Manuel Trajtenberg, *General Purpose Technologies: "Engines of Growth"?*, 65 J. ECONOMETRICS 83, 84 (1995) ("Most [general purpose technologies] play the role of 'enabling technologies,' opening up new opportunities rather than offering complete, final solutions.").

284. And harder still to fund lobbying efforts on behalf of technology yet-to-be-invented.

285. Cf. FREDERICK J. TURNER, *THE FRONTIER IN AMERICAN HISTORY* 25–32 (1921) (attributing democratic success to the availability of a "public domain" of free land).

the American West. In the course of exploration and mapping the space, innovators may make unexpected discoveries and find new sources of value. Even if their hoped-for gold rush does not pan out, their exploration of the space may pave the way for other innovations of great aggregate importance.

Anticircumvention encourages copyright holders to stake out the frontier of technological innovation with “no trespassing” signs, not because they have explored and settled the territory to develop it productively, but because they feel threatened if others do so. Anticircumvention sends a message to developers, both commercial and user-innovators, that certain activities and opportunities are off limits, that even if it is technically feasible to improve interoperation with a wide variety of media, they are forbidden from doing so without advance permission. The vagaries (and transaction costs) of the permission-granting mechanisms deter innovators, as do the prospects of being forced to share the rents.

Anticircumvention serves as public law, enforcing private law, to forbid tinkering and block distributed user innovation. As a matter of regulatory design, this kind of architectural regulation externalizes costs. It lets those who benefit, a set of incumbent copyright holders, pretend that the imperfect DRM is good, while imposing a mode-of-development tax on the entire public. In the full cost-benefit analysis of anticircumvention, the loss to open user innovation outweighs the gains from this imperfect mechanism of copyright enforcement. Treating code literally as law leaves the law with too many harmful side effects.

# LAW, TECHNOLOGY, AND SHIFTING POWER RELATIONS

*Bert-Jaap Koops*<sup>†</sup>

“I thought he was going to get pneumonia, but actually he said in his letter it wasn’t the cold that bothered him, it was being watched all the time. The eye in the door.” . . . This eye, where no eye should have been, was deeply disturbing to Prior. . . . “That’s horrible,” he said, turning back to Beattie. “’S not so bad long as it stays in the door.” She tapped the side of her head. “You start worrying when it gets in here.”<sup>1</sup>

## TABLE OF CONTENTS

I.	<b>INTRODUCTION</b> .....	974
II.	<b>PRELIMINARIES</b> .....	976
	A. POWER RELATIONS .....	976
	B. LEGAL PROTECTION: INEQUALITY COMPENSATION.....	977
	C. TECHNOLOGY .....	978
III.	<b>LAW ENFORCEMENT–CITIZEN</b> .....	979
	A. CASE STUDY 1: DNA FORENSICS .....	980
	B. CASE STUDY 2: INTERCEPTION OF TELECOMMUNICATIONS .....	984
	C. CASE STUDY 3: PASSENGER NAME RECORDS.....	987
	D. DISCUSSION.....	989
IV.	<b>EMPLOYER–EMPLOYEE</b> .....	996
	A. CASE STUDY 1: WORKPLACE MONITORING.....	996
	B. CASE STUDY 2: LOCATION MONITORING .....	1000
	C. DISCUSSION.....	1001
V.	<b>BUSINESS–CONSUMER</b> .....	1006

---

© 2010 Bert-Jaap Koops.

<sup>†</sup> Professor of Regulation & Technology, Tilburg Institute for Law, Technology, and Society (TILTS), Tilburg University, the Netherlands. MSc (mathematics) and MA (literature), Groningen University, PhD (law), Tilburg University. The research for this Article was funded by the Netherlands Organisation for Scientific Research (NWO), whom I thank for its generous support. This Article presents the conclusions of a five-year project on law, technology, and shifting balances of power; it builds on the results of subprojects and hence will refer relatively frequently to earlier publications by me and my co-researchers in the project, Dr. Colette Cuijpers and Dr. Merel Prinsen.

1. PAT BARKER, THE EYE IN THE DOOR 36 (Plume 1995) (1993).

A.	CASE STUDY 1: PROFILING AND BEHAVIORAL ADVERTISING.....	1006
B.	CASE STUDY 2: BUYING ONLINE.....	1010
C.	DISCUSSION.....	1014
VI.	<b>THE IDENTITY OF THE CITIZEN–CONSUMER– EMPLOYEE</b> .....	1018
A.	ROLE-PLAYING, IDENTITY, AND SELF-DEVELOPMENT.....	1019
B.	A DIGITAL IDENTITY CRISIS?.....	1021
C.	PANOPTICISM AND NORMALIZED IDENTITY.....	1023
VII.	<b>CONCLUSIONS AND OUTLOOK</b> .....	1024
A.	SHIFTS IN POWER RELATIONS .....	1025
B.	CONSEQUENCES OF LEGAL PROTECTION.....	1027
C.	BEYOND CONTEXT-SPECIFIC INEQUALITY COMPENSATION .....	1029
D.	TWO DIRECTIONS TO EMPOWER PERSONS.....	1031
	1. <i>The Orthodox View: Resistance by Data Limitation and User Control</i> .....	1031
	2. <i>The Radical View: Resistance by Data Proliferation and Looking in Return</i> .....	1032
E.	CONCLUSION: NO MIDDLE WAY .....	1033
VIII.	<b>POSTSCRIPT: UMBERTO ECO’S ANOPTICON</b> .....	1034

## I. INTRODUCTION

Law and power are closely connected. In a lawless society, power will reign supreme, while in a society of rule of law, power is reined in by the law. However, law also establishes and validates power. The dual face of law—establishing and restraining power—is particularly relevant in unequal power relations, where the law both consolidates the power of strong parties and restricts their power by providing the weak parties with rights, in order to prevent them from being subjected to exercises of brute power. In this respect, inequality compensation is a key legal mechanism to regulate power relations. The law treats certain categories of people, including citizens, criminal suspects, employees, and consumers, as systematically weak parties relative to parties that are considered strong, such as the government, employers, and businesses. To balance these unequal power relationships weak parties are granted various rights in the domains of constitutional, administrative, labor, contract, and tort law. Examples include information rights, benefit-of-the-doubt and burden-of-proof rules, access to justice, and compensational rights.

This classic account of inequality compensation in legal domains is challenged by technology. With the advent of computers, the Internet, genetic profiling, and other information-related technologies, power relations

start to shift. Since “knowledge is power,” as the adage holds, both strong and weak parties use information-related technologies to improve their respective information positions. For example, consumers can now use the Internet to search for the lowest prices and can use ratings websites to inform one another of their experiences with particular products and services. Such practices free the consumer from the monopoly power of the shop around the corner. However, the gain in power that technology has provided to consumers is paralleled by the gains that technology provides to businesses. For instance, technology enables e-businesses to gather more information about consumers—using cookies, web forms, and profiling techniques—than the classic brick-and-mortar shop. They can then use this information to target consumers with increasing sophistication.

The gross outcome of such shifts in power relations is unclear: changes occur in different directions and sometimes along different dimensions. Parties may become stronger in one way, weaker in another, or both, depending on the circumstances. These shifts in power along different axes complicate the traditional *ex ante* model of inequality compensation which is based on the idea that certain parties are intrinsically stronger than others and must always be restrained by legal norms.

This Article aims, first, to explore the technology-related shifts in power relations that are occurring in the domains of law enforcement, labor, and commerce. Second, it aims to identify and examine the consequences of these shifts for the legal protection of weak parties, particularly for existing mechanisms of inequality compensation in the associated legal domains.

The domains of law enforcement, labor, and commerce exemplify unequal power relations and together cover a wide range of public and private law. Furthermore, technology is associated with significant shifts in the ways in which power is exercised today in these domains. Although the mechanism of inequality compensation will occur in most modern legal systems, the analysis is limited to the legal systems of the United States and the Netherlands and the concrete examples of legal protection that they supply.<sup>2</sup> The Article will explore these countries, with their different common law and civil law traditions, in a roughly comparative approach to discover differences and commonalities in the compensation granted to weak parties for inequalities in power relations.

---

2. Where Dutch law is based on European Union (E.U.) law, I will focus on the E.U. law. I will also occasionally mention developments in U.K. law where these are illustrative, particularly in the area of criminal law.

After some preliminaries in Part II that further introduce the notions of power relations, inequality compensation, and technology, the Article examines the relationship between law enforcement and citizens in Part III, employers and employees in Part IV, and businesses and consumers in Part V. The analysis is grounded in case studies to illustrate the effect of information-related technologies on shifting power relations. These case studies form the basis for a more general discussion of shifts in the power relation and their consequences for legal protection of weak parties. Then, Part VI provides an integrated view of citizens, employees, and consumers and the way in which the different roles of individuals are becoming intertwined in the information society. This forms the basis for drawing some conclusions in Part VII, not only about the distinct areas of law for specific categories of people, but also about the overall legal protection of individuals in the information society.

## II. PRELIMINARIES

### A. POWER RELATIONS

Power is complex and multifaceted. The term “power” comprises a wide array of notions bearing Wittgensteinian family resemblances.<sup>3</sup> It even comes close to being an “essentially contested concept,” that is, a concept “the proper use of which inevitably involves endless disputes about their proper uses on the part of their users.”<sup>4</sup>

For the purposes of this Article, the working definition of power is drawn from Dahl’s conceptualization of power relations: “A has power over B to the extent that he can get B to do something that B would not otherwise do.”<sup>5</sup> This definition shows exactly why weak parties are given legal

---

3. The many notions of power are connected by a series of overlapping similarities, but no one feature is common to all. For an introduction into the concept of “family resemblances,” see generally LUDWIG WITTGENSTEIN, *PHILOSOPHICAL INVESTIGATIONS* (P.M.S. Hacker & Joachim Schulte eds., G.E.M. Anscombe et al. trans., Blackwell Publishing Ltd. 2009) (1953).

4. W. B. Gallie, *Essentially Contested Concepts*, 56 *PROC. ARISTOTELIAN SOC’Y* 167, 169 (1956); see also Eugene Garver, *Rhetoric and Essentially Contested Arguments*, 11 *PHIL. & RHETORIC* 156 (1978) (connecting Gallie’s essentially contested concepts to Aristotle’s account of rhetorical argument). For overviews of the many notions of power, see generally MARK HAUGAARD, *POWER: A READER* (2002) and JOHN SCOTT, *POWER* (2001). It is not possible to discuss the concept itself in this Article. Fortunately, there is no need to; since this Article looks at power relations from the perspective of legal protection of weak parties against strong parties, it suffices to provide a working definition that fits in this context. The Article, after all, aims to reflect on legal protection of weak parties rather than on power relations per se.

5. Robert A. Dahl, *The Concept of Power*, 2 *BEHAV. SCI.* 201, 202–03 (1957); Robert A.

protection in power relations. From the perspective of autonomy—a key value underlying modern Western legal systems—B should be able to decide without undue restrictions what she wants to do, and not merely because A makes her do so.

This working definition can be enriched with some insights that refine Dahl's conceptualization. Bachrach and Baratz have called attention to a second dimension of power by pointing out that power can be exercised indirectly and passively by limiting the scope of decision-making to exclude issues of relevance to B, for example, by (non-)agenda setting.<sup>6</sup> Lukes has added a third dimension, namely, the bias in a system sustained “by the socially structured and culturally patterned behaviour of groups, and practices of institutions.”<sup>7</sup> Foucault has provided an important variation of Lukes' third dimension with his insights into the power mechanism of surveillance architecture. This is famously illustrated by the Panopticon, a mechanism where watched people (prisoners) aware of the continuous gaze of the watcher (the prison guard) internalize the value and knowledge system of the watcher, disciplining themselves according to the dominant discourse in society.<sup>8</sup>

This Article studies power relations in which A can get B to do something which B would not otherwise do, with a broad interpretation of “getting to do” that includes non-decision-making as well as cultural, institutional, and architectural mechanisms that have a disciplining effect on B.

#### B. LEGAL PROTECTION: INEQUALITY COMPENSATION

The legal phenomenon of inequality compensation that is embedded in the law is based on the idea that in society there are specific parties that have a structural, systematic advantage over other specific parties. This was a

---

Dahl, *Power*, in INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL SCIENCES (David L. Sills ed., 1968).

6. Peter Bachrach & Morton S. Baratz, *Two Faces of Power*, 56 AM. POL. SCI. REV. 947, 948 (1962).

7. See generally STEVEN LUKES, POWER: A RADICAL VIEW 26 (2005) (1974). Note, however, Haugaard's critique, HAUGAARD, *supra* note 4, at 38–40, that Lukes' stress on socially constituted bias makes it difficult to distinguish power from structural constraint. Haugaard suggests that an integrated theory of power and structure needs to be developed. *Id.*

8. For a critical discussion of panopticism, see generally David Lyon, THEORIZING SURVEILLANCE: THE PANOPTICON AND BEYOND (Devon Cullompton ed., 2006). See also MICHAEL FOUCAULT, POWER: ESSENTIAL WORKS OF FOUCAULT 1954–1984 (James D. Faubion ed., 2000); MICHEL FOUCAULT, SURVEILLER ET PUNIR: NAISSANCE DE LA PRISON (1975).

natural and valid assumption to make when the current systems were shaped during the nineteenth and most of the twentieth century. The relationship between law enforcement and citizens, between employers and employees, and between enterprises and consumers was by and large clear: the former could easily impose their will on the latter, unless something—such as legal norms—prevented or corrected them.

It is an important function of the law to compensate for such structural inequalities. Criminal, consumer, and labor law have developed to regulate structural imbalances by protecting the weak party against abuse of power by the strong party. The protection takes the form of inequality compensation which imposes duties on strong parties and grants rights to weak parties. Examples include rights related to information provision, notification duties, supervision mechanisms, and access to justice. The legal system views citizens, employees, and consumers as intrinsically disadvantaged parties that require structural inequality compensation. These legal-protection rules are triggered by the mere fact of belonging to the class of the weak party at issue, irrespective of the specific manifestation of the power relation in concrete circumstances.

### C. TECHNOLOGY

In a society as complex as the modern information and network society,<sup>9</sup> it may no longer be valid to assume that traditionally dominant parties remain more powerful than other parties. Partly through the influence of new technologies, most notably information and communication technology (ICT), but also genetic and surveillance technologies, unequal relationships seem to be shifting. This happens in subtle and often contradictory ways. Parties that were once considered weak by the very nature of the power relationship may emerge as the stronger party in certain circumstances. Alternatively, they can find themselves even weaker than they were before.

Technology plays a significant role in these shifts. For instance, increasingly sophisticated technology enables criminals to protect their communications from police surveillance and store incriminating electronic evidence in a data haven abroad, outside the reach of mutual legal assistance. However, technology also facilitates criminal investigation by supplying unprecedented surveillance tools, such as, microscopic sensors, smart cameras, and keyboard sniffers (i.e., software that secretly records keystrokes and sends these to the police). In the field of commerce, e-businesses can collect much more data about customers using technology such as cookies

---

9. *See generally* MANUEL CASTELLS, *THE RISE OF THE NETWORK SOCIETY* (Blackwell, 1996) (charting the social and economic relations of the global information economy).

and loyalty schemes. Consequently, these businesses are in a better position than ever to exploit their information advantage over the customer. At the same time, e-consumers can search the web for the lowest prices, participate in collective-buying activities, and set up grudge websites<sup>10</sup> to force a company to change its policy.

Admittedly, the role of technology in shifts in power relations is not always clear or easy to isolate from other factors. After all, power relations develop in social, economic, cultural, political, and architectural contexts. Technology is sometimes a sufficient cause for a certain development, sometimes a necessary cause, sometimes both, and at other times neither. Technological developments interact with other societal developments, in a process of mutual shaping where both developments influence each other.<sup>11</sup>

This Article does not aim to determine the causal influence of technology on power relations as such; rather, it limits its inquiry to describing shifts in power relations in which technology plays some role. This includes circumstances where technology opens up new possibilities for a strong party to exercise power, where it creates new opportunities for weak parties to resist the power of a strong party, or even where it blurs the very distinction between a strong and a weak party.

Because power relations hinge on knowledge and information<sup>12</sup> it will be important to examine information technologies. However, while the case studies in this Article primarily involve ICT, technologies relating to genetic information have also contributed to technology-related shifts in power relations, particularly through the advent of DNA forensics.

### III. LAW ENFORCEMENT—CITIZEN

This Part examines technology-related changes in the power relation of law enforcement and citizens, and assesses the consequences of these changes for the legal protection of citizens. The analysis begins with three case studies: DNA forensics, interception of telecommunications, and

---

10. *See, e.g.*, Wakeup Walmart.com: America's Campaign to Change Wal-Mart, <http://www.wakeupwalmart.com> (last visited Apr. 4, 2010).

11. For example, the introduction of mobile telephones has significantly changed the way people communicate. Specifically, mobile telephone users started using the SMS function of mobile telephones on a large scale and in ways completely unforeseen by its developers, thereby changing the technology. For a discussion of the mutual shaping of technology and society, see *SHAPING TECHNOLOGY/BUILDING SOCIETY: STUDIES IN SOCIOTECHNICAL CHANGE* (Wiebe E. Bijker & John Law eds., 1992).

12. *See* FOUCAULT, *POWER*, *supra* note 8, at 133 (arguing that power hinges on the political, economic, and institutional regime of the production of truth).

Passenger Name Records. Based on these case studies, a general discussion follows, outlining major developments and showing that the government, in its role as the protector of law and order, has embraced the enormous increase in technology-enabled tracing capacity. This development has not been offset by counter-developments of citizen empowerment. The resulting shift in power relation involves two types of problems—citizens being wrongly involved in a government investigation and a potentially disciplining effect of surveillance architectures—which seem to require new forms of legal protection.

#### A. CASE STUDY 1: DNA FORENSICS

Since the invention of DNA fingerprinting in the 1980s, DNA forensics have contributed to a gradual expansion in investigation powers.<sup>13</sup> Different types of DNA research have been developed and used in criminal investigation including DNA databasing, DNA phenotyping, mass screening, and familial searching.

The rise of DNA databasing is most visible in England and Wales, where the U.K. database, National DNA Database (NDNAD), has expanded enormously over the past decade. In 2007, it contained up to four million profiles (around six percent of the population), which were gathered through routine sampling and profile retention from arrestees as well as victims, consenting witnesses, and volunteers.<sup>14</sup> The U.S. national database, Combined DNA Index System (CODIS), was originally smaller in size, but outgrew the U.K. database in 2007, with 4.6 million profiles (around 1.5

---

13. See generally NUFFIELD COUNCIL ON BIOETHICS, THE FORENSIC USE OF BIOINFORMATION: ETHICAL ISSUES (2007) (examining the balance between police powers and individual rights to autonomy and privacy and offering recommendations to minimize misuses); MEREL M. PRINSEN, FORENSISCH DNA-ONDERZOEK: EEN BALANS TUSSEN OPSPORING EN FUNDAMENTELE RECHTEN (2008) (critically assessing Dutch and U.K. approaches to forensic DNA legislation); ROBIN WILLIAMS ET AL., GENETIC INFORMATION AND CRIME INVESTIGATION: SOCIAL, ETHICAL AND PUBLIC POLICY ASPECTS OF THE ESTABLISHMENT, EXPANSION AND POLICE USE OF THE NATIONAL DNA DATABASE (2004) (discussing the evolution in the use of genetic information in criminal investigations from case-by-case use to extensive and routine practice).

14. NUFFIELD COUNCIL ON BIOETHICS, *supra* note 13, at 9. Note that the current English and Welsh practice of retaining profiles and samples from unconvicted offenders should be changed in light of the European Court of Human Rights' judgment in *S. and Marper v. United Kingdom*, 2008 Eur. Ct. H.R. 1581; see HOME OFFICE, KEEPING THE RIGHT PEOPLE ON THE DNA DATABASE: SCIENCE AND PUBLIC PROTECTION (2009) (reporting on the government's consultation process launched in May 2009). The proposed "change" comprises retention of profiles from unconvicted people for six years for less serious crimes or twelve years for serious crimes, which does not seriously alter the policy of storing data from non-criminal citizens.

percent of the population).<sup>15</sup> In addition to CODIS, DNA databases exist in the United States at state and local levels, often containing more profiles. U.S. states are continually expanding their databases, allowing DNA samples to be taken from convicts and profiles to be stored for an increasing variety of crimes as well as for groups of citizens charged but not convicted.<sup>16</sup>

In the Netherlands, the power to take a DNA sample from a suspect was introduced in 1994, and then expanded in 2001 to allow for DNA collection in more types of crime and without a magistrate's warrant. In 2004, the DNA Convict Sampling Act allowed the Public Prosecutor to take samples from convicts in the interest of deterrence and to ensure more future matches with repeat offenders.<sup>17</sup> The Dutch database is smaller than the U.K. and U.S. databases; nevertheless, since the 2004 Act it has exploded, growing from 6,000 individual profiles in early 2005, to 45,000 in December 2007, to over 99,000 (around 0.6 percent of the population) in May 2010.<sup>18</sup>

As the use of DNA forensics has become more common, new qualitative methods of DNA analysis have also developed. For example, forensic DNA phenotyping, a relatively recent development, uses personal characteristics determined from crime scene DNA to trace unknown suspects.<sup>19</sup> This can help limit the pool of possible suspects so that law enforcement officials can conduct a mass-screening investigation. Alternatively, it can help exclude certain groups of people from further investigation. In England and Wales, the Forensic Science Service can determine the rough geographical ancestry of the DNA sample donor,<sup>20</sup> and at one time it offered a service to check for

---

15. NUFFIELD COUNCIL ON BIOETHICS, *supra* note 13, at 9.

16. See generally Aaron P. Stevens, *Arresting Crime: Expanding the Scope of DNA Databases in America*, 79 Tex. L. Rev. 921 (2001) (describing the history of DNA databases and their expansion to include more classes of criminals); Bonnie L. Taylor, *Storing DNA Samples of Non-Convicted Persons & the Debate over DNA Database Expansion*, 20 T.M. Cooley L. Rev. 509 (2003) (arguing that the national trend of expanding DNA databases to include more unconvicted individuals violates the Fourth Amendment and privacy rights).

17. Wet DNA-onderzoek in strafzaken, Staatsblad van het Koninkrijk der Nederlanden [Stb.] 596 (1993) (Neth.); Wet van 5 juli 2001 tot wijziging van de regeling van het DNA-onderzoek in strafzaken, Staatsblad van het Koninkrijk der Nederlanden [Stb.] 335 (2001) (Neth.); Wet DNA-onderzoek bij veroordeelden, Staatsblad van het Koninkrijk der Nederlanden [Stb.] 465 (2004) (Neth.).

18. DNA: Sporen Naar de Toekomst, <http://www.dnasporen.nl> (last visited July 1, 2010). See also the comparison of U.K. and Dutch developments in PRINSEN, *supra* note 13.

19. For technical and regulatory discussions see Bert-Jaap Koops & Maurice Schellekens, *Forensic DNA Phenotyping: Regulatory Issues*, 9 COLUM. SCI. & TECH. L. REV. 158 (2008); Pilar N. Ossorio, *About Face: Forensic Genetic Testing for Race and Visible Traits*, 34 J.L. MED. & ETHICS 277 (2006).

20. This is contested, first because individuals' DNA shows more variation than the variations of geographic groups of people, and second, because race is a social rather than a

red hair and light skin pigment.<sup>21</sup> Determining geographical ancestry or ethnic background is becoming more popular as scientific knowledge about DNA evolves.<sup>22</sup> As genetic knowledge advances, other phenotypical characteristics, such as hair, form, or height, may become available. In common law systems, such as those in the United States and the United Kingdom, use of a new technology is allowed until legislation or case law dictates otherwise.<sup>23</sup> In contrast, in civil law systems, new investigation techniques can usually only be used when legislation specifically allows it. The Netherlands, for example, enacted a law allowing phenotyping for geographic ancestry and gender. Other features, such as hair color, however, must be designated by an Order in Council (i.e., a lower-order regulation based on the statute) before the police can derive them.<sup>24</sup>

A third development is the use of DNA mass screening, or dragnet investigations, in which a group of people who match a suspect description are asked to voluntarily provide a DNA sample for profiling. This method was first used in the United States in 1990, when over 800 men in San Diego were tested in connection with a sextuple murder, and it has been used many times since.<sup>25</sup> Dragnet investigations have raised constitutional concerns where the volunteers' consent to DNA testing was given under police coercion. A paradigmatic example would be when a police officer gives an

---

genetic concept. See NUFFIELD COUNCIL ON BIOETHICS, *supra* note 13, at 80–81.

21. FORENSIC SCIENCE SERVICE, FACT SHEET: COMMONPLACE CHARACTERISTICS (2004), available at <http://www.forensic.gov.uk>. This service was discontinued due to insufficient demand from the police. For recent technological developments in deriving visible traits from crime-scene DNA, see Manfred Kayser & Peter M. Schneider, *DNA-Based Prediction of Human Externally Visible Characteristics in Forensics: Motivations, Scientific Challenges, and Ethical Considerations*, 3 FORENSIC SCI. INT'L: GENETICS 154 (2009); Fan Liu et al., *Eye Color and the Prediction of Complex Phenotypes from Genotypes*, 19 CURRENT BIOLOGY 192 (2009).

22. See Mark D. Shriver & Rick A. Kittles, *Genetic Ancestry and the Search for Personalized Genetic Histories*, 5 NATURE REV. GENETICS 611 (2004).

23. See Michelle Hibbert, *DNA Databanks: Law Enforcement's Greatest Surveillance Tool?*, 34 WAKE FOREST L. REV. 767, 791–92 (1999); Koops & Schellekens, *supra* note 19, at 27–32. In the United States, only Indiana, Rhode Island, and Wyoming have outlawed forensic phenotyping. See IND. CODE ANN. § 10-13-6-16 (West 2004); R.I. GEN. LAWS § 12-1.5-10 (2007); WYO. STAT. ANN. § 7-19-404 (2007).

24. Wet van 8 mei 2003 tot wijziging van de regeling van het DNA-onderzoek in strafzaken in verband met het vaststellen van uiterlijk waarneembare persoonskenmerken uit celmateriaal [Act of May 8, 2003], Staatsblad van het Koninkrijk der Nederlanden [Stb.] 201 (2003) (Neth.).

25. Philip P. Pan, *Pr. George's Chief Has Used Serial Testing Before; Farrell Oversaw DNA Sampling of 2,300 in Fla.*, WASH. POST, Jan. 31, 1998, at B1 (“One of the first agencies to experiment with the [DNA dragnet] was the San Diego police department, which tested about 800 men during its search for a serial killer who stabbed six women to death in their homes between January and September 1990.”).

individual the “choice” between providing a sample voluntarily or being subjected to a court order for DNA testing and enduring the publicity that such an order would generate.<sup>26</sup>

The Netherlands has also used DNA mass screenings in a score of cases, starting with the 1999 testing of 115 men in the still unsolved case of a serial rapist in Utrecht.<sup>27</sup> Between 1999 and 2004, approximately 4,600 people were asked to volunteer a DNA sample in fourteen cases.<sup>28</sup> The Dutch practice has to conform to policy criteria—provided by the Minister of Justice—that favor restricted use. These are generally reserved for the most serious crimes that cause significant social unrest,<sup>29</sup> and must be made with authorization from the Board of Procurators-General, the highest body within the Public Prosecutor. In 2007, the policy was expanded. Mass screening is no longer a tool of last resort, but rather part of the reasonable effort extolled by law enforcement during investigations.<sup>30</sup>

The fourth, and final recent innovation in DNA forensics is familial searching. This involves searching a database for partial matches of DNA profiles that suggest that the unknown person who left the stain at the crime scene is closely related to a known person whose DNA is stored in the DNA database. Familial searching was first used in England in 2002 in solving a 1973 double homicide case.<sup>31</sup> The crime scene stains had been profiled with Low Copy Number analysis, a new technique so sensitive that it can yield a DNA profile from only a few body cells, and yielded a partial match with a profile in the database.<sup>32</sup> Ultimately the father of the partial match, then deceased, was identified as the perpetrator.<sup>33</sup>

---

26. See Sepideh Esmaili, *Searching for a Needle in a Haystack: The Constitutionality of Police DNA Dragnets*, 82 CHI. KENT L. REV. 495 (2007).

27. CHRISTIANNE J. DE POOT & EDWIN W. KRUISBERGEN, KRINGEN ROND DE DADER GROOTSCHALIG DNA-ONDERZOEK ALS INSTRUMENT IN DE OPSPORING [CIRCLES AROUND THE PERPETRATOR: LARGE-SCALE DNA-ANALYSIS AS A TOOL IN CRIMINAL INVESTIGATION] 33 (2006).

28. *Id.* at 203.

29. Kamerstukken II, 2000-2001, 27 400 VI, No. 49 (Neth.).

30. Kamerstukken II, 2007-2008, 34 415, No. 1 (Neth.).

31. Robin Williams & Paul Johnson, *Inclusiveness, Effectiveness and Intrusiveness: Issues in the Developing Uses of DNA Profiling in Support of Criminal Investigations*, 33 J.L. MED. ETHICS 545, 554 (2005); Robin McKie, *Did a Killer Evade Justice Due to Withheld Evidence? The Collapse of the Case Against Angus Sinclair was a Bitter Blow to a Scientist Whose DNA work was not Fully Presented in Court*, THE OBSERVER, Sept. 16, 2007, at 17.

32. Williams & Johnson, *supra* note 31, at 554.

33. *Id.* In the Netherlands, familial searching requires a statutory basis which does not yet exist, but which has been proposed by the government. Kamerstukken II, 2007-08, 34 415, No. 1 (Neth.); see also Merel M. Prinsen, *DNA-verwantschapsonderzoek. Familie van de verdachte?*, 6 STRAFBLAD 242 (2008) (discussing pros and cons of familial searching in the

Familial searching provides an interesting expansion of policing, since it “effectively increases police scrutiny and interest in people based on their relatives’ past involvement with the criminal justice system.”<sup>34</sup> This practice raises three main concerns. First, it could have “differential effects on groups in American society.”<sup>35</sup> Second, it raises questions about whether the consent given by volunteers in a mass screening is truly informed. That is, are volunteers sufficiently informed that permitting their DNA to be included in a forensic database can also affect their relatives? Finally, familial searching unearths ethical concerns in situations where individuals are not aware that their social family is not their biological family, for example, when the assumed father turns out not to be the biological father.<sup>36</sup>

#### B. CASE STUDY 2: INTERCEPTION OF TELECOMMUNICATIONS

Over the past decades, interception of communications has expanded greatly in the United States and, particularly, in the Netherlands. In the United States, interception of phone (wire) communications was regulated<sup>37</sup> after *Katz* interpreted the Fourth Amendment to protect telephone communications.<sup>38</sup> In 1986, the Electronic Communications and Privacy Act (ECPA) enabled the interception of electronic communications under less strict conditions than those that govern wire interception.<sup>39</sup> ECPA also allowed wiretapping for more types of crimes and introduced “roving” interception, which involves following the targeted suspect rather than focusing on fixed phone lines or places.<sup>40</sup> The USA Patriot Act of 2001 allowed interception for even more crimes, and it transferred voice mail from

---

Dutch context).

34. Henry T. Greely et al., *Family Ties: The Use of DNA Offender Databases To Catch Offenders’ Kin*, 34 J.L. MED. & ETHICS 248, 255 (2006).

35. *Id.*

36. See Williams & Johnson, *supra* note 31, at 554–56 (assessing the effects of the recent innovative use of DNA databasing for “familial searching” and the way it has unsettled agreed understandings about appropriate uses of DNA). See generally Erica Haimes, *Social and Ethical Issues in the Use of Familial Searching in Forensic Investigations: Insights from Family and Kinship Studies*, 34 J.L. MED. & ETHICS 263 (2006) (exploring the socio-ethical concerns raised by familial searching of forensic databases in criminal investigations).

37. Omnibus Crime Control and Safe Streets Act, 18 U.S.C. § 2510 (2006).

38. See *Katz v. United States*, 389 U.S. 347, 353 (1967) (“The Government’s activities in electronically listening to and recording the petitioner’s words violated the privacy upon which he justifiably relied while using the telephone booth and thus constituted a ‘search and seizure’ within the meaning of the Fourth Amendment.”).

39. Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848 (codified as amended at 18 U.S.C. § 2510 (2006)).

40. *Id.*

the wiretap regime to the less protected communications storage regime.<sup>41</sup> The Netherlands experienced similar expansion.<sup>42</sup> Particularly significant was the Special Investigation Powers Act of 2000 that broadened the scope of police powers to allow interception of the connection not only of suspects, but also of non-suspects, provided that such interception could benefit the investigation.<sup>43</sup>

While the law in the books broadened over time, an equally important expansion of interception powers occurred in practice. In the United States, the number of interception authorizations (for criminal investigation, not for intelligence or national security) tripled since 1987 (from 673 authorizations in 1987 to 1,891 in 2008), and the average number of intercepted communications in each case doubled (from 1,299 communications in 1987 to 2,707 in 2008), so that the total amount of communications intercepted sextupled.<sup>44</sup> In the Netherlands, the available figures are much higher: in 1993, 3,619 interception authorizations were granted for criminal investigation (more, in absolute terms, than in the United States),<sup>45</sup> rising to 10,000 in 1999, and 26,425 authorizations in 2008.<sup>46</sup> This does not mean that over a decade, ten times more people have been under wiretap. Authorizations are given for separate connections, such as fixed and mobile phones, and criminals today have substantially more phones; nevertheless, the trend is undeniably upwards. Furthermore, given the enormous increase

---

41. Uniting and Strengthening America by Providing Appropriate Tools Required to Interrupt and Obstruct Terrorism (USA PATRIOT ACT) Act of 2001, Pub. L. No. 107-56, 115 Stat. 272 (2001).

42. *See generally* ARNO HUBERTUS HENRICUS SMITS, STRAFVORDERLIJK ONDERZOEK VAN TELECOMMUNICATIE (2006) (surveying the historical development of Dutch law in the area of criminal investigation of telecommunications).

43. *See* WETBOEK VAN STRAFVORDERING [CRIM. PROC. CODE] art. 126n, 126u, introduced by the Wet bijzondere opsporingsbevoegdheden [Special Investigatory Powers Act], Staatsblad van het Koninkrijk der Nederlanden [Stb.] (1999) 245 (Neth.).

44. ADMIN. OFFICE OF THE U.S. COURTS, 2008 WIRETAP REPORT 7 (2008), *available at* <http://www.uscourts.gov/wiretap08/contents.html> [hereinafter 2008 WIRETAP REPORT]; ADMIN. OFFICE OF THE U.S. COURTS, 1997 WIRETAP REPORT 30 (1999), *available at* <http://www.uscourts.gov/wiretap/contents.html>. Note, however, a slight decrease in recent years: the 2007 numbers were higher, with 2,208 authorizations and 3,106 intercepts on average. 2008 WIRETAP REPORT, *supra*, at 32.

45. Wiretapping for intelligence versus criminal investigation purposes may yield a quite different picture, although it is difficult to compare these due to the official secrecy associated with intelligence practice.

46. The 1993 figure is from Z. REIJNE, TAPPEN IN NEDERLAND (1996). The 2000 figure is mentioned in Grootchalig af luisteren van moderne telecommunicatiesystemen, Kamerstukken II, 2000-2001, 27 591, No. 2. The 2008 figure is given in Kamerstukken II, 2009-2010, 30 517, No. 13 (Neth.)—the first annual systematic Dutch wiretap figures to be officially published.

in communications generally, particularly since the advent of mobile phones with short message service (SMS) capability and the Internet, much more data have become available to the police through intercepts. This is not only a quantitative increase, but also a qualitative increase. New types of data, such as internet browsing and location data, now allow long distance glimpses of human life that were previously hidden to the police, or observable only with significant effort and costs.

Another development has occurred in interception: the introduction of mandatory interceptability. Until the early 1990s telephone communications were easily interceptable. Then, because of an increase in market parties and diversification of telecom technologies, the police began to encounter difficulty intercepting. Concerned that a major investigation tool might be lost, governments passed laws forcing telecommunication providers to build in interceptability. The U.S. Communications Assistance for Law Enforcement Act (CALEA) of 1994<sup>47</sup> and Chapter 13 of the Dutch Telecommunications Act of 1998 imposed obligations on telecommunications carriers to ensure interceptability.<sup>48</sup> A statement by Dutch Member of Parliament highlights the reasoning behind these laws: “The traditional form of telephony can be intercepted. An alternative must be the same. We find that being interceptable is an *inseparable* part of the phenomenon of telephony in our country.”<sup>49</sup>

---

47. Communications Assistance for Law Enforcement Act, 47 U.S.C. § 1008(b)(1) (2006); *see also* AskCALEA: Communications Assistance for Law Enforcement Act, <http://www.askcalea.net> (last visited Jan. 31, 2010) (acting as a resource and information clearinghouse for individuals and organizations with an interest in the Communications Assistance for Legal Enforcement Act of 1994).

48. Telecommunicatiewet [Telecommunications Act], Staatsblad van het Koninkrijk der Nederlanden [Stb.] 610 (1998) (Neth.); *see also* Council Resolution 96/C329/01, 1996 O.J. (C 329) 1 (EC) (addressing the lawful interception of telecommunications). For a comparison of the U.S. and Dutch legislation, which shows a more liberal approach to impositions on market parties in the United States, see Bert-Jaap Koops & Rudi Bekkers, *Interceptability of Telecommunications: Is US and Dutch Law Prepared for the Future?*, 31 TELECOMM. POL'Y 45 (2007). This analysis shows that the U.S. approach, as laid down in CALEA, is essentially more flexible and balanced than the Dutch approach. CALEA already presumes some form of trade-off, through the crucial provision of 47 USC § 1008(b)(1), which lists ten factors to be taken into account in determining the reasonableness of requiring a particular telecom provider to build in interceptability, is flanked by several other checks and balances that ensure enhanced cost-effectiveness. The authors conclude that the rigid Dutch law cannot handle such a trade-off since it requires tout court that new telecommunications networks and services are made interceptable and that the providers fund these measures, regardless of the costs or the effects on security, privacy, or innovation.

49. Handelingen II, Oct. 25, 1995, 17-1123 (Neth.) (translation by Bert-Jaap Koops, emphasis added). In a related development, when technologists discovered that cell telephones were capable of “knowing” their location, governments started to mandate that

C. CASE STUDY 3: PASSENGER NAME RECORDS<sup>50</sup>

Apart from developments in regular criminal investigation discussed in the previous case studies, anti-terrorism developments in the periphery of criminal law also merit attention. In this area, the mandatory exchange between countries of Passenger Name Records (PNR) of air travelers constitutes an interesting case study. PNR include information such as a passenger's name and address, birth date, passport details, payment data, emergency contacts, and meal and seating preferences. After the 9/11 terrorist attacks on New York and Washington, D.C., in 2001, the United States believed that processing PNR might help to keep terrorists out of the country. The Bureau of Customs and Border Protection (CBP) started asking airlines to provide the government with access to their PNR records. For European airlines, this processing of personal data for purposes other than the original purposes for which the data were collected violated data protection legislation if conducted without a legal ground. Consequently, the European Union (E.U.) made an agreement with the United States to authorize the provision of PNR.<sup>51</sup> The agreement did not authorize the *exchange* of PNR data, but only the one-way access of U.S. government agencies to European data.

The PNR agreement, however, was controversial. The European Parliament felt that it had been outmaneuvered as protector of the privacy of European citizens and challenged the underlying documents<sup>52</sup> before the European Court of Justice. The Court struck down the Commission's and Council's decisions, finding that they were based on the wrong legal ground, and thereby effectively annulled the PNR agreement.<sup>53</sup> However, the

---

mobile phones be made with the ability to make their location known in case an emergency number was called. *See generally* David J. Phillips, *Privacy and Data Protection in the Workplace: The US Case*, in REASONABLE EXPECTATIONS OF PRIVACY? 39 (Sjaak Nouwt et al. eds., 2005) (concluding that although the impetus for these mandates was not crime-control but safety concerns, as a result locatability has become an inherent feature of mobile phones, and as a consequence, generated location data will routinely be available for criminal investigation purposes).

50. This Section builds on Vagelis Papakonstantinou & Paul De Hert, *The PNR Agreement and Transatlantic Anti-Terrorism Co-Operation: No Firm Human Rights Framework on Either Side of the Atlantic*, 46 COMMON MKT. L. REV. 885 (2009).

51. On the European side, the agreement was backed up by two official documents: Commission Decision 2004/535, 2004 O.J. (L 235) 11 (EC), and Council Decision 2004/496, 2004 O.J. (L 183) 83 (EC).

52. *Id.*

53. Joined Cases C-317/04 & 318/04, *Parliament v. Council*, 2006 E.C.R. I-04721 (holding that the Agreement had been passed as a measure in the area of justice and home affairs (where the European Commission and European Council take decisions), while it should have been passed as a measure related to economic, social, and environmental

European Parliament's action backfired, by triggering a renegotiation with the United States on a second PNR agreement, in which the United States emerged even stronger.

The second PNR agreement was concluded on June 29, 2007 and memorialized in three documents.<sup>54</sup> It comprises fewer passenger records than the first agreement—nineteen instead of thirty four—but since the records contain the same types of information, only in a different format, this was only a cosmetic reduction. The records can be retained significantly longer than before—fifteen instead of three and a half years—and it is not guaranteed that the records will be destroyed after this period.<sup>55</sup> Moreover, a wider range of U.S. agencies, not only customs, can access the data, and the data may be transferred to other countries at the discretion of the U.S. Department of Homeland Security (DHS).<sup>56</sup>

Perhaps most importantly for our purposes, although the PNR Agreements were initiated in the post 9/11 wave of anti-terrorism measures, the PNR data may be used by the U.S. government for combating or preventing not only terrorism, but also

other serious crimes, including organized crime, that are transnational in nature. PNR may be used where necessary for the protection of the vital interests of the data subject or other persons, or in any criminal judicial proceedings, or as otherwise required by law. DHS will advise the EU regarding the passage of any U.S. legislation which materially affects the statements made in this letter.<sup>57</sup>

In other words, personal data of European citizens collected to facilitate air travel accommodations are now mandatorily provided to the U.S. government in the interest of counter-terrorism, and can simultaneously be

---

policies (where the European Parliament has co-decision power alongside the Commission and Council)).

54. The three documents are (1) an agreement signed by both parties, (2) a U.S. letter to the E.U. assuring how it will handle European PNR data in the future, and (3) a letter from the E.U. to the United States acknowledging receipt of this letter. *See* Council Decision 2007/551/CFSP/JHA, 2007 O.J. (L 204) 16 (EU). The relationship between the three documents makes the agreement uncertain, thereby complicating the assessment of its exact legal status and contents. *See* Papakonstantinou & De Hert, *supra* note 50, at 908–19.

55. Papakonstantinou & De Hert, *supra* note 50, at 912 (citing Ch.VII, U.S. Letter to the European Union in Council Decision 2007/551, 2007 O.J. (L 204) 16–25 (EC)) (“We expect that EU PNR data shall be deleted at the end of this period; questions of whether and when to destroy PNR data collected in accordance with this letter will be addressed by DHS and the EU as part of future discussions.”).

56. *Id.* at 911.

57. *Id.*

used for combating serious crime, protecting vital interests, or any other purposes currently or later stipulated by U.S. law. “Function creep” seems a bland description of this deviation from the principle of purpose specification and use limitation that is ingrained in European data protection law.<sup>58</sup> The rapid expansion of PNR functionality seems better captured by the term “function rush.”

The developments in PNR are not solely products of U.S. political pressure. Several E.U. countries have also started to require access to PNR and store these data for anti-terrorism or other purposes, and a Framework Decision is being proposed to introduce PNR processing throughout the E.U.<sup>59</sup> Interestingly, air carriers already must communicate Advance Passenger Information (API) to authorities of E.U. Member States for fighting illegal immigration;<sup>60</sup> “[t]he added value of PNR is that it helps identify unknown people and develop risk indicators.”<sup>61</sup> Some member states, including the United Kingdom, would like to see the purpose of PNR processing extended from fighting terrorism and organized crime to other purposes as well.<sup>62</sup>

#### D. DISCUSSION

Broad use of DNA forensics and interception of communications is representative of a wide range of advances in criminal investigation using new technologies or new applications of existing technologies. The means for searching computers, collecting traffic data, ordering the production of computer data, employing camera and olfactory surveillance,<sup>63</sup> and utilizing forensic chemistry have developed significantly over the past two decades.<sup>64</sup>

58. See Directive 95/46/EC, art. 6(1)(b), 1995 O.J. (L 281) 31 (EC).

59. Press Release, European Commission, Proposal for a COUNCIL FRAMEWORK DECISION on the Use of Passenger Name Record (PNR) for Law Enforcement Purposes, Memo/07/449 (Nov. 6, 2007), available at <http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/07/449&format=HTML&aged=0&language=EN&guiLanguage=en>.

60. Directive 2004/82/EC, 2004 O.J. (L 261) 24 (EU).

61. Press Release, European Commission, *supra* note 59.

62. See SEC’y OF STATE FOR THE HOME DEP’t, THE PASSENGER NAME RECORD (PNR) FRAMEWORK DECISION: THE GOVERNMENT REPLY TO THE FIFTEENTH REPORT FROM THE HOUSE OF LORDS EUROPEAN UNION COMMITTEE SESSION 2007-08 HL PAPER 106, 1–2 (2008) (recommending that PNR be extended to “serious crimes”).

63. Olfactory surveillance is conducted by detecting scents, for example, with sniffer dogs or wasps to detect drugs or chemicals. See Amber Marks, *Drug Detection Dogs and the Growth of Olfactory Surveillance: Beyond the Rule of Law?*, 4 SURVEILLANCE & SOC’y 257 (2007) (discussing the expansion of olfactory surveillance in the United Kingdom through increased use of drug detection dogs and arguing that it sets a dangerous precedent for the regulation of other surveillance technologies).

64. See KRISTIE BALL ET AL., A REPORT ON THE SURVEILLANCE SOCIETY: FOR THE

The PNR case study, moreover, is emblematic of government anti-terrorism measures and efforts in the fringes of crime-fighting. This includes a host of administrative or pseudo-criminal measures—incorporating biometrics in travel documents, increased identification duties, preventative frisking, and scanning of laptops at customs—taken to scan and store data about groups of people to prevent potentially dangerous activities.

Evident in these case studies is a consistent pattern of increasing traces. Citizens leave digital traces when exploring the Internet, using automatic teller machines (ATMs) and point-of-sale terminals, entering secured buildings, walking the streets under the watchful eyes of closed circuit television (CCTV); they leave physical traces when walking around, touching objects, smoking cigarettes, combing their hair, or drinking beer—all of which leaves behind enough body cells to enable DNA profiling.<sup>65</sup> Not all of these traces are new. Fingerprints, for example, have long been available for government scrutiny. But many traces have only come into existence through the advent of ICT, while others can only be considered as traces because technological developments have enabled their identification as such. The increase in traces is enormous when we compare the digital and physical footprint of today's citizens with the footprint of citizens two decades ago, both in quantity and in quality.

Moreover, network technologies and digitization have enabled connecting these traces in many ways, effectively making citizens into digital persons<sup>66</sup> living their lives in databases.<sup>67</sup> The interconnection of traces can

---

INFORMATION COMMISSIONER BY THE SURVEILLANCE STUDIES NETWORK (David M. Wood ed., 2006), available at [http://www.ico.gov.uk/upload/documents/library/data\\_protection/practical\\_application/surveillance\\_society\\_full\\_report\\_2006.pdf](http://www.ico.gov.uk/upload/documents/library/data_protection/practical_application/surveillance_society_full_report_2006.pdf); Ben Bowling et al., *Crime Control Technologies: Towards an Analytical Framework and Research Agenda*, in REGULATING TECHNOLOGIES 51 (Roger Brownsword & Karen Yeung eds., 2008); JAMES C. FRASER & ROBIN WILLIAMS, HANDBOOK OF FORENSIC SCIENCE (2009); Bert-Jaap Koops, *Technology and the Crime Society: Rethinking Legal Protection*, 1 L. INNOVATION & TECH. 93 (2009).

65. Current DNA profiling requires only some dozens of picograms (i.e.,  $10^{-12}$  or a millionth of a millionth of a gram) of body material, the equivalent of four or five body cells, to make a DNA profile, provided the material is not contaminated. DNA profiles can therefore be made from material collected from single strands of hair, toothbrushes, cigarette butts, or smudges on a glass. See Peter Gill & Tim Clayton, *The current status of DNA profiling in the UK*, in HANDBOOK OF FORENSIC SCIENCE 29, 49 (Jim Fraser & Robin Williams eds., 2009).

66. See generally DANIEL J. SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE (2004) (arguing that the existing privacy regulatory regime is uneven, overly complex, and ineffective at addressing the expansion in data compilation and retention on individuals).

67. See generally SIMSON GARFINKEL, DATABASE NATION: THE DEATH OF PRIVACY IN

also yield surprising results hitherto unimaginable, as illustrated by the development of DNA familial searching and digital profiling,<sup>68</sup> in which information about citizens is created using only data from other persons.

Within the power relations of law enforcement and citizens, the enormous increase in tracing capacity enabled by technology has been liberally embraced by the government in its role as law and order protector. Surfing the wave of the post-9/11 climate of fear,<sup>69</sup> as well as the more general and somewhat older wave of the risk society (i.e., a society that frames problems in terms of risks and that deals with hazards through systematic risk assessment and risk management),<sup>70</sup> the U.S. government has significantly broadened its surveillance powers over the past two decades. This has enabled the United States to surveil all citizens, in the dual sense of *sur-veiller* (i.e., “watching over”): care and control.<sup>71</sup> It has eagerly accepted the possibilities of the ever-increasing availability of personal data stored in existing databases, which are accessible to police and intelligence agencies through liberal data-ordering powers. It has also started to mandate the storage of personal data that would otherwise be deleted.<sup>72</sup> Furthermore, it has created extensive databases itself, such as DNA and PNR databases, which store data not about suspects of concrete crimes, but of varying collections of citizens who are, in varying degrees, seen as potential perpetrators of crime or terrorism. Technology is thus facilitating what is effectively a paradigm shift in the government’s role in combating crime

---

THE 21ST CENTURY (2000) (discussing how advances in technologies endanger personal privacy).

68. See generally Mireille Hildebrandt, *Profiling and the Identity of the European Citizen*, in *PROFILING THE EUROPEAN CITIZEN: CROSS-DISCIPLINARY PERSPECTIVES* 303 (Mireille Hildebrandt & Serge Gutwirth eds., 2008) (describing how the proliferation of automatically generated profiles in an increasingly networked society can affect the lives of ordinary citizens).

69. Cf. JONATHAN SIMON, *GOVERNING THROUGH CRIME: HOW THE WAR ON CRIME TRANSFORMED AMERICAN DEMOCRACY AND CREATED A CULTURE OF FEAR* (2007) (arguing that governing through crime fuels a culture of fear and control that in turn lowers the threshold of fear); CASS R. SUNSTEIN, *LAWS OF FEAR: BEYOND THE PRECAUTIONARY PRINCIPLE* (2005) (discussing problems in individual and social judgments that can make people more fearful than is warranted).

70. See generally ULRICH BECK, *RISK SOCIETY: TOWARDS A NEW MODERNITY* (1992) (arguing that in a risk society, the “logic” of risk production outweighs the “logic” of wealth production).

71. DAVID LYON, *SURVEILLANCE SOCIETY: MONITORING EVERYDAY LIFE* 3 (2001).

72. Most notable is the mandatory retention of telecommunications traffic data in Europe. See Directive 2006/24/EC, 2006 O.J. (L 105) 54 (EU).

from an ex post, incidental, and last resort type of criminal law to an ex ante, comprehensive, and first resort type of criminal law.<sup>73</sup>

Now, how exactly does this affect the power relation of law enforcement and citizen? Law enforcement has acquired and is exercising considerably more power over ordinary citizens. The most poignant way the government can get a citizen to do what he would not otherwise do—by incarcerating him—has gained considerable momentum in the climate of “penal harshness” that has accompanied the risk society, particularly in the United States, but also in the United Kingdom and the Netherlands, a country that was once renowned for its mild and humane penal approach.<sup>74</sup> The expanded footprint of substantive law, constituted by the rise of regulatory crimes<sup>75</sup> and the criminalization of banal offenses or antisocial behavior,<sup>76</sup> implies that the punishing power of government is now exercised against wider circles of citizens.

However, what is more important for our analysis is that power is being exercised in new ways, beyond simply imprisoning or fining people. This is the architectural component of the surveillance society. Society’s information processes are being structured in such a way as to enable continuous scrutiny of citizens for early warnings of abnormal and potentially dangerous behavior. As soon as the system gives off warning signals, restraint is exercised in ways more subtle than mere physical incapacitation, for example, by tracking rather than confining potentially dangerous subjects.<sup>77</sup>

---

73. Koops, *supra* note 64, at 117 (arguing that criminal law is shifting from a last resort to a primary tool of social control).

74. *See generally* DAVID GARLAND, *THE CULTURE OF CONTROL: CRIME AND SOCIAL ORDER IN CONTEMPORARY SOCIETY* (2001) (arguing that changes in criminal justice in the United States and the United Kingdom in the last twenty-five years are attributable to the social organization of modernity and the neoconservative politics that dominated in the 1980s); NICOLA LACEY, *THE PRISONERS’ DILEMMA: POLITICAL ECONOMY AND PUNISHMENT IN CONTEMPORARY DEMOCRACIES* (2008) (discussing how British criminal justice policy has become increasingly politicized); Michael Tonry & Catrien Bijleveld, *Crime, Criminal Justice, and Criminology in the Netherlands*, in *CRIME AND JUSTICE IN THE NETHERLANDS 1* (Michael Tonry & Catrien Bijleveld eds., 2007) (surveying the Dutch criminal justice system).

75. *See generally* Andrew Ashworth, *Is the Criminal Law a Lost Cause?*, 116 L. Q. REV. 225 (2000) (critically examining the expansion of criminal offenses to gain political favor); Robert Baldwin, *The New Punitive Regulation*, 67 MOD. L. REV. 351 (2004) (discussing evidence of a drift towards punitive approaches to regulation and more frequent imposition of criminal sanctions).

76. *See generally* Stuart Macdonald, *A Suicidal Woman, Roaming Pigs and a Noisy Trampolinist: Refining the ASBO’s Definition of ‘Anti-Social Behaviour,’* 69 MOD. L. REV. 183 (2006) (discussing the definition of antisocial behavior employed by the Crime and Disorder Act of 1998 for the purposes of the Anti-Social Behaviour Order).

77. *See* Erin Murphy, *Paradigms of Restraint*, 57 DUKE L.J. 1321 (2008) (arguing that legal

Foucault's recollection of Bentham's Panopticon as a paradigmatic way of disciplining people truly seems visionary: digital citizens in today's database nation cannot help but be aware of the watchful eye of the government's guards. It is not the fact of *being* watched, but the fact that at any moment they *can* be watched, that has a potentially disciplining effect on citizens. Perhaps it is this as much as any other factor that triggers the ubiquitous "I have nothing to hide" response<sup>78</sup>: a psychological mechanism of citizens to rationalize and therewith get to grips with the government's panoptic gaze. The implicit implication of "I have nothing to hide" is that "I don't mind being watched because I'm doing nothing wrong," and this precisely constitutes the normalizing, disciplining effect that Foucault's analysis of power elucidates. Through the panoptic power of surveillance architecture, citizens embrace society's prevalent paradigm of normality. This is not in itself good or bad, but it is an exercise of power in the relationship between government and citizen that must not be overlooked.

The increase in government power through the enlarged footprint of criminal law and the establishment of surveillance architectures is not offset by counter-developments that empower citizens. It is obvious that technology also opens up new paths for citizens, but these lie in the sphere of participatory democracy and electronic service delivery, and they do not generally affect the power relation of citizens with law enforcement. Technology does offer some options to citizens for shielding information, potentially more securely than is possible physically (e.g., with strong cryptography). However, on balance, technology facilitates the investigative ability of law enforcement and intelligence agencies much more than it enhances citizens' ability to evade authorities.<sup>79</sup> Simple privacy enhancing technologies (PETs) like drawing the curtains or whispering used to be quite effective against peeping Toms or eavesdroppers, but they are insufficient against modern home and body monitoring devices.<sup>80</sup> Furthermore, PETs for digital security are usually more complex and difficult to use than physical security devices. To be sure, some groups—organized and calculating criminals and terrorists—do benefit from technologies that allow them to

---

scrutiny of targeted forms of non-physical control has been overlooked).

78. Cf. Daniel J. Solove, *"I've Got Nothing to Hide" and Other Misunderstandings of Privacy*, 44 SAN DIEGO L. REV. 745 (2007) (critically examining the argument that no privacy problem exists if a person has nothing to hide).

79. Koops, *supra* note 64, at 101.

80. See generally Bert-Jaap Koops & Merel M. Prinsen, *Houses of Glass, Transparent Bodies: How New Technologies Affect Inviolability of the Home and Bodily Integrity in the Dutch Constitution*, 16 INFO. & COMM. TECH. L. 177, 180 (2007) (discussing how technological developments in ICT and DNA research pose a threat to home and bodily integrity).

hide from government scrutiny, and it is these groups that the new surveillance measures aim to combat. But compared to the traditional application of criminal law, the new measures are much less targeted and narrowly tailored. Thus, they are more likely to affect all citizens rather than small groups of suspects or would-be terrorists. In the arms race between governments and organized crime and terrorist groups, however legitimate and necessary it might be, ordinary citizens suffer massive collateral damage.

This collateral damage has two faces, each of which seems to require new forms of legal protection if a reasonable balance of power is to be maintained between government and citizens. First, citizens risk being wrongly involved in a government investigation. Some errors will always happen, because of human or technical imperfections, or due to the fact that profiling always involves some false positives, i.e., people who happen to fit a certain profile when in fact they do not belong to the category of people the profile aims at identifying. Errors may also occur because of criminal identity theft, which is a serious problem in both the United States and the Netherlands.<sup>81</sup> The risk of errors in crime fighting is not new, but the magnitude of the risk has grown with the rise of penal harshness and the expansion of surveillance databases. More importantly, it also involves other types of risk: the potential harm for citizens is not so much incarceration or even severe physical or emotional damage to home, body, or close relationships, rather it involves vague, invisible, and long-term forms of harm resulting from “data shadows” lingering in public and private databases. Perhaps the core vulnerability is no longer sending an innocent person to jail, but labeling the digital persona of an innocent citizen with a stamp that significantly lowers the quality of her future social life. Besides safeguards for proportional investigation and a fair

---

81. For the United States, see Michael W. Perl, *It's Not Always About the Money: Why the State Identity Theft Laws Fail to Adequately Address Criminal Record Identity Theft*, 94 J. CRIM. L. & CRIMINOLOGY 169 (2003) (discussing the inadequacy of state identity theft laws to protect against criminal record identity theft in which an identity thief obtains a victim's personal information then commits crimes while acting as the victim). For the Netherlands, note the case of Mr. K, who was registered in government databases for over thirteen years as a serious drug criminal as the result of identity theft by a drug addict. Mr. K suffers significant obstacles in daily life; he is frequently held up at Schiphol Airport, receives numerous tickets for dodging transport fares, has had difficulty obtaining a mortgage, and has been subjected to a search in his home by thirty-five armed investigation officers, which induced him to move because all of his neighbors shunned him. The National Ombudsman castigated the government for their consistent failure to remove the man's registration data from its databases. See DUTCH NATIONAL OMBUDSMAN REPORT 2008/232 (2008), available at <http://www.ombudsman.nl/nieuws/persberichten/2008/documents/Rapport20080232.pdf>.

trial, new protection mechanisms should be introduced in the form of structural organized distrust within the neo-criminal justice system itself.<sup>82</sup>

The second face of the collateral damage to citizens is the disciplining effect of surveillance architectures. New forms of restraint, more subtle and varied than physical imprisonment, are imposed on groups encompassing more than just sophisticated criminals and terrorists. For example, ethnic minorities may be disproportionately stopped, frisked, and asked for identification in public spaces; antisocial people may be forced to comply with conditions of “Anti-Social Behaviour Orders”;<sup>83</sup> and perpetrators of sexual offenses may be required to register for life in a sexual offender registry with community notification,<sup>84</sup> to name but a few affected groups.<sup>85</sup> Legal protection for these new forms of restraint is significantly underdeveloped.<sup>86</sup> This is eloquently illustrated by the U.S. Supreme Court’s statement that “a statute that requires people to report for the rest of their lives to the government each time that they change hair color does not even invoke any constitutional scrutiny.”<sup>87</sup> Moreover, citizens who are not directly restrained because they happen not to belong to a hapless category of “abnormal” people, are nevertheless affected by the government’s panoptic power and may discipline themselves to conform to the prevalent paradigm of normality. Should this shift in the power relation between government and citizen not also be balanced by some new form of legal protection? This

---

82. Koops, *supra* note 64.

83. Anti-Social Behaviour Orders (ASBOs) allow authorities in the United Kingdom to impose an injunction on someone to refrain from further “antisocial” behavior, a breach of which is a criminal offense. *See* Crime and Disorder Act, 1998, c. 37 (Eng.); *see also* Macdonald, *supra* note 76 (arguing that ASBOs should be limited to repeat criminal offenders).

84. *Cf.* Jill S. Levenson & David A. D’Amora, *Social Policies Designed to Prevent Sexual Violence: The Emperor’s New Clothes?*, 18 CRIM. JUST. POL’Y REV. 168 (2007) (arguing that sex offender registration and notification laws have not achieved their goals). Note that certain non-sexual offenders also end up in sexual offender registries. *See* Ofer Raban, *Be They Fish or Not Fish: The Fishy Registration of Nonsexual Offenders*, 16 WM. & MARY BILL RTS. J. 497, 499 (2007) (“[A] textbook example of negligent policymaking supported by faulty data and upheld by often poor judicial reasoning.”).

85. *Cf.* DAVID LYON, SURVEILLANCE STUDIES: AN OVERVIEW 40 (2007) (“Controls are sought especially against ‘undeserving’ claimants and ‘dangerous’ offenders—and, even more, ‘terrorists’—with the result that it is the poor and the marginal who are most deeply affected.”).

86. *See* Murphy, *supra* note 77 (arguing that technologies of restraint are imposed without necessary procedural safeguards).

87. *Id.* at 191 (referring to an earlier discussion of *Connecticut Department of Public Safety v. Doe*, 538 U.S. 1 (2003), addressing Connecticut’s “Megan’s Law” that establishes a publicly available on-line sex-offender registry with photographs showing, among other things, the offender’s hair color).

question will be revisited after first examining how other power relations are shifting.

#### IV. EMPLOYER–EMPLOYEE<sup>88</sup>

This Part examines technology-related changes in the power relation of employers and employees, and assesses the consequences of these changes for the legal protection of employees. While advances in ICT in recent years have lifted workplace constraints for many employees, these advances have also subjected workers to increased scrutiny. The following two case studies—workplace monitoring and location monitoring—suggest that the limits of employer surveillance will have to be renegotiated. It is questionable, however, whether current legal-protection mechanisms, which are largely based on transparency and consent, will suffice to empower employees to engage in renegotiation.

##### A. CASE STUDY 1: WORKPLACE MONITORING

The workplace has changed drastically with the introduction of ICT. Contrary to early fears—or hopes—that many workers would become redundant through the automation of office tasks, ICT has not led to the replacement of workers, but rather to significant changes in the nature and organization of work processes. The advent of the Internet, in particular, and the attendant introduction of e-mail as a standard tool for communication have changed the nature of the work floor. Cyberspace has emerged alongside physical space as the place where work is carried out and has led to a rise in telecommuting from home. Moreover, the walls of the workspace have become permeable: employees at the office are regularly in contact with the outside world without immediately visible or audible signs.

The introduction of ICT in the workplace has affected the power relation between employers and employees in different ways. At the empowering end of the spectrum, ICT has enabled employees to conduct activities they could not do before, or could only do to a limited extent, during working hours or from the office. For instance, employees can now make an appointment with the dentist, order groceries online, chat with a friend at the other side of the world, download pornography, or search the web for more interesting jobs.

---

88. See generally Colette Cuijpers, *ICT and Employer-Employee Power Dynamics: A Comparative Perspective of United States' and Netherlands' Workplace Privacy in Light of Information and Computer Technology Monitoring and Positioning of Employees*, 25 J. MARSHALL J. COMPUTER & INFO. L. 37 (2007) (examining how ICT has affected the power balance between employer and employee and how adequately the existing legal framework has dealt with the resulting shifts in power).

The power of employers to make employees do what they would not otherwise do (e.g., work) has diminished somewhat as a result. More importantly, the power of employers to prevent employees from doing harm to the company has diminished. The huge number of outgoing SMS messages, e-mails, chats, and tweets, often drafted in informal language, could contain statements that are embarrassing or outright harmful for the company should they become public. Accounts of employees viewing or e-mailing pornography during work hours could also be damaging to a company's reputation. Finally, the risk that confidential business secrets or confidential documents may be leaked to third parties has grown substantially.<sup>89</sup> In these respects, ICT has weakened the power of employers.

In response, employers have taken countermeasures to rebalance the power relation. Primarily, they have started to routinely and extensively monitor employee communications. Workplace surveillance, by empowering the employer with new means of exercising control over employees, constitutes a shift in the power relation at the other end of the spectrum. A large majority of companies digitally monitor employee communications and activities.<sup>90</sup> Unsurprisingly, they often discover that employees are engaging in inappropriate activities and thereafter dismiss the employees.<sup>91</sup> Dismissal, of course, is one of the most far-reaching instruments of power employers possess (particularly during credit crunch crises), and the ability to dismiss

---

89. For an overview of liability risks for employers, see generally Michele Colucci, *The Impact of the Internet and New Technologies on the Workplace: A Legal Analysis from a Comparative Point of View*, in BULLETIN OF COMPARATIVE LABOUR (Roger Blanpain ed., 2002).

90. A 2007 survey by the American Management Association of 304 American companies showed that sixty-six percent monitor internet connections (and sixty-five percent block "inappropriate" websites); forty-five percent monitor computer activity, i.e., content, keystrokes, and time spent at the computer; forty-three percent monitor e-mail (over forty percent of which assign an individual to read e-mail); forty-five percent monitor telephones for time spent and numbers called, and sixteen percent record phone conversations; nine percent monitor voicemail; forty-eight percent use video surveillance to counter theft, violence, or sabotage, and seven percent use video surveillance to monitor on-the-job performance. Press Release, Am. Mgmt. Ass'n, 2007 Electronic Monitoring and Surveillance Survey: Over Half of All Employers Combined Fire Workers for E-Mail & Internet Abuse (Feb. 28, 2008), available at <http://press.amanet.org/press-releases/177/2007-electronic-monitoring-surveillance-survey>.

91. The Am. Mgmt. Ass'n survey showed that thirty percent of companies have fired employees for internet misuse, largely for viewing, downloading, or uploading inappropriate or offensive content (eighty-four percent), violation of any company policy (forty-eight percent), or excessive personal use (thirty-four percent); twenty-eight percent have fired employees for e-mail misuse, largely for violation of any company policy (sixty-four percent), inappropriate or offensive language (sixty-two percent), excessive personal use (twenty-six percent), or breach of confidentiality rules (twenty-two percent); and six percent have fired employees for misuse or private use of office phones. *Id.*

following workplace monitoring shows that technology is significantly empowering employers by strengthening the tools at their disposal.

In the United States, legal protection for the traditionally weaker party, employees, is found in privacy and employment law. However, the Fourth Amendment and the privacy tort of intrusion into seclusion have almost no bearing in light of the “reasonable expectation of privacy” doctrine. This is because the workplace is rarely considered a space where individuals may have any reasonable expectation of privacy, particularly when it comes to use of communication facilities provided by the employer.<sup>92</sup> ECPA protects communications privacy, but provides ample exceptions for employers, including the “provider exception,” the “normal course of employment,” and obtaining (implied) consent of the employee.<sup>93</sup> Employment law does not provide significant protection to employees against dismissal, because the doctrine of at-will employment still prevails.<sup>94</sup> Furthermore, exceptions to this doctrine developed in case law only apply in situations involving serious breaches of privacy.<sup>95</sup> Coupled with the absence of substantial privacy

---

92. See Cuijpers, *supra* note 88 (examining how ICT has affected the power balance between employer and employee and how adequately the existing legal framework has dealt with the resulting shifts in power); Ariana R. Levinson, *Industrial Justice: Privacy Protection for the Employed*, 18 CORNELL J.L. & PUB. POL’Y 609, 620, 619 (2009) (“[T]he employee’s right of privacy is a hollow shell against the lead weight of the employer’s claim to run his business as he pleases.” (quoting Clyde W. Summers, *Individualism, Collectivism and Autonomy in American Labor Law*, 5 EMPLOYEE RTS. & EMP. POL’Y J. 453, 468 (2001))) (“Most people ‘think they enjoy certain privacy protections when they are at work’ but they do not.”); Michael L. Rustad & Sandra R. Paulsson, *Monitoring Employee E-Mail and Internet Usage: Avoiding the Omniscient Electronic Sweatshop: Insights from Europe*, 7 U. PA. J. LAB. & EMP. L. 829 (2005) (discussing the prevalence of employer monitoring of employees’ e-mail and Internet use).

93. The “provider exception” allows any private employer who stores e-mail communications on her computer or network to access these communications. The “normal course of employment” exception applies when an employer can show that monitoring was necessary in the regular practice of employment, for example, to protect her company’s property or to provide the internal communication service in a proper manner. The “consent” exception allows employers to monitor communications with the consent of employees, which includes implicit consent that may be affected when an employer gives prior notice to her employees that she will monitor e-mail communications. *Cf.* Cuijpers, *supra* note 88 (examining how ICT has affected the power balance between employer and employee and how adequately the existing legal framework has dealt with the resulting shifts in power); Rustad & Paulsson, *supra* note 92 (discussing the prevalence of employer monitoring of employees’ e-mail and Internet use).

94. Katherine V.W. Stone, *Revisiting the At-Will Employment Doctrine: Imposed Terms, Implied Terms, and the Normative World of the Workplace*, 36 INDUS. L.J. 84, 85 (2007) (“[T]he contract is moment to moment for dismissal purposes but ongoing in relation to certain employer-imposed terms.”).

95. Cuijpers, *supra* note 88 (examining how ICT has affected the power balance between employer and employee and how adequately the existing legal framework has dealt with the resulting shifts in power).

protection of employees, the current state of employment law offers no safeguard for employees against dismissal when workplace monitoring shows inappropriate conduct or breach of company policy.

On paper, employee protection in the Netherlands is more robust. European privacy law applies in the workplace context,<sup>96</sup> and European data protection legislation provides strict rules for processing personal data from employees.<sup>97</sup> Furthermore, frequent involvement of the Works Council in defining monitoring policies provide stronger checks on what a company can define as proper use of ICT. However, privacy law is seldom invoked in dismissal cases in practice;<sup>98</sup> courts usually take recourse in employment law under the general standards of “good employership” and “good employeeship.”<sup>99</sup> Dismissal based on workplace monitoring takes into account three aspects: the grounds for workplace monitoring, the general principles of proportionality and subsidiarity, and the presence of a company policy with regard to Internet and e-mail use and monitoring. Even if the employer’s monitoring was considered illegitimate or disproportionate, the dismissal is usually condoned by the courts if the employee’s conduct did not conform to “good employeeship” or if the relationship between employer and employee has been seriously disrupted, which is usually the case.<sup>100</sup> Thus, even if employers abuse their power for workplace monitoring, employees who do not comply with company standards have little recourse to legal protection: they will not be reinstated in their job, nor will they get damages for breach of privacy.<sup>101</sup>

In summary, employees have gained new possibilities for communicating with the world outside the workplace and thus new opportunities for conducting personal activities during work hours. However, employer surveillance of employee communications has created a considerable risk of dismissal for employees if the employer decides that the activities observed are inappropriate, unlawful, or embarrassing to the company, or contrary to company policy. Legal protection, via privacy and employment law, provides employees with little recourse against dismissal if they have not conformed to what the employer has unilaterally defined as proper use of company

---

96. *See* Niemietz v. Germany, 16 Eur. Ct. H.R. 97 (1992); Halford v. United Kingdom, 24 Eur. Ct. H.R. 523 (1997).

97. Directive 95/46/EC, *supra* note 58.

98. *Cf.* Frank Hendrickx, *Privacy and Data Protection in the Workplace: The Netherlands, in REASONABLE EXPECTATIONS OF PRIVACY?: ELEVEN COUNTRY REPORTS ON CAMERA SURVEILLANCE AND WORKPLACE PRIVACY* 115, 139 (Sjaak Nouwt et al. eds., 2005).

99. BURGERLIJK WETBOEK [BW] book 7, art. 611 (Neth.).

100. Cuijpers, *supra* note 88, at 55.

101. *Id.*

facilities. The Netherlands' legal protection appears to provide stronger checks on what a company can define as proper use than that of the United States, primarily due to the standard of good employership and the Works Council's hand in limiting monitoring policies.

#### B. CASE STUDY 2: LOCATION MONITORING

New ICTs not only enable monitoring of communications, but also facilitate monitoring individuals' locations. In recent years, the market for tracking and tracing devices has boomed and applications in the private sector have grown dramatically. However, employers' use of employee localization services is still limited when compared to communications monitoring.<sup>102</sup> The fact that location tracking may also take place outside of company premises and beyond working hours makes it a poignant new technology in the employer–employee relationship. Cuijpers distinguishes four technologies that can be used for location monitoring: video surveillance, radio frequency identification (RFID),<sup>103</sup> mobile phone cell ID, and the Global Positioning System (GPS) for use outside of company premises. For employers whose businesses involve transport, these technologies are particularly intriguing. This includes not only taxi or cargo companies, but also businesses with company vehicles or company mobile phones that can have an interest in monitoring their employees' whereabouts.

There are few specific laws regulating location monitoring of employees. The legal framework for location monitoring in the United States is much the same as that for communications monitoring.<sup>104</sup> There is little to no privacy protection of employees, despite the fact that GPS can provide detailed records of privacy-sensitive locations visited, such as doctors' offices, casinos, striptease clubs, or labor rallies.<sup>105</sup> Employers can easily nullify reasonable expectations of privacy by notifying employees of their location monitoring policy.<sup>106</sup> The fact that monitoring has become more intrusive due to location monitoring of vehicles, phones, and other company-provided gadgets, which easily extend outside of working hours and off premises, does

---

102. The AMA survey of 2007 showed that eight percent of companies use GPS for tracking company vehicles, three percent use GPS to monitor cell phones, and less than one percent use GPS for monitoring employee smartcards. Press Release, Am. Mgmt. Ass'n, *supra* note 90. However, fifty-two percent use smartcards to control physical security and access to buildings and data centers, which may also involve some form of location tracking. *Id.*

103. RFID uses radio waves for identifying and tracking objects or persons, which are particularly useful within company premises.

104. *See supra* Section IV.A.

105. *Cf. State v. Jackson*, 76 P.3d 217 (Wash. 2003).

106. Cuijpers, *supra* note 88, at 70–71.

little to alter the legal status of such monitoring. Absent specific legislative protection for employees,<sup>107</sup> companies can impose any policy governing use of the equipment they provide to employees.<sup>108</sup>

In the Netherlands, more specific rules apply for processing location data, as provided in the Telecommunications Act. These rules, however, apply only to publicly provided communications networks or services, and companies will often use private networks or services for intra-company monitoring. Moreover, the data protection rules for location services are an extremely complex amalgam, which makes it difficult for both providers and data subjects to interpret which rules apply to which data in which situations.<sup>109</sup> In the absence of workable specific rules for location data, the legal status of location monitoring in the Netherlands, like in the United States, is much the same as for communications monitoring.<sup>110</sup> This means that although protection for employees may be better on paper than in the United States, it remains to be seen whether this makes a material difference in practice.

### C. DISCUSSION

The shifts in the relationship between employers and employees are more straightforward and less multifaceted than the shifts in the relationship between governments and citizens; the context in which this relationship takes shape, after all, is much smaller and simpler. The two case studies covered here—communication and location monitoring—address important aspects of employment, illustrating how intensively ICT has affected the nature of the workplace and the power relation between employer and employee. Two types of shifts take place in this power relation. On the one hand, employers lose power to control employees due to the ICT-facilitated permeability of the workplace, which allows employees to conduct more on-duty, non-work-related activities that pose higher risks for causing harm to the company. On the other hand, employers gain power to control employees by using comprehensive monitoring of communications and, increasingly, movement patterns, which may extend to off-duty and off-premises activities.

---

107. Statutory protection in federal and state law is limited and largely related to some specific activities, such as off-duty smoking. See Levinson, *supra* note 92, at 619.

108. See Press Release, Am. Mgmt. Ass'n, *supra* note 90 and accompanying text.

109. Collette Cuijpers & Bert-Jaap Koops, *How Fragmentation in European Law Undermines Consumer Protection: The Case of Location-Based Services*, 33 EUR. L. REV. 880 (2008).

110. Cuijpers, *supra* note 88 at 73.

The shifts in the power relation are symmetrical in that the increased surveillance of the workspace is a direct result of the increased permeability of the workspace. To the extent that employees benefit from new opportunities offered by ICT in the employment context, they are also under increased scrutiny when they use these new opportunities. Although this might imply that the shifts counterbalance each other, some aspects must be taken into account before we can draw conclusions about the legal protection of the traditionally weak party in this context.

First, the shifts in the power relation, even if they are symmetrical, broaden the context of employment considerably. The permeability of the workspace comes along with a blurring, both in spatial and in temporal terms, of work and private life. The exercise of power by the employer therefore gains a wider scope of application. Perhaps unlawful or inappropriate behavior by employees, even off-duty, has always been sufficient cause for dismissal, but the chances of observing such behavior and collecting demonstrable evidence of it are considerably higher as monitoring of employee behavior widens.

Second, notification duties play a crucial role. Since most of the ICT monitoring is invisible, employees may not be aware of the monitoring unless the employer has told them in advance. Because serious potential consequences—notably dismissal—can follow depending on what the monitoring uncovers, it is justified to at least notify employees of the monitoring system and the associated policy. Yet in the United States, only a handful of states have legislation requiring notification of electronic monitoring.<sup>111</sup> In contrast, in the Netherlands, notification is one of the core principles of data protection legislation<sup>112</sup> and a guideline in the Dutch Data Protection Authority's Framework Policy for E-Mail and Internet Use.<sup>113</sup> A further mechanism for alerting employees about employee monitoring is included in the Dutch Works Councils Act, which requires that personnel tracking systems [*personeelsvolgsystemen*] are approved by the Works Council.<sup>114</sup> Admittedly, having a notification duty on paper does not necessarily guarantee that employees are actually made aware; that will depend on the

---

111. Phillips, *supra* note 49; Levinson, *supra* note 92, at 622 (referring to CONN. GEN. STAT. § 31-48(d) (2009); DEL. CODE ANN. tit. 19, § 705 (2002)).

112. See Wet bescherming persoonsgegevens [Personal Data Protection Act], Staatsblad van het Koninkrijk der Nederlanden [Stb.] 302 (2000), ch. 5 (Neth.).

113. See College Bescherming Persoonsgegevens, *Raamregeling Voor Het Gebruik van E-mail en Internet* (2002), available at [http://www.cbpweb.nl/downloads\\_av\\_sv/AV21\\_raamregeling.pdf?refer=true&theme=purple](http://www.cbpweb.nl/downloads_av_sv/AV21_raamregeling.pdf?refer=true&theme=purple).

114. See Dutch Works Councils Act art. 27(1)(1).

implementation in practice, for example, in what form and when the notification takes place.<sup>115</sup>

A third relevant aspect is whether the company policy governing the use of company facilities, including computers, Internet, mobile phones, and vehicles, strikes a fair balance between employer and employee interests. Again, there seem to be better legal safeguards in place in the Netherlands. For instance, the Works Councils Act requires works councils to approve monitoring policies and a provision in the criminal law prohibits “obvious misuse” by the employer of his right to monitor employee communications.<sup>116</sup> In the United States, collective bargaining mechanisms and the National Labor Relations Act may provide some safeguards, but these are limited in scope and scale.<sup>117</sup> Levinson has argued that the “law of the shop” as applied by arbitrators should be used more widely to provide better safeguards to employees against arbitrary monitoring.<sup>118</sup>

Together, these aspects caution against concluding that the empowering and disempowering shifts in the employer–employee relationship counterbalance one another. The broadening of the scope of employer monitoring, both within the workplace and off-duty and off-premises, is not necessarily balanced by the mechanisms for curbing the employer’s power, if notification and supervision of the fairness of monitoring policies are largely absent, which seems the case in the United States but also, perhaps, in the Netherlands due to the lack of enforcement of privacy protection in dismissal cases.<sup>119</sup> Moreover, even if the policy is fairly balanced and employees are duly notified of it, the stakes are higher for employees nonetheless: employees are under more prevalent scrutiny, and at times and in places that used to be reserved for purely private, non-work-related activities.

---

115. For example, in small print in a leaflet about company policies given to new employees on their first day of work or hidden somewhere in an attic room of the company’s internal homepage—neither of which is very likely to truly inform employees—or, at the other extreme, by a notice appearing on the screen each time the employees log in on their computer.

116. DUTCH CRIMINAL CODE art. 139c(2)(2).

117. Levinson, *supra* note 92 at 622; cf. M.W. Finkin, *Information Technology and Workers’ Privacy: The United States Law*, 23 COMP. LAB. L. & POL’Y J. 471, 498 (2002) (noting that “the percentage of the civilian labor force eligible for union representation that actually is unionized, has fallen to a post-War low of 9.4%”).

118. Levinson, *supra* note 92, at 639; see also National Academy of Arbitrators, <http://www.naarb.org/> (last visited Sept. 25 2009).

119. See *supra* note 88 and accompanying text.

This parallels the two types of collateral damage to citizens caused by the arms race in criminal investigation: the risk of interpretation errors and the mass disciplining and normalization of behavior by employees.<sup>120</sup>

Increased monitoring of employee behavior has amplified the risk of errors. For example, if a civil servant searching for public policy information absent-mindedly types in [www.whitehouse.com](http://www.whitehouse.com) (instead of [.gov](http://www.whitehouse.gov)) or [www.amsterdam.com](http://www.amsterdam.com) (instead of [.nl](http://www.amsterdam.nl)) and finds himself confronted with flashy X-rated pictures,<sup>121</sup> can the monitoring system distinguish him from an employee who is actually surfing for adult material? If an employee has to deliver a package in the center of Amsterdam and, due to road renovations and subway constructions, gets lost and ends up in the red-light district, does he have a defense against the proof of the car tracking system? Or what if his nineteen-year-old son “borrowed” the car for an excursion into red-light nightlife? Such errors in interpretation, including technical errors, can and will usually be redressed somewhere in the procedure. Sometimes, however, it might be too late; the procedure itself may have a negative effect on the employer–employee relationship.<sup>122</sup> For example, an employer’s false accusation may have induced the employee to circulate enraged messages among his colleagues with insulting remarks about his employer.

More important is the mechanism, comparable to the increased footprint of criminal law to cover trivial offenses or antisocial types of behavior, of sanctioning employees for undesirable off-duty or off-premises activities that do not cause significant harm to the employer, but that nonetheless fall within the ambit of “unacceptable” behavior as defined by the employer. Although off-duty and off-premises monitoring does not yet take place on a large scale, when it happens it has significant potential effects on the freedom of employees to behave as they wish.

Here, we encounter the second face of collateral damage to citizens: the potentially disciplining effects of ubiquitous surveillance. Although stricter limits apply to the monitoring powers of employers when it comes to off-duty behavior, monitoring is allowed to a degree by employment or privacy

---

120. *Supra* Section III.D.

121. Both sites nowadays are relatively respectable-looking websites, but they started out as porn websites. *See, e.g.*, Lodewijk F. Asscher, *Schuldige domeinnamen*, 10 *COMPUTERRECHT* 186 (2003); Jeff Peline, *Whitehouse.com Goes to Porn*, *CNET NEWS*, Sept. 5, 1997, <http://news.cnet.com/2100-1023-202985.html>.

122. *See generally* Cuijpers, *supra* note 88 (referring to J. Yung, *Big Brother IS Watching: How Employee Monitoring in 2004 Brought Orwell's 1984 to Life and What the Law Should do About It*, 36 *SETON HALL L. REV.* 163, 180 (2005), noting the risk of abuse of power: employers could use a relatively harmless incident to fire an employee if they want to get rid of her for other reasons that might be less successfully argued in court).

law, particularly if it concerns employer-provided equipment, such as a cell phone or computer.<sup>123</sup> Awareness that the employer is potentially monitoring activities conducted with such equipment could lead to (over)cautiousness in the use of the equipment. Although an employee could refrain outright from using company equipment for personal purposes, that option does not align with the reality of the “new economy.” Private use of employer property

is often regarded as a kind of fringe benefit. Employees expect some leniency with regard to the private use of company assets, and the employer often encourages this use to circumvent employees’ nine-to-five mentality. A lot of companies provide employees with home computers or mobile telephones which they can use for private purposes. The added value for employers lies in the fact that the employee can be reached for business related purposes 24-hours-a-day.<sup>124</sup>

In this new constellation of blurred boundaries between office and home, and working hours and spare time, the limits of employer surveillance will have to be renegotiated. It is not clear, however, that current legal mechanisms for employee protection provide employees with sufficient bargaining power in this renegotiation process. The easy way out may well be to accept the increased monitoring scope and to normalize private behavior in a self-disciplining act of conforming to the company standards for acceptable behavior.

Internalizing acceptability standards contributes to the erosive effect of technology on privacy.<sup>125</sup> David Phillips calls attention to the “vicious circularity” of ever more invasive surveillance techniques at the workplace:

courts have found that employees reduce or extinguish their reasonable expectation of privacy when they explicitly consent to employers’ search policies. Employers, then, demand such consent as a matter of standard business practice. That standard practice then becomes implicit in the community norms generally governing the workplace surveillance. Eventually, consent to search becomes implicit in the employment relationship.<sup>126</sup>

---

123. Levinson, *supra* note 92.

124. Cuijpers, *supra* note 88, at 85.

125. *See generally* Bert-Jaap Koops & Ronald Leenes, “Code” and the Slow Erosion of Privacy, 12 MICH. TELECOMM. & TECH. L. REV. 115 (2005) (finding that technology generally makes privacy violations easier and erodes reasonable expectations of privacy).

126. Phillips, *supra* note 49, at 60.

Communications monitoring can already be called a standard practice; location monitoring is only beginning to be included in employer policies.<sup>127</sup> If both types of monitoring continue in the vicious circularity of ingrained, normalized standards, then the workplace will acquire distinct characteristics of the Panopticon and Foucauldian self-disciplining effects on employees. Current protection mechanisms based on transparency (notification) and consent (employment at will) seem inadequate to deal with such a shift in the power relationship between employers and employees.

## V. BUSINESS–CONSUMER<sup>128</sup>

This Part examines technology-related changes in the power relation of businesses and consumers, and assesses the consequences of these changes for consumer protection. The two following case studies—profiling and behavioral advertising and buying goods or services online—suggest that, although both consumers and businesses gain substantially in their information position, businesses gain considerably more power to seduce consumers to buy goods that they would not otherwise buy. The subsequent general discussion will analyze whether current legal measures of consumer protection are equipped to deal with businesses' exercise of power. Particularly relevant here is the second dimension of power: the indirect influencing of consumers' actions through “agenda-setting” mechanisms of targeted advertising and website design.<sup>129</sup>

### A. CASE STUDY 1: PROFILING AND BEHAVIORAL ADVERTISING

Commerce often starts with advertising. Traditionally, this is a fairly crude mechanism. Advertisements are directed at a large group of people who happen to read the same newspaper or watch the same television channel; and businesses can only target their advertising to the extent that they know something about the average consumer of the medium. ICT is enabling the use of profiling techniques to offer personalized advertising to consumers. This mechanism is relatively new and qualitatively different from the traditional information position of businesses. Profiling provides a new type of knowledge, based on patterns discovered by correlating data in

---

127. See generally Cuijpers, *supra* note 88.

128. This Section builds on Colette Cuijpers, *The Influence of ICT on Consumer Protection: Empowerment or Impairment of the Consumer?* (TILT Law & Tech. Working Paper Series, Paper No. 015/2009, 2009), available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1515790](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1515790).

129. See *supra* note 6 and accompanying text for a discussion of the second dimension of power.

databases.<sup>130</sup> Information about consumption patterns, such as different products that certain types of consumers often buy together or types of books that people with certain characteristics are more likely to buy, can be used to target advertising to specific consumers in concrete contexts. Rather than displaying any advertising banner, a website may choose a specific advertisement based on the prospective buyer's clickstream, search words, zip code, or other data that the user has filled in on a web form.

Such personalized or behavioral advertising is a promising innovation in commerce. Since it customizes the advertising to align with the consumer's inferred interests, it can be more effective for both businesses and consumers. Indeed, personalization can be "an effective tool to achieve an efficient market."<sup>131</sup> However, the promise does not come without threats. Behavioral advertising is usually based on group profiles, which will almost always be probabilistic and non-distributive, i.e., not all members of the group defined by the profile will share all the attributes of the group profile.<sup>132</sup> In other words, false positives are bound to occur: someone who has bought Koontz's *Mr. Murder* and Weldon's *The Cloning of Joanna May* because she is interested in fiction about clones may not be at all interested in other pulp thrillers or feminist novels. Some targeted advertisements will therefore miss their mark, through false positives and false negatives. Although this may lead to "unanticipated encounters" in which consumers are confronted with undesirable or irritating information that they have not sought out,<sup>133</sup> this is not generally a serious threat. Compared to traditional, non-personalized advertising, the error rate, particularly of false positives, will be much lower and missed opportunities for advertising do not impair the ability of consumers to buy goods of their own initiative.<sup>134</sup>

---

130. Mireille Hildebrandt, *Defining Profiling: A New Type of Knowledge?*, in PROFILING THE EUROPEAN CITIZENS 17 (Mireille Hildebrandt & Serge Gutwirth eds., 2008).

131. Simone van der Hof & Corien Prins, *Personalisation and Its Influence on Identities, Behaviour and Social Values*, in PROFILING THE EUROPEAN CITIZEN 111 (Mireille Hildebrandt & Serge Gutwirth eds., 2008).

132. See BART HERMAN MARIA CUSTERS, THE POWER OF KNOWLEDGE: ETHICAL, LEGAL, AND TECHNOLOGICAL ASPECTS OF DATA MINING AND GROUP PROFILING IN EPIDEMIOLOGY 61 (2004) (discussing non-distributivity).

133. Cass R. Sunstein, *The Daily We: Is the Internet Really a Blessing for Democracy?*, BOSTON REV., Summer 2001, available at <http://bostonreview.net/BR26.3/sunstein.html>.

134. There may be a concern, however, if the targeted advertising is based on sensitive data, even if the consumer is not personally identifiable; consumers can then "view it as invasive or, in a household where multiple users access one computer, it may reveal confidential information about an individual to other members." Press Release, Fed. Trade Comm'n, Online Behavioral Advertising: Moving the Discussion Forward to Possible Self-Regulatory Principles 5 (Dec. 20, 2007), available at <http://www.ftc.gov/os/2007/12/P859900stmt.pdf>. *But cf.* van der Hof & Prins, *supra* note 131.

There is a more subtle and serious threat being exercised in behavioral advertising, which relates to the second dimension of power.<sup>135</sup> By personalizing the offers shown to online consumers, businesses influence the horizon of consumers' interest: it is a form of agenda-setting. This "may force individuals into restrictive two-dimensional models,"<sup>136</sup> reducing the consumer's areas of interest into simplified machine-readable patterns and resulting in a potential loss of nuances and of occasional side-steps into marginal or new areas of interest. It may also lead to normalization of consuming behavior, through the panoptic logic of "the system":

[w]hen the system seems to know what you want better and earlier than you do, how can you know where these desires really come from? . . . [P]rofiles will begin to normalize the population from which the norm is drawn. The observing will affect the observed. The system watches what you do; it fits you into a pattern; the pattern is then fed back to you in the form of options set by the pattern; the options reinforce the patterns; the cycle begins again.<sup>137</sup>

In other words, if you are persistently being offered pulp thrillers and feminist novels because your online bookshops think you should be interested in them, you might as well give it a try because there should be some merit to the recommendations (why else would the system give them?), thereby reinforcing your profile and leading to more of the same offers. Thus, you might well end up reading only these types of books because it is—by the system and through panoptic logic by yourself—expected of you.

How do current legal-protection mechanisms deal with behavioral advertising? In general, data protection cannot be invoked as long as group profiles are being used to show ads to unidentifiable online consumers.<sup>138</sup> However, if a European consumer is identifiable, for example by an IP address,<sup>139</sup> when showing her an ad based on a group profile of certain online

---

135. See *supra* note 6 and accompanying text for a discussion of the second dimension of power.

136. Van der Hof & Prins, *supra* note 131, at 121.

137. LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE 154 (1999).

138. This constitutes one of the weaknesses of the Data Protection Directive in an online environment, since potential privacy or discrimination threats to web users often do not depend on their being identifiable as Jill the Plumber from Tuscaloosa, Alabama, but on their being recognized as being the same person as an earlier website visitor or their being traced throughout a session. See Ronald E. Leenes, *Do You Know Me? Decomposing Identifiability* (Tilburg Univ. Legal Studies, Working Paper No. 001/2008, 2008) (distinguishing between L-identifiability ("looking-up"), R-identifiability ("recognition"), and S-identifiability ("session")).

139. IP addresses can be considered as personal data. See ARTICLE 29 DATA PROTECTION WORKING PARTY, OPINION 4/2007 ON THE CONCEPT OF PERSONAL DATA

behavior or personal characteristics, her personal data are covered within the ambit of the Data Protection Directive. The Directive also applies to the collection of behavioral or personal information from identifiable consumers. This theoretically could provide some legal protection, for example, against disproportionate collection of data or application of a “wrong” group profile; but it is highly dubious whether the Directive can actually be enforced in such a context.<sup>140</sup>

In the United States, no specific data protection rule seems to apply to the collection of data and the display of advertisements in this context. The Federal Trade Commission (FTC), however, has recommended that

[e]very website where data is collected for behavioral advertising should provide a clear, concise, consumer-friendly, and prominent statement that (1) data about consumers’ activities online is being collected at the site for use in providing advertising about products and services tailored to individual consumers’ interests, and (2) consumers can choose whether or not to have their information collected for such purpose.<sup>141</sup>

Much of the efficacy of such a recommendation will depend on how consumers are being informed and given a choice—only a handful of knowledgeable and privacy-aware consumers might be able to understand and act upon the issue.<sup>142</sup> Moreover, it remains to be seen whether business self-regulation as advocated by the FTC really works in this area.

---

16–17 (2007), available at [http://ec.europa.eu/justice\\_home/fsj/privacy/docs/wpdocs/2007/wp136\\_en.pdf](http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf). This is not uncontested in the literature, but Opinions of the Article 29 Data Protection Working Party serve as important guidelines for courts in the E.U. to interpret the Data Protection Directive.

140. One wonders what Amazon.co.uk would include in a reply to a request from a Dutch customer such as:

Could you inform me on what basis you recommended E.M. Forster’s *Maurice* and Baldwin’s *Giovanni’s Room* to me (*see* Wet bescherming persoonsgegevens [Dutch Data Protection Act], art. 35), whether this relates to my having bought Leavitt’s *The Lost Language of Cranes* last month (*see id.* at art. 33) and whether you have therefore profiled me as being interested in homosexuality, which is sensitive personal data (*see id.* at art. 16), and would you please delete all these data since it is unlawful for you to process them (*see id.*) and irrelevant for my buying books with you (*see id.* at art. 36(1)), and can you inform me in writing of your having deleted my sensitive personal data (*see id.* at art. 36(2))?

141. Press Release, Fed. Trade Comm’n, *supra* note 134, at 3.

142. *Id.* (noting that “panelists recognized that many consumers do not read privacy policies and raised a genuine question about consumers’ willingness and ability to read and understand long disclosures about privacy”).

Specifically relevant in this case study is the legal protection against deceptive advertising. This, after all, is the true concern about enticing consumers to buy goods they would not otherwise buy. Deceptive advertising is thus seen as an abusive exercise of power by businesses; to protect consumers, this is prohibited both in European and U.S. legislation.<sup>143</sup> However, as with data protection rules, this legal protection will be difficult to enforce; consumers are often unaware of the practice, and even if they are, the damage for each individual consumer deceived by a behavioral advertisement will usually be relatively low compared to the time, money, and tools they have to invest in pursuing a contract or tort claim.<sup>144</sup> More importantly, however, it is not clear whether behavioral advertising is at all deceptive. The purpose, after all, is to better target the offer to the consumer's own preferences, and the personalization as such does not make it manipulative. Only the agenda-setting and interest-shaping aspect of behavioral advertising could be considered deceptive for consumers who embrace the profile underlying the advertisement through panoptic logic,<sup>145</sup> even though they had no obvious prior interest in the offered product. But it is unlikely that this subtly manipulative effect, which works through an act of double anticipation by the consumer herself,<sup>146</sup> is sufficient to fall within the scope of deceptive advertising rules.

#### B. CASE STUDY 2: BUYING ONLINE

The advent of e-commerce has provided consumers not only with a new means to do old business—buying goods or services—but also opened up a far wider range for conducting this business. Rather than leaving their homes to shop locally, consumers can now search for goods around the world. With automated search engines, websites comparing products and prices, and

---

143. See, e.g., Directive 2005/29/EC, 2005 O.J. (L 149) 22 (EC); Federal Trade Commission Act, 15 U.S.C. §§ 42–58 (2006); Lanham Act, 15 U.S.C. § 1125(a) (2006).

144. See Marla Pleyte, *Online Undercover Marketing: A Reminder of the FTC's Unique Position to Combat Deceptive Practices*, 6 U.C. DAVIS BUS. L.J. 14 (2006); Willem van Boom & Marco Loos, *Effective Enforcement of Consumer Law in Europe: Private, Public, and Collective Mechanisms*, in COLLECTIVE ENFORCEMENT OF CONSUMER LAW 231 (Willem H. Van Boom & Marco Loos eds., 2007).

145. See *supra* note 137 and accompanying text.

146. I.e., the consumer changes her preferences—buying a product she would not otherwise have bought—in anticipation of the interest profile she thinks is reflected in the advertisement that anticipates her preferences. On the similar mechanism of double anticipation in identity building, see WP7, D7.14A: WHERE *IDEM*-IDENTITY MEETS *IPSE*-IDENTITY: CONCEPTUAL EXPLORATIONS (Mireille Hildebrandt et al. eds., 2008), available at [http://www.fidis.net/fileadmin/fidis/deliverables/fidis-WP7-del7.14a-idem\\_meets\\_ipse\\_conceptual\\_explorations.pdf](http://www.fidis.net/fileadmin/fidis/deliverables/fidis-WP7-del7.14a-idem_meets_ipse_conceptual_explorations.pdf). See also *infra* Section VI.A.

auction and trading sites, consumers have an array of tools available to help them find the best products for the lowest prices.

There are more aspects in which ICT is empowering consumers. They can form ad hoc online collectives that use mass buying to obtain the lowest possible price from e-sellers.<sup>147</sup> Similarly, businesses are being profiled by ad hoc collections of consumers who together build and maintain ratings websites with assessments of businesses' quality, service level, and reliability. A hotel owner now must not only be friendly to Mr. Michelin or Miss Lonely Planet when they visit once a year, but to each and every customer, or risk receiving unfavorable reviews on a rating website.

This does not imply, however, that the Internet constitutes a Shangri-La of pervasive consumer power. Rating and experience-sharing websites are not necessarily reliable:

[t]his valuable source of information is diminished when online undercover marketers are allowed to surreptitiously infiltrate such sites and plant self-interested messages about their products. These advertisers are well-funded and sophisticated enough to craft messages that are extremely believable and likely to induce consumer reliance. As a result, these practices turn a valuable source of information into a source of disinformation for consumers.<sup>148</sup>

In other words, the potentially most powerful tool for consumer empowerment—peer review—may backfire, since businesses can turn it to their own advantage to seduce consumers to buy their products.

When we focus on the actual online buying process, we observe that consumers no longer depend on the local bookstore or camera shop; they also can shop from their desk chair for the best deal in as large a region as they care to explore. Moreover, websites allow for a full presentation of the general terms and conditions, rather than a scant reference to paper documents that can be inspected somewhere or snail-mailed upon request. E-consumers thus have, in principle, a much wider scope for buying as well as better knowledge of the product and the terms and conditions covering the sale.

---

147. Robert J. Kauffman & Bin Wang, *Bid Together, Buy Together: On the Efficacy of Group-Buying Business Models in Internet-Based Selling*, in *THE E-BUSINESS HANDBOOK* 99 (Paul Benjamin Lowry et al. eds., 2002). For examples of collective buying sites, see, for example, Groupon Deal of the Day: Find Great Deals on Fun Things to Do in San Francisco, <http://www.groupon.com/> (last visited Sept. 25, 2009) and Pingel Partner, <http://www.pingelpartner.nl/> (last visited Sept. 25, 2009).

148. Pleyte, *supra* note 144.

Several obstacles, however, decrease the empowering potential of online buying. A website's design can make a business' terms and conditions difficult to find. Alternately, consumers are increasingly presented with terms and conditions before they can make a purchase. However, because the terms and conditions are obfuscated by legalese and small print, few e-consumers will read and understand these terms. Furthermore, a physical product cannot be seen, let alone touched and immediately taken away. Thus, the consumer has to rely on pictures and on the seller's reliability to send the correct product depicted in the image.

This uncertainty is offset by increased options for redress. The European Distance-Selling Directive, for example, allows consumers to return online-purchased goods without providing a reason within seven days.<sup>149</sup> Still, returning a defective good will not always be cost-effective for consumers if the defect is relatively small—receiving a yellow coffee machine when you thought you were ordering an orange one—particularly if the consumer needs to spend precious time on repackaging and going to the post office. Finally, “scattered damage” (i.e., many trifling losses that are too minor for individuals to seek redress for, but that constitute a significant loss on a collective scale) is a problem that will occur more frequently as e-commerce expands.

Perhaps the biggest obstacle is that these downsides are exacerbated in the cross-border context of online commerce. Terms and conditions are not necessarily available in one's own language (particularly for native speakers of relatively small languages, such as Dutch, Italian, or Hungarian, not to mention minority languages like Frisian or Kwakiutl). The contract may be embedded in foreign legal systems with possibly unfamiliar rules and presuppositions. And redress is more costly and cumbersome when returning a package to businesses abroad. Some mechanisms for cross-border redress are starting to emerge, for example, in the network of European Consumer Centres.<sup>150</sup>

Another relevant aspect is what happens with the consumer data that are gathered by businesses throughout the process. ICT can empower businesses, who can gather huge amounts of information about consumers through mechanisms such as cookies and web forms. Privacy statements will inform consumers about the purposes and conditions for processing the personal data collected during e-commerce activities, but it is unclear how

---

149. Directive 97/7/EC, art. 6, 1997 O.J. (L 144) 19 (EC).

150. See ECC-Net, [http://ec.europa.eu/consumers/redress\\_cons/index\\_en.htm](http://ec.europa.eu/consumers/redress_cons/index_en.htm) (last visited Feb. 4, 2010).

many consumers actually find, read, and understand privacy statements. Even if they do, it is not clear to what extent they can effectively resist undesirable provisions—such as selling data to third parties—in a market that is dominated by information brokers.

The legal protection of consumers has already been adapted in several respects to the new reality of e-commerce, at least in Europe. The E.U. “Consumer Acquis” consists of many directives with consumer-protection rules.<sup>151</sup> Particularly relevant here are the Unfair Commercial Practices Directive and the Distance Selling Directive.<sup>152</sup> These contain numerous information obligations, such as requirements to provide information beforehand in a comprehensible and durable form,<sup>153</sup> and balancing requirements to enhance the fairness of terms and conditions.<sup>154</sup> The current framework of European consumer protection, nevertheless, is very fragmented, and a new Directive on Consumer Rights has been proposed that aims to bring together and harmonize key consumer rights.<sup>155</sup> In contrast to the European legislative approach, “[t]he U.S. legal system has tried, at times awkwardly, to fit the new transactions into existing doctrinal categories,”<sup>156</sup> which leaves consumer protection primarily to the market.

Altogether, the information gains that the Internet allows constitute a significant shift towards consumer empowerment. However, this empowerment is lessened by several factors, such as the risk of information overload and the non-transparent nature of many information providing websites that may manipulate results for commercial reasons. Nevertheless the information position of ICT-savvy consumers is superior to their information position in traditional, physical-space commerce. The scope for buying goods has also expanded enormously, and despite obstacles for e-commerce, notably in cross-border contexts, this can likewise be seen as a

---

151. The “Consumer Acquis” is an umbrella term used to indicate the widespread collection of consumer-protection rules in E.U. legislation. *See generally Proposal for a Directive of the European Parliament and of the Council on Consumer Rights*, COM (2008) 614 final (Oct. 8, 2008) (reviewing the Consumer Acquis, including a number of directives on consumer protection, and proposing changes to simplify and harmonize the current fragmented regulatory framework).

152. Directive 2005/29/EC, *supra* note 143; Directive 97/7/EC, *supra* note 149.

153. *See* Directive 97/7/EC, *supra* note 149, at art. 4–5; Directive 2000/31, art. 5, 6, 10, 2000 O.J. (L 178) 1 (EC).

154. *See* Directive 2005/29/EC, *supra* note 143, at art. 3.

155. *Proposal for a Directive of the European Parliament and of the Council on Consumer Rights*, COM (2008) 614 final (Oct. 8, 2008).

156. Jane K. Winn & Brian H. Bix, *Diverging Perspectives on Electronic Contracting in the U.S. and the EU*, 54 CLEV. ST. L. REV. 175, 190 (2006).

significant empowerment of consumers, supported as they are by new consumer-protection rules in legislation or case-law doctrine.

Some caution is warranted, however, when it comes to assessing the overall effect on consumers at large: not all consumers benefit equally from the new possibilities. These new possibilities are perhaps real options only for experienced ICT users with a sufficiently perceptive and critical attitude to web-based information sources. Furthermore, the information position of businesses is also strengthened considerably, through information- collecting, sharing, and profiling tools. They can use this information to better attract and bind consumers to them. Just as consumers have access to a broader array of businesses, businesses also have significantly increased their capability for finding customers. With the varied effect on different types of consumer groups, this implies that less ICT-savvy and less critical consumers may now be more vulnerable to abusive exercises of power by businesses.

### C. DISCUSSION

The commerce context presents perhaps the most empowering potential of ICT yet encountered in the case studies. The Internet has opened up a wide range of opportunities for consumers to counter businesses' efforts to seduce them into buying their products and services. Gathering information, shopping irrespective of place, and forming consumer collectives are important consumer-empowering mechanisms. One might question to what extent all consumers benefit from these possibilities: perhaps only the technology-savvy consumers are exploiting them in practice. Compared to the other power relations studied here, consumers benefit more from new technology-facilitated opportunities than citizens in their relationship with the government, but perhaps less than employees in their relationship with the employer. Determining to what extent the "average" consumer actually makes use of information-gathering and collective-pressure mechanisms is a matter for further study.<sup>157</sup>

---

157. The current legal status of the average consumer with respect to her ICT awareness or tech-savviness is indeterminate. For example, European Court of Justice case-law on the free movement of goods and services seems to assume a relatively high level of activity and knowledge of consumers, whereas the Consumer Acquis—the fragmented system of consumer protection in many first-pillar Directives—seems to treat the consumer as relatively passive and poor-informed. See Hannes Unberath & Angus Johnston, *The Double-Headed Approach of the ECJ Concerning Consumer Protection*, 44 COMMON MKT. L. REV. 1237 (2007); Vanessa Mak, *Harmonisation Through 'Directive-Related' and 'Cross-Directive' Interpretation: The Role of the ECJ in the Development of European Consumer Law* (Filburg Inst. of Comparative & Transnational Law, Working Paper No. 2008/8, 2008).

At the same time, ICT also has significant empowering effects for businesses, who can gather huge amounts of information about consumers, both specifically through tools like cookies and web forms and generically through profiling. They can also target consumers cost-effectively with advertisements and offers, on a massive scale unimaginable in the pre-ICT era (i.e., spam) but also on an individual, personalized level (i.e., behavioral advertising).

Both consumers and businesses have thus gained substantially in their information position. Like the case of employers and employees, it is far from evident that these trends counterbalance each other, particularly since the trends are less clearly intertwined than in the workplace monitoring cases.<sup>158</sup> With some justification, Dholakia and Zwick conclude that “[t]he power balance has shifted to the marketers.”<sup>159</sup> After all, the strengthened position of businesses gives them considerable power to seduce consumers to buy goods they would not otherwise buy.

This increased power can only be outweighed by the empowering information-gathering possibilities for consumers, provided that the business activities are sufficiently transparent for the consumer. This is a key issue both in behavioral advertising and in online buying with impenetrable terms and conditions and privacy statements. Does the consumer in these situations know on what basis she is being offered something, or what will happen with the personal data that are collected when she buys something online? The legal protection of consumers has been adapted in some respects to the new reality of e-commerce, through information obligations aimed at enhanced transparency.<sup>160</sup> Such measures surely help to balance the power relation, but they are probably insufficient to protect consumers in all

---

158. See *supra* Section IV.C.

159. Nikhilesh Dholakia & Detlev Zwick, *Privacy and Consumer Agency in the Information Age: Between Prying Profilers and Preening Webcams*, 1 J. RES. FOR CONSUMERS 18–19 (2001), available at [http://www.jrconsumers.com/academic\\_articles/issue\\_1?f=5800](http://www.jrconsumers.com/academic_articles/issue_1?f=5800). As they observe,

real-time customization of interactive messages can actually limit the ability of the consumer to shape his or her ideas of market prices, product variability, and quality, among other things. In such a scenario—of which we can see the first signs in the electronic marketplace—real-time interactivity does not enable consumer choice and informed decision-making, but delimits consumer freedom and unrestrained agency in the market.

*Id.* at 12–13; cf. LYON, *supra* note 71, at 127–28 (noting that “while the public awareness of consumer surveillance may be rising, it is undoubtedly doing so at a rate far slower than the opportunities for consumer surveillance are being exploited”).

160. See *supra* note 151 and accompanying text.

respects. Several authors stress that more legal protection is needed to decrease the information asymmetry between consumers and business,<sup>161</sup> to increase transparency,<sup>162</sup> and to allow more room for collective action in order to address the problem of large amounts of small individual damages where access to justice for individual consumers is unattractive.<sup>163</sup>

Two issues also call for attention, relating closely to the two faces of collateral damage identified in the government–citizen power relation shifts.<sup>164</sup> First, the consumer has become more vulnerable through the increased collection and storage of personal data. Not only can these data be used for other purposes—for example, when sold to third parties—but they can also be leaked through inadequate security measures, and subsequently be used for financial identity theft.<sup>165</sup> The lack of a potent and practically enforceable data protection regime<sup>166</sup> is apparent. Also, the occurrence of interpretation errors may be relevant; for example, the product searched for or bought might be for someone else—a gift, a purchase for a bed-ridden neighbor, or a family member to whom you lent your credit card. This will not always lead to concrete damage, but the erroneous profile thus

161. *E.g.*, Hildebrandt, *supra* note 68, at 308; Els Soenens, *Web Usage Mining for Web Personalisation in Customer Relation Management*, in *PROFILING THE EUROPEAN CITIZEN: CROSS-DISCIPLINARY PERSPECTIVES* 175, 180 (Mireille Hildebrandt & Serge Gutwirth eds., 2008).

162. Detlev Zwick & Nikhilesh Dholakia, *Whose Identity Is It Anyway? Consumer Representation in the Age of Database Marketing*, 24 *J. MACROMARKETING* 31, 40–41 (2004) (arguing that “the power to constitute consumer identity . . . is located within the database and that ‘only if consumers are given full access to companies’ customer databases can they maintain a sense of control over their identities in the marketplace”).

163. Van Boom & Loos, *supra* note 144.

164. *See supra* Section III.D; *see also supra* Section IV.C.

165. *See, e.g.*, Jennifer A. Chandler, *Negligence Liability for Breaches of Data Security*, 23 *BANKING & FIN. L. REV.* 223 (2008) (discussing the need for civil liability to increase data security and problems plaintiffs face in civil lawsuits); Sasha Romanosky et al., *Do Data Breach Disclosure Laws Reduce Identity Theft?* (Sept. 16, 2008) (unpublished manuscript, on file with author), available at <http://ssrn.com/abstract=1268926>.

166. Note the overall conclusion of Neil Robinson et al., *Review of the European Data Protection Directive* (2009) that

as we move toward a globally networked society, the Directive as it stands will not suffice in the long term. While the widely applauded principles of the Directive will remain as a useful front-end, they will need to be supported by a harms-based back-end in order to cope with the growing challenge of globalisation and international data flows.

*Id.* at vii; *see also* F. FABBRINI ET AL., *COMPARATIVE LEGAL STUDY ON ASSESSMENT OF DATA PROTECTION MEASURES AND RELEVANT INSTITUTIONS* (EUI, 2009) (noting several deficiencies in compliance); *cf. supra* note 140 and accompanying text.

established could lead to future disadvantages, for example, when profile-based price discrimination is introduced.<sup>167</sup>

Second, there is the same potential disciplining effect resulting from awareness of being watched as in the previous power relations. As various authors have noted, “private entities are happily and busily creating their own independent Panopticons,”<sup>168</sup> and surreptitiously, “consumers are being disciplined *by consumption itself* to obey the rules, to be ‘good’ not because it is morally preferable to being ‘bad’ but because there is no conceivable alternative to being good, other than being put outside the reach of benefits.”<sup>169</sup> The second and third dimensions of power (the indirect influencing of people’s actions)<sup>170</sup> that are at play here surely call for reflection on the legal protection of consumers, since most consumer protection mechanisms focus on the exercise of the most visible first dimension of power (directly causing the consumer to do something which she would not otherwise do).

Before we take an integrated look at the three power relations we have studied, however, it is interesting to note that, perhaps more than in the previous power relations, the consumer domain itself shows signs of resistance against the panoptic gaze of e-businesses. Successful grassroots campaigns have fought many sometimes absurdly privacy-invasive applications proposed by companies, such as Intel’s “Big Brother inside” chip,<sup>171</sup> Sony’s rootkit,<sup>172</sup> information-hungry RFID chips,<sup>173</sup> NebuAd,<sup>174</sup> and

---

167. On price discrimination, see, for example, Rajiv Dewan et al., *Product Customization and Price Competition on the Internet*, 49 MGMT. SCI. 1055 (2003) (examining the effect of product customization on price in markets with monopolies and duopolies). While price discrimination based on an incorrect profile could equally benefit the consumer, from a Rawlsian perspective of fairness and consumer protection, the possible disadvantage of being offered a higher price based on incorrect data carries more weight than the possible advantage of being offered a lower price based on incorrect data.

168. Paul M. Schwartz, *Internet Privacy and the State*, 32 CONN. L. REV. 815, 853 (2000).

169. REGINALD WHITAKER, *THE END OF PRIVACY: HOW TOTAL SURVEILLANCE IS BECOMING A REALITY* 142 (1999); see also Dholakia & Zwick, *supra* note 159, at 10 (“The superpanopticon erected by the new information entrepreneurs allows personal data to play a distinctive role in the modern STP (segmenting, targeting, and positioning) marketing process” and this “bestows consumer-friendly concepts like ‘customization’ and ‘personalization’ with the dark aura of totalitarian control.”); Koops & Leenes, *supra* note 125, at 118, 129–32.

170. See *supra* Section II.A.

171. Intel’s new Pentium III chip was proposed to include a unique Processor Serial Number, enabling identification of each and every single computer. Big Brother Inside, Latest News, April 28, 2000, <http://bigbrotherinside.org/>.

172. Sony’s rootkit consisted of spyware on music CDs that was automatically installed on computers and created security vulnerabilities for the computers. See Wikipedia, Sony BMG CD Copy Protection Scandal, [http://en.wikipedia.org/wiki/Sony\\_BMG\\_CD\\_copy\\_](http://en.wikipedia.org/wiki/Sony_BMG_CD_copy_)

Webwise.<sup>175</sup> An unorthodox, if perhaps rather desperate, anti-surveillance strategy could also be perceived in digital exhibitionism:

[u]ltraexhibitionism, we argue, is not a negation of privacy but an attempt to *reclaim some control over the externalization of information*. As such, ultraexhibitionism is to be understood as an act of resistance against the surreptitious modes of profiling, categorization, and identity definition that are being performed *by others* on the consumer whenever he or she enters the electronic “consumptionscape.”<sup>176</sup>

Thus, the active disclosure of large amounts of detailed personal information on the Internet could emerge as a new strategy of consumers to counter the risks of errors and panopticism associated with the increasing profiling power of businesses.

## VI. THE IDENTITY OF THE CITIZEN–CONSUMER–EMPLOYEE

The previous Parts have discussed three distinct power relations involving different roles of individuals. Several similarities can be observed in the developments of the domains discussed, which suggest that certain general conclusions can be drawn on technology-related shifts in power relations and their consequences for legal protection.<sup>177</sup> Before drawing such conclusions, however, we must face a new issue that emerges from the discussion. Some shifts in power relations broaden or blur contexts, particularly with the creation and interconnection of public and private databases that cross the boundaries of law enforcement, employment, and commerce. This calls into question the sectoral approach to inequality compensation. Is it enough to protect individuals in their role as citizen, employee, and consumer, or should we also seek legal protection for

protection\_scandal (last modified Feb. 5, 2010).

173. RFID chips are “smart” chips for wireless identification of objects at small distances, enabling for example tracing of consumer products. *See* CONSUMERS AGAINST SUPERMARKET PRIVACY INVASION & NUMBERING ET AL., RFID POSITION STATEMENT OF CONSUMER PRIVACY AND CIVIL LIBERTIES ORGANIZATIONS (2003), <http://www.privacyrights.org/ar/RFIDposition.htm>.

174. NebuAd was a company offering a service for behavioral advertising with ISPs transferring user communications to NebuAd for real-time profile-based advertising. Wikipedia, NebuAd, <http://en.wikipedia.org/wiki/NebuAd> (last modified May 20, 2010).

175. Webwise is a profile-based advertising system from Phorm, similar to NebuAd. *See* Antiphorm, <http://www.antiphorm.co.uk/> (last visited Feb. 10, 2010); Wikipedia, Phorm, <http://en.wikipedia.org/wiki/Phorm> (last modified Nov. 13, 2009).

176. Dholakia & Zwick, *supra* note 159, at 13.

177. *See infra* Sections VII.A & VII.B.

individuals regardless of their role? To answer these questions, we need to broaden our view and look deeper into the identity of today's individuals.

This Part will start with a description of how identity construction takes place, in a context-dependent presentation of someone's self in her various roles in everyday life. The discussion will draw upon the findings of the previous Parts to argue that the shifts in power relations of the citizen–employee–consumer result in a mixing of contexts, which could well become responsible for a digital identity crisis. Moreover, another development affecting identity construction is taking place—panopticism. These insights in identity construction of today's individuals will subsequently be of use when overall conclusions are drawn on legal protection of weak parties in the next and final Part.

#### A. ROLE-PLAYING, IDENTITY, AND SELF-DEVELOPMENT

Until this point, we have encountered three characters in search of legal protection: the citizen, the employee, and the consumer. They are embodied in a single person, manifestations of an actor playing different roles in different contexts to which different areas of the law apply—constitutional and administrative law, labor law, and contract and tort law, respectively. I purposefully use the imagery of the stage here; Erving Goffman has shown how the presentation of the self in everyday life builds on the ability to set the stage that defines the situation in which others form opinions of oneself and to act a part that conveys the most favorable impressions of oneself.<sup>178</sup> The ability to influence the conduct of others by this role-playing and stage-setting is a crucial part of social life.<sup>179</sup>

Self-presentation is equally crucial for identity-building and self-development, since the sense of self develops according to how we perceive others to perceive us. We construct our identities by anticipating how others are profiling us.<sup>180</sup> For example, Johnny believes himself to be a cool guy, not because wearing Calvin Klein intrinsically makes him cool, but because Johnny thinks his peer group will think him cool when they see him wearing Calvin Klein underpants.

In power relations between A and B, the ability of B to control the presentation of self—and therefore construct an identity—is impaired, since it is usually A rather than B who sets the stage. This is why many legal mechanisms to protect weak parties aim at enhancing their ability to control the situation, by decreasing the information asymmetry, giving them access to

---

178. ERVING GOFFMAN, *THE PRESENTATION OF SELF IN EVERYDAY LIFE* (1959).

179. *Id.*

180. WP7, *supra* note 146, at 15–17.

mechanisms for redress, or increasing their ability to make autonomous choices. Data protection is often presented as informational self-determination: the ability of data subjects to control the dissemination of information about themselves.<sup>181</sup>

The process of identity construction is dynamic and time-dependent, building a continuously adaptive narrative of “who I am.”<sup>182</sup> It is also context-specific; my sense of self is not a single, clearly definable “I,” but a complex amalgam of different “mes,” which come to the forefront in different settings.<sup>183</sup> Within the context of this Article, the fact that I am a researcher at the Tilburg Institute for Law, Technology, and Society is more relevant than my being a Dutch citizen or an online buyer of books, and consequently, I present more credentials relating to my scholarship than to my political inclinations or my customer profile at Amazon.<sup>184</sup> In other contexts, however, my role as a citizen or a consumer may be more prominent, and—just like my anticipation of the reactions to this Article influence my self-presentation and self-image as a scholar—my identity as a citizen or a consumer will be influenced by what happens to me, and by what I make happen, in those contexts.

These insights into role-playing and identity construction are presented here to illustrate a crucial point for this Article’s theme. The legal protection of weak parties in power relations is defined by the roles these parties play, and these roles are played out in separate contexts regulated by separate areas of the law. However, having observed some shifts in power relations taking place that broaden or blur contexts, we should ask whether these contexts

---

181. See *seminally*, ALAN F. WESTIN, *PRIVACY AND FREEDOM* (1967), and in the European context the Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court], Dec. 15, 1983, 1 BvR 209 (F.R.G.), *available at* <http://beck-online.beck.de/?vpath=bibdata%2fents%2flmr%2f1983%2fcont%2fLMRR.1983.0056.htm>. Whether data protection as embodied in the American fair information processing principles or the European Data Protection Directive actually effectuates informational self-determination is another matter. The competing interest of the free flow of information and services in the internal E.U. market have also set their stamp on data protection legislation, leaving the data subject with relatively toothless tools to control the flow of her personal data in today’s information economy and database nations. *See supra* note 166.

182. WP7, *supra* note 146, at 15–17.

183. On the different narratives that can be told about one’s life and how these narratives complicate the construction of a singular identity, see KAREL ČAPEK, *AN ORDINARY LIFE* (M. & R. Weatherall trans., 1936).

184. Admittedly, literature can teach important lessons about scholarly questions, so that my roles as a researcher and as a reader of fiction intermingle, making some of my Amazon profile shimmer through in this article. *See* JOHN GIBSON, *FICTION AND THE WEAVE OF LIFE* (2007); *cf.* ČAPEK, *supra* note 183.

are still sufficiently separate, or separable, to determine the role and consequent legal status of an individual in the information society.

B. A DIGITAL IDENTITY CRISIS?

The shifts in power relations in the case studies partly occur within each separate context. Both the strong and the weak parties have in some sense gained power in relation to the other. In fact, today the same person can play different roles on both sides of the power relation. The most notable of the dual roles is the “prosumer,” a person who is a consumer acting as a producer.<sup>185</sup> This also occurs in the rise of citizens participating in public policy-making<sup>186</sup> and of the self-employed worker. Some individuals, but by no means all, thus gather experience playing the role of the traditionally strong party on some occasions, which may help them when they act in their role as the traditionally weak party in other situations.

However, the shifts in power relations also have effects across contexts. In today’s technology-mediated world, the person enacting citizenship, employeeship, and consumption is becoming a digital persona living in a myriad of databases, who may have less control over the specific role she is playing in different contexts. The labor context extends to non-labor time and space, where private activities may be monitored by employers enforcing their company policy of having “good” employees.<sup>187</sup> The consumer is digitized into interesting information segments that are distributed across networks and used to build profiles.<sup>188</sup> Citizens’ activities are recorded and stored in databases regardless of whether there is a preexisting suspicion that they engage in criminal acts. Frequently, these databases are controlled by private parties such as telecom providers or airline carriers. In this way, private sector architectures for doing business are being adapted to meet public policy goals of crime and terrorism fighting.<sup>189</sup> Internet Service

---

185. See ALVIN TOFFLER, *THE THIRD WAVE* 282–305 (1980).

186. On participatory governance, see, for example Cary Coglianese, *Citizen Participation in Rulemaking: Past, Present, and Future*, 55 *DUKE L.J.* 943 (2006).

187. See *supra* Section IV.C.

188. Dholakia and Zwick state that

With privacy dispossession, the consumer most significantly loses the power over his or her *representation* as consumer in the market. Someone else’s image of what the consumer *might be* takes on a *real* existence. These synthesized representations of the consumer “self” are being distributed through information entrepreneurs to the databases of the world.

Dholakia & Zwick, *supra* note 159, at 17; see also *supra* Section V.C.

189. See *supra* Section III.D.

Providers are increasingly harnessed as nodal points to monitor and intervene in the enforcement of private and public rights and duties.<sup>190</sup>

In this world of interconnected or interconnectable databases, digital representation is slowly but surely overtaking physical presentation in face-to-face contacts: the “data double is more real tha[n] the person behind it.”<sup>191</sup> The digital persona of the citizen–employee–consumer increasingly functions as her interface in the power relations with the government–employer–producer, resulting in decisions based on data stored in databases.<sup>192</sup> These databases, however, generally do not intrinsically retain the original context of the data stored in them. When function creep leads to data being exchanged and used in other contexts, the primary context often is lost. The power and attractiveness of databases lie not only in their persistence and comprehensiveness, but also in their multifunctionality. The logic of a world that thrives on databases is therefore at odds with purpose specification and use limitation, two important principles of the data protection framework. Today, I seriously doubt that purpose specification and use limitation continue to play a substantial role in practice.

The transformation of a person playing different roles in context-rich, face-to-face situations into a person interacting in different power relations through the interface of a context-poor but potentially information-rich digital persona has important implications for self-presentation and identity construction.<sup>193</sup> Goffman describes how roles are typically played before the same or similar audiences:

---

190. Koops & Leenes, *supra* note 125, at 118 (“An interesting aspect of this Internet Panopticon is that the state shifts the responsibility of enforcement to entities in the private sector, such as Internet service providers (ISPs).”); *see also* EGBERT DOMMERING, GEVANGEN IN DE WAARNEMING. HOE DE BURGER DE COMMUNICATIEMIDDELEN OVERNAM EN ZELF OOK DE BEWAKING GING VERZORGEN (2008); Jonathan Zittrain, *Internet Points of Control*, 44 B.C. L. Rev. 653 (2003).

191. Maria Los, *Looking Into the Future: Surveillance, Globalization and the Totalitarian Potential*, in THEORIZING SURVEILLANCE 69, 86 (David Lyon ed., 2006).

192. SOLOVE, *supra* note 66; *see* Roger Clarke, *The Digital Persona and Its Application to Data Surveillance*, 10 INFO. SOC’Y 77 (1994); Los, *supra* note 191, at 87 (noting that “both actuarial calculations and data-matching procedures constantly produce real consequences for individuals represented by their ersatz doubles”).

193. Los, *supra* note 191, at 78. Los notes that the fragmented, decontextualized information, collected for many specific purposes, may acquire a multitude of completely different meanings depending on its particular compilation, re-contextualization and application. As well, because of the ramified nature of data networks, it appears practically impossible to correct erroneous or twisted information. In this context, the notion of biographical truth loses any meaning.

[d]efining social role as the enactment of rights and duties attached to a given status, we can say that a social role will involve one or more parts and that each of these different parts may be presented by the performer on a series of occasions to the *same kinds of audience* or to an *audience of the same persons*.<sup>194</sup>

How can a person present herself—or her self—as a digital person when it is almost impossible to know from the outset what her audience will be? Which social role can or should she play in her guise as a digital persona performing multiple functions in public–private database conglomerations? And can the presentation of the self be at all controlled by the person when the presentation is digitized in databases? Altogether, the shifts in power relations of the citizen–employee–consumer and the associated mixing of contexts could well become responsible for a digital identity crisis.<sup>195</sup>

### C. PANOPTICISM AND NORMALIZED IDENTITY

A digital identity crisis is, however, not the necessary outcome of the shifts in power relations studied. Besides the shattering of the person across databases, a second development is taking place: panopticism.<sup>196</sup> In all three power relations, to greater or lesser degrees, this mechanism at play is distinguished. The citizen–employee–consumer is increasingly being watched in and across the different contexts in which she acts. The awareness that any activity may be observed has a potentially self-disciplining effect, through which the person embraces society’s paradigm of normality and starts to behave accordingly.<sup>197</sup>

Panopticism also affects identity construction. Profiling is a key technology that causes shifts in all three power relations studied, because it is associated with panopticism and may, through panoptic logic, affect the freedom of persons to construct their identities.<sup>198</sup> Through the double

---

*Id.*

194. GOFFMAN, *supra* note 178, at 16 (emphasis added).

195. *Cf.* Los, *supra* note 191, at 85 (“The new logic of late-modern surveillance, typified by the data double, dehumanization of freedom and de-socialized criteria of sorting, suggests a special form of biographical uprooting, whereby for many people a caring relationship with their peripatetic, de-contextualized virtual double(s) is likely to become a major preoccupation.”).

196. *See supra* Section II.A.

197. It should be noted that awareness of the average person of being watched through data surveillance may currently be fairly low. It is expected to rise, however, with the increase of personal experiences, media stories, and growing intrusiveness of surveillance practices. *See* Los, *supra* note 191, at 77, 80–81.

198. *See supra* note 137 and surrounding text; *see also* Hildebrandt, *supra* note 68, at 305–11.

anticipation that is at work in identity building, the panoptic embracing of the “system’s” paradigm of normality has a major impact on the resulting identity. The digital personae that represent the person in today’s database-based power relations *constitute* their identity as much as they are *constituted by* the person’s self-presentation in everyday life. Where panopticism is at work, there is little difference between an imposed persona (i.e., the representing profile imposed by a counter-party) and a projected persona (i.e., the self-representing profile controlled by the person herself)<sup>199</sup>: both reflect the prevalent paradigm of how someone is supposed to behave.<sup>200</sup>

Here, we can observe the third dimension of power at work: the socially structured and culturally patterned practices of institutions of government, labor, and consumption reinforce existing imbalances in the power relations. The weak parties see no other option but to embrace their (self-)imposed normalized digital personae as constituting who they are, rather than challenging their digital personae to represent the persons they want to be.

## VII. CONCLUSIONS AND OUTLOOK

After the tour d’horizon of developments in the power relations between government–citizen, employer–employee, and business–consumer and the integrated vision of the identity of today’s citizen–employee–consumer, it is time to return to the questions posed at the outset. What technology-related shifts occur in power relations in the domains of government, labor, and commerce? And what are the consequences of these shifts for the legal protection of weak parties, in particular, for existing mechanisms of inequality compensation in the associated legal domains?

This Part begins with a summary of the shifts in power relations discussed previously. Next, a discussion follows of the consequences for legal protection: first, within the realms of criminal, labor, and consumer law, and subsequently, beyond context-specific forms of inequality compensation. The analysis highlights the importance of having a comprehensive data protection framework. This Article concludes with a sketch of two alternative directions for such a framework to protect citizens in today’s cross-context database society.

---

199. Clarke, *supra* note 192.

200. *Cf.* MANUEL CASTELLS, THE POWER OF IDENTITY 7 (1997) (“Although . . . identities can also be originated from dominant institutions, they become identities only when and if social actors internalize them, and construct their meaning around this internalization.”). Panopticism precisely has such an internalizing effect. *See also* Los, *supra* note 191.

## A. SHIFTS IN POWER RELATIONS

With the development of new technologies—in particular ICT, but also genetic applications—power relations shift in various ways. Most shifts relate to an increase in information: both the traditionally strong parties (i.e., governments, employers, business) and the traditionally weak parties (i.e., citizens, employees, consumers) use ICT, and sometimes DNA techniques, to improve their information position. These shifts do not counterbalance each other. Rather, they are asymmetrical: they involve different types of information, different situations, and—particularly in the government context, but probably also in commerce—different sub-groups that are affected. Although the weak parties can use their improved information position to resist or bypass the power exercise of the strong parties, in many cases, the strong parties can use their improved information position to exercise power even more strongly and in different ways.

Particularly in the government and employment context, the empowerment of the strong party fundamentally affects the character and scope of the power relation. Criminal law is shifting from a reactive, incidental, last-resort mechanism to a preventative, comprehensive, and primary regulatory mechanism. Since this reshaping of criminal law involves massive-scale data collection, storage, and profiling of unsuspected citizens, the nature of the government–citizen relationship is slowly changing. Citizens are treated less as *prima facie* trustworthy subjects and more as *prima facie* risk objects.

A similar development, although much smaller in scale and scope, can be seen in the employment context, where employers are now monitoring their employees on a routine basis and also increasingly in off-duty situations. A crucial consequence of the changed nature of these power relations is that contexts are broadened and become intertwined: the data involved in the “risk management” of citizens and employees are stored in interconnected or interconnectable databases. Here, the commerce sector also enters the picture, since several of these databases are outsourced to third parties. These third-party information brokers or intermediaries then fill or merge the data of such databases with that of the commercial databases. Thus, database and profiling technologies are facilitating the rise of comprehensive monitoring systems that move between and across different contexts.

An important feature of the technology-facilitated changes in the nature of power relations is the rise of “governing through crime,” which implies that the contexts of crime (and criminal law) and other sectors of society (and their associated areas of law) are increasingly overlapping. Particularly in the United States, but also visible on a smaller scale in the Netherlands, many

types of relationships are increasingly cast in risk discourse and governed by mechanisms derived from or based in criminal law. These mechanisms include, for example, crime language, offender statistics, and sanctioning policies.<sup>201</sup> Many common spaces—schools, the family, the workplace—are adopting “practices suggestive of the penal aspects of criminal justice.”<sup>202</sup> For example, the U.S. Safe Schools Act of 1994 created a “national model of crime governance for schools,” encompassing zero-tolerance policies, disciplinary violations categorized as quasi-crimes, in-school detention systems, and data collection systems for (quasi-)crime monitoring.<sup>203</sup> In divorce cases, allegations of crimes committed by the partner have emerged as a primary argument in contested child custody and property distribution proceedings.<sup>204</sup> The Anti-Drug Abuse Act of 1988 comprises a “one strike and you’re out” standard to evict tenants from public housing when she or “any guest or other person under the tenant’s control” commits a drug-related offense on- or off-premises.<sup>205</sup> It is a strict liability standard that would permit a landlord to evict a mother for drugs her daughter possessed.<sup>206</sup> The “governance through crime” mechanism here is that the “exclusionary power associated with criminal designation [is also used] to accomplish other organizational goals (like ridding schools of poor test takers or ridding public housing of waiting lists) . . . .”<sup>207</sup>

In this altering landscape of context-crossing power relations, two overarching trends stand out: the use of digital personae as a substitute for the physical persons of the weak parties in power relations, and the creation of panoptic risk-governing architectures that have a potentially self-disciplining effect on the weak parties. Some individuals from the category of traditionally weak parties may use the new technological opportunities to effectively resist the power exercised by the traditionally strong party. However, the combination of these trends implies that many, if not most, individuals within the weak-party categories face new, difficult to counter, and more diffuse, context-crossing, and subtle forms of power exercise by the strong parties.

---

201. Simon, *supra* note 69, at 221.

202. *Id.*

203. *Id.* at 214–31. Illustrative is the Ruffner Middle School (Norfolk, Virginia) mandatory uniform policy. “Students who come to school without a uniform are subject to in-school detention,” which is reported as successful in improving student behavior: “throwing objects is down 68 percent and fighting has decreased by 38 percent.” *Id.* at 225.

204. *Id.* at 192.

205. *Id.* at 194–95.

206. *Id.*

207. *Id.*

## B. CONSEQUENCES OF LEGAL PROTECTION

The shifts in power relations raise questions about the ability of current mechanisms to compensate weak parties for structural inequalities. The case studies in this Article have uncovered insufficiencies or inadequacies in current legal protection. Each legal field—criminal, constitutional, administrative, labor, and consumer law—requires adaptation to meet the new reality. This has already partly been achieved, for example, with the introduction of consumer protection rules in e-commerce legislation and the creation of guidelines for responsible monitoring of employees.<sup>208</sup>

However, much remains to be updated. For example, adequate protection against the use of new applications, such as familial DNA searching or behavioral advertising, must be devised. This is a matter of course: the law usually lags behind technological developments, and it is unsurprising that legal-protection mechanisms will eventually be adapted or created for such new developments.

The U.S. legal system is arguably better equipped than the Dutch legal system to achieve this because its reliance on case-law besides statutory law ensures that it can relatively swiftly adapt to new technological realities. Constitutional review allows modern-day re-interpretation of age-old constitutional protection provisions.<sup>209</sup> Furthermore, its statutes frequently contain open norms that can be flexibly interpreted by the courts,<sup>210</sup> and its greater reliance on the market also allows for more flexibility.

Nevertheless, the flexibility and use of open, re-interpretable norms in the U.S. approach also have drawbacks for legal protection, since the legal norms can easily and docilely follow technological and market developments rather than actively shape these developments. For example, this is visible in

---

208. See *supra* notes 113–14, 153–56 and accompanying text.

209. In the Netherlands, constitutional review by the courts is unconstitutional. GRONDWET VOOR HET KONINKRIJK DER NEDERLANDEN [GW.] [Constitution] art. 120 (Neth.). A Bill is pending to amend this (Kamerstukken II, 2001–2002, 28 331 (Neth.)), but it is dubious whether this Bill will be adopted in the foreseeable future. The Dutch constitutional system is therefore much more rigid, leading to technology-specific constitutional provisions, like the freedom of the printing press and the secrecy of telegraphy, becoming outdated without legal certainty with respect to “new” technologies like the Internet. For a comparison of the Dutch and American approaches to “digital constitutional rights,” see Bert-Jaap Koops & Marga Groothuis, *Constitutional Rights and New Technologies in the Netherlands*, in CONSTITUTIONAL RIGHTS AND NEW TECHNOLOGIES: A COMPARATIVE STUDY (R. Leenes et al. eds., 2008) and Susan W. Brenner, *Constitutional Rights and New Technologies in the United States*, in CONSTITUTIONAL RIGHTS AND NEW TECHNOLOGIES: A COMPARATIVE STUDY (R. Leenes et al. eds., 2008).

210. For example, in the requirements for interceptability of telecommunications, see *supra* Section III.B.

the erosion of privacy protection through the technology-facilitated erosion of reasonable expectations of privacy.<sup>211</sup> Technology and the market are not usually allies of weak parties, and hence, the law should have firm mechanisms in place when it is to provide protection to weak parties in the face of technology-related shifts in power relations. It remains to be seen whether, in general, the more flexible and responsive, but also more fluid and market-oriented character of the U.S. approach is better able to meet the challenges of providing legal protection in power relations than the slower and more rigid, but also more principled and paternal character of the Dutch approach.

Regardless of the specifics of legal systems and concrete changes that may need to be made in specific areas of law, two general conclusions can be drawn on legal protection of weak parties. One problem resides in the shift from a reactive, incident-driven approach to a preventative, comprehensive approach. It is most obvious in criminal justice, but it is also visible in other contexts where risk governance is gaining ground. Legal protection of weak parties in a reactive, incident-driven system tends to focus on preventing or redressing grave errors that may incidentally occur, for example, sending an innocent person to jail or having an ignorant consumer declared bankrupt after being lured into buying a high-risk financial product. When the system becomes preventative and comprehensive, however, the vulnerability does not lie solely in incidental major errors or injustices, but in frequent minor errors or injustices. Examples include wrongly blacklisted passengers being detained at airports for a few hours, employees forced to explain with occasional embarrassment what they (or their cars) were doing at dubious locations, children's rights organizations finding their websites blocked by overzealous child pornography filtering systems. Such relatively small inconveniences, with minor damage, will not be set right by legal-protection mechanisms based on the old paradigm of addressing incidental grave errors. The law should therefore be supplemented with new forms of legal protection that can prevent or redress adequately the overall functioning of comprehensive risk-governing systems. This means introducing more administrative and accountancy-type checks and balances, such as regular audits by independent supervisors monitoring the fairness of policies and practices, low-threshold complaint mechanisms with teeth to call the strong

---

211. Koops & Leenes, *supra* note 125 (finding that technology does not incorporate privacy norms and erodes reasonable expectations of privacy); Phillips, *supra* note 49, at 59 (“This reliance on the market as a policy mechanism for privacy protection reinforces and exacerbates unequal power relations between employers and employees.”).

parties to order, and liberal compensation mechanisms for people suffering inconveniences and smaller injustices.<sup>212</sup>

A second overarching problem that needs to be addressed is the gap between law in the books and law in action. This problem occurs most notably in data protection but also in employment law (e.g., the difficulty of achieving redress) and consumer law (e.g., the lack of individual access to justice through dispersion of damage). The gaps in data protection and privacy law are systemic, and in the era of databases, profiling, ubiquitous computing, and Ambient Intelligence,<sup>213</sup> law in the books has reached the limits of its powers. As a result, legal protection should not only be articulated in written law, but also in the socio-technical infrastructure itself. Both PETs and transparency enhancing technologies (TETs) must be developed that embed legal rules in present and future ubiquitous technologies.<sup>214</sup> The same may well apply to other areas of legal protection, including legal mechanisms in labor and consumer law, which are difficult to enforce in today's technology-pervaded world. "Code as law" will be required to supplement law in the books if weak parties are to be effectively protected.<sup>215</sup>

### C. BEYOND CONTEXT-SPECIFIC INEQUALITY COMPENSATION

It is insufficient to adapt legal protection, as outlined in the previous Section, only within each specific area of law where inequality-compensating protection mechanisms are found. The search for legal protection of weak parties should not be restricted to the specific context of their concrete role as citizen, employee, or consumer. On the contrary, the key challenge of updating inequality compensation in light of technology-related shifts in power relations lies in finding ways to empower individuals with means to develop themselves and to construct their identities in a technology-mediated

---

212. G.G. Fuster & P. De Hert, *PNR and Compensation: How to Bring Back the Proportionality Criterion*, in ARE YOU WHO YOU SAY YOU ARE? THE EU AND BIOMETRIC BORDERS 101, 108 (2007) ("The [compensation] mechanisms as envisaged in this paper will offer potentially many more benefits than mere individual redress. Their efficiency in generating collective benefits, however, relies ultimately on the generosity of the compensation. . . ."); Koops, *supra* note 64.

213. Ambient Intelligence refers to the concept of sensor-equipped environments that respond in real-time to the people moving around in them. *See generally* THE NEW EVERYDAY: VIEWS ON AMBIENT INTELLIGENCE (Emile Aarts & Stefano Marzano eds., 2003) (discussing Ambient Intelligence, its potential, implications and potential problems).

214. M. Hildebrandt & Bert-Jaap Koops, *The Challenges of Ambient Law and Legal Protection in the Profiling Era*, 73 MOD. L. REV. 428 (2010).

215. Employing "code as law" is easier said than done if it is to be both effective and legitimate. *See id.* This is one of the major challenges for future interdisciplinary research. *Id.*

world that obfuscates the audiences for which they play their different roles as citizens, employees, and consumers.

The main trends in the shifts in power relations are the use of context-poor digital personae and the creation of panoptic monitoring architectures.<sup>216</sup> The resulting vulnerabilities are increased risks of errors in interpretation, as well as a normalizing, self-disciplining effect of the panoptic architectures on individuals' behavior and identity construction. Both of these trends and vulnerabilities are related to the fact that myriads of personal data are stored and processed in diverse power relations, and possibly exchanged across contexts. The key issue is who is allowed to process which data for which purposes and under which conditions. In other words, data protection turns out to be the major mechanism that deals with the key issue in reducing vulnerabilities emerging through the shifts in power relations. If people are to be protected against abuses of power in the era of databases and profiling, then some form of data protection is crucial. Furthermore, this data protection should be generic rather than context-specific because the audiences of digital personae are far less clearly distinguishable than the audiences for physical, role-playing persons.

The European, general approach to data protection is more adequate in that respect than the context-specific, piecemeal approach of the United States, but the comprehensive European approach to data protection also faces considerable challenges. For instance, as discussed above, the European approach must deal with the sustainability of the purpose-specification and use-limitation principles in a database-pervaded world,<sup>217</sup> the gap between law in the books and law in action, and the consequent need to build in PETs and TETs in socio-technical architectures.<sup>218</sup>

Can a comprehensive data protection framework actually meet all these challenges, in order to provide cross-context inequality compensation to the citizen–employee–consumer, as represented by her proxy, the digital persona? In other words, can data protection be made to empower people to control their digital personae to such an extent that they can resist the abuse of power by different strong parties in diverse and opaque situations? In theory, it can. *How* it can achieve this, however, is a matter of debate. The literature seems to suggest two radically different directions for empowering persons.

---

216. *See supra* Section VII.A.

217. *Supra* Section VI.B.

218. *Supra* Section VII.B.

## D. TWO DIRECTIONS TO EMPOWER PERSONS

1. *The Orthodox View: Resistance by Data Limitation and User Control*

The first way in which persons can be empowered to resist the exercise of power by diverse strong parties in the face of ubiquitous databases and profiling, is to make the current data protection framework more effective. This approach uses a two prong strategy to empower data subjects to gain and retain a substantial level of control over their personal data in the world's myriad databases, as well as over their digital personae. The first prong requires that purpose-specification and use-limitation principles remain cornerstones of data protection, and hence, limits are set on who can access and process which personal data for which purposes. To meet the realities of today's database world, however, these limits may be less strict in terms of preventing data processing. This is then compensated by stronger requirements for making the processing, particularly if used for other purposes, more transparent to and challengeable by data subjects. The other prong is that these limits and requirements are to be enforced more effectively in practice than is the case today, particularly by using PETs and TETs.

Advocates for this approach are typically the data-protection community—a loose network of professionals and scholars aiming to develop and preserve data protection, including Data Protection Authorities and Information Commissioners, privacy advocates, and experts in the field of data security.<sup>219</sup> This approach builds on Nissenbaum's notion of "contextual integrity," which presents a context-sensitive "justificatory framework for prescribing specific restrictions on collection, use, and dissemination of information about people."<sup>220</sup> This approach explores possibilities to achieve "privacy in the clouds," where cloud computing and Web 2.0 call for identity-management systems that are under the control of users; these possibilities include technical and organizational solutions, such as open-source software, federated identity management, multiple and partial identities, audit tools, and data-centered or "sticky" policies.<sup>221</sup> The vision of

---

219. The latter is exemplified in the title of the German journal *Datenschutz und Datensicherheit* [Data Protection and Data Security], the authorship and readership of which constitute a significant part of the continental data protection community.

220. Helen Nissenbaum, *Privacy as Contextual Integrity*, 79 WASH. L. REV. 119, 155 (2004); see also Avner Levin & Patricia Sánchez Abril, *Two Notions of Privacy Online*, 11 VAND. J. ENT. & TECH. L. 1001, 1051 (2009) (arguing that even when people voluntarily disclose personal information on the web such as on social network sites, a reasonable expectation of privacy remains as long as the information remains inside the network in which it was disclosed).

221. Ann Cavoukian, *Privacy in the Clouds*, 1 IDENTITY INFO. SOC'Y 89 (2008); see also Jan

Ambient Intelligence, in this approach, is to be realized by architecture-embedded Ambient Law in the form of PETs and TETs.<sup>222</sup> In short, this approach holds that by harnessing technology through technology, a new and more effective generation of context-sensitive data protection can be achieved.

2. *The Radical View: Resistance by Data Proliferation and Looking in Return*

The second approach to empowerment is more unorthodox and radically different. The power of panoptic architectures can be resisted, in this approach, by beating the observers at their own game. This resistance can take a number of manifestations. For example, using the method of the “Jam Echelon Day,” in which the Anglo-Saxon intelligence snooping network Echelon was to be clogged by including in each e-mail message a signature with fifty “red-flag” words,<sup>223</sup> panoptic observers can be overwhelmed by creating such enormous haystacks of personal data that the needles are hopelessly lost. Another manifestation of this type of resistance is exhibitionism—disclosing a complete digital persona online that is fully visible in order to preempt others from constructing a digital persona for you. By anticipating imposed personae and exhibiting “adult” versions of their projected persona, people can retain a sense of control in the construction of their identity.<sup>224</sup>

Unorthodox as this approach may be, it aligns with other developments in the network society, such as crowdsourcing, viral marketing, free

Camenisch et al., Privacy and Identity Management for Everyone, in PROCEEDINGS OF THE 2005 WORKSHOP ON DIGITAL IDENTITY MANAGEMENT 20–27 (2005); Marco Casassa Mont et al., *Towards Accountable Management of Identity and Privacy: Sticky Policies and Enforceable Tracing Services*, in PROCEEDINGS OF THE 14TH INTERNATIONAL WORKSHOP ON DATABASE AND EXPERT SYSTEMS APPLICATIONS 377 (2003); PRIME-Privacy-Enhanced Identity Management in Europe, <http://www.prime-project.eu> (last visited Sept. 25, 2009).

222. Mireille Hildebrandt, *A Vision of Ambient Law*, in REGULATING TECHNOLOGIES: LEGAL FUTURES, REGULATORY FRAMES AND TECHNOLOGICAL FIXES 175, 188–89 (Roger Brownsword & Karen Yeung eds., 2008).

223. An example of an Echelon-jam-generated e-mail signature is: “ATF DOD WACO RUBY RIDGE OKC OKLAHOMA CITY MILITIA GUN HANDGUN MILGOV ASSAULT RIFLE TERRORISM BOMB DRUG KORESH PROMIS MOSSAD NASA MI5 ONI CID AK47 M16 C4 MALCOLM X REVOLUTION CHEROKEE HILLARY BILL CLINTON GORE GEORGE BUSH WACKENHUT TERRORIST.” Chris Oakes, *Monitor This, Echelon*, WIRED, Oct. 22, 1999, <http://www.wired.com/print/politics/law/news/1999/10/32039/>.

224. Dholakia & Zwick, *supra* note 159, at 1 (noting that “exhibitionism and voyeurism seem to offer new tools for consumer resistance against the electronic surveillance systems in networked markets and are inextricably interwoven with consumers’ desire for control over their information”); *see also supra* note 176 and accompanying text.

distribution of goods as a new business model, and a “pirate’s” approach to information dissemination,<sup>225</sup> all of which involve an extreme proliferation of data and a re-evaluation of the value attached to those data.

Besides data proliferation, the most pertinent manifestation of this type of resistance is the vision of David Brin, who proposes the strategy of looking in return.<sup>226</sup> In Brin’s view, checks and balances in a panoptic surveillance society are to be found in monitoring the monitors, not only by independent supervisors, but also and more importantly, by the people monitored: “we may not be able to eliminate the intrusive glare shining on citizens of the next century, but the glare just might be rendered harmless through the application of more light aimed in the other direction.”<sup>227</sup> By increasing transparency on all sides, a bottom-up coalition of amateur watchers will scrutinize the monitoring practices of the powerful: “the cameras *are* coming. You can rail against them, shaking your fist in futile rage at all the hovering lenses. Or you can join a committee of six billion neighbors to control the pesky things, making each one an extension of your eyes.”<sup>228</sup> In this radical view, the glaring light of ubiquitous transparency is not incompatible with privacy; on the contrary, it safeguards privacy by its unique power to hold accountable those who violate privacy.<sup>229</sup> The power of knowledge may be wielded by data-collecting, strong parties, but abuse of power will immediately be brought to light by the power of numbers of weak parties who can scrutinize all the strong parties’ actions.

#### E. CONCLUSION: NO MIDDLE WAY

This Article argues that technology-related shifts in power relations call for revision of the legal protection of weak parties, in particular, of mechanisms of inequality compensation. Part of this should be achieved by updating existing mechanisms in the associated legal domains, by critically reviewing existing provisions in criminal, administrative, labor, and consumer

---

225. *See, e.g.*, CHRIS ANDERSON, *FREE: THE FUTURE OF A RADICAL PRICE* 3 (2009) (discussing a new business model in which many people are “making lots of money charging nothing”); CHARLES LEADBEATER & PAUL MILLER, *THE PRO-AM REVOLUTION: HOW ENTHUSIASTS ARE CHANGING OUR SOCIETY AND ECONOMY* (2004) (discussing the counter-trend of Pro-Ams, “innovative, committed and networked amateurs working to professional standards,” and their impact on society); MATT MASON, *THE PIRATE’S DILEMMA: HOW YOUTH CULTURE IS REINVENTING CAPITALISM* (2008) (describing how youth culture and trends have influenced society).

226. DAVID BRIN, *THE TRANSPARENT SOCIETY: WILL TECHNOLOGY FORCE US TO CHOOSE BETWEEN PRIVACY AND FREEDOM?* (1998).

227. *Id.* at 23.

228. *Id.* at 333.

229. *Id.* at 334.

law, to assess their ability to protect weak parties in light of shifts in power relations that empower the strong party, particularly when power is exercised in new ways.<sup>230</sup> However, sector-specific legal protection is not enough. We also need a comprehensive approach to protect individuals in a world where they interact with different strong parties through digital personae across contexts and in pervasively monitoring architectures. Such a comprehensive approach is most likely to be found in data protection: enforceable rules that regulate who can process which data for which purposes under which conditions.

However, it is unclear how empowerment of data subjects can best be achieved to meet the reality of today's database and profiling era. The dominant, orthodox strand in the literature favors the current European approach to data protection, focusing on data limitation and user control, and suggesting a concerted attempt to make this work in practice by implementing context-sensitive PETs and TETs. In contrast, a subsidiary, radical strand in the literature favors user-generated data maximization and counter-surveillance strategies based on transparency to resist the exercise of panoptic power.

Both directions for the next generation of data-protection frameworks have potential, and either will be a challenge to achieve. However, there is no middle path: the approaches of data limitation and data proliferation are incompatible with one another. We will have to choose between the orthodox and the radical approach. And while the debate continues, digital personae and panoptic architectures will continue to proliferate, playing into the hands of the powerful, and the citizen–employee–consumer of the database era will face an ever harder job to resist the exercise of those powers. A consistent approach to achieve effective data protection must be decided upon soon.

If the orthodox way does not prove successful in the coming years, then, perhaps, we should collectively shift to the radical way.

### VIII. POSTSCRIPT: UMBERTO ECO'S ANOPTICON

The radical way will, like all radical suggestions, seem far-fetched and fraught with questions of feasibility. I will leave aside discussing these questions here, as they require considerable further research, and end instead with a visionary metaphor for the radical way, which can serve as a welcome

---

230. See *supra* Section VII.B.

contribution to the post-Foucauldian literature on panopticism and how to resist it: Umberto Eco's Anopticon.<sup>231</sup>

The Anopticon is a hexagonal building which effectuates "the principle of 'being able to be seen by everyone without seeing anyone.'" <sup>232</sup> The Anopticon's subject is a prison guard who lives in a closed, central hexagonal room, illuminated by a few conical embrasures which allow some light to shine through from above, but which do not permit the prison guard to see anything other than a small circular portion of sky. Around the prison guard's room are the prison's corridors where the prisoners can walk around freely.

From these corridors encircling the central room, the prisoners can watch the prison guard through conical embrasures, in such a way that the prison guard can not know when he is being observed, nor by whom. In fact,

[t]he Anopticon does not allow the prison guard to have any control over the rest of the prison: he can not surveil the prisoners, he can not prevent their escape, he can not even know if there are any prisoners left nor whether anyone is watching him, and, supposing that someone were watching him, the prison guard is not capable of knowing whether it is a prisoner or an occasional visitor of this *machine-à-laisser-faire* (see also the married machines and *The virgin dressed by her other spouses*).<sup>233</sup>

As in the radical view of a maximally transparent society, Umberto Eco's Anopticon provides a new answer to the traditional question "Quis custodiet custodes?" It is we, the watched, who should watch the watchers.

---

231. Umberto Eco, *L'Anopticon*, in *IL SECONDO DIARIO MINIMO* 176 (1992).

232. *Id.* (translation by Bert-Jaap Koops).

233. *Id.* (translation by Bert-Jaap Koops). *The virgin dressed by her other spouses* is an inverse reference to Marcel Duchamp, *The Bride Stripped Bare by Her Bachelors, Even (The Large Glass)* (sculpture) (1923), which recalls associations of the work's machine-like appearance and of its Panopticon-suggestive nickname, "The Large Glass."

1036

**BERKELEY TECHNOLOGY LAW JOURNAL** [Vol. 25:973

# WHAT PAYMENT INTERMEDIARIES ARE DOING ABOUT ONLINE LIABILITY AND WHY IT MATTERS

Mark MacCarthy<sup>†</sup>

## TABLE OF CONTENTS

I.	<b>INTRODUCTION</b> .....	1038
II.	<b>INDIRECT INTERMEDIARY LIABILITY</b> .....	1043
	A. INDIRECT LIABILITY REGIMES.....	1043
	B. A FRAMEWORK FOR ANALYSIS .....	1046
	1. <i>Market Failure Analysis</i> .....	1047
	2. <i>Cost–Benefit Analysis</i> .....	1051
	3. <i>Equity Analysis</i> .....	1055
III.	<b>APPLYING THE FRAMEWORK TO PAYMENT INTERMEDIARIES</b> .....	1059
	A. INTERNET GAMBLING LEGISLATION.....	1062
	1. <i>Implementation Challenges with the Internet Gambling Act</i> .....	1066
	2. <i>Internet Gambling Assessment</i> .....	1068
	B. CHILD PORNOGRAPHY, CONTROLLED SUBSTANCES, AND ONLINE TOBACCO .....	1074
	1. <i>Child Pornography</i> .....	1074
	2. <i>Controlled Substances</i> .....	1078
	3. <i>Online Tobacco</i> .....	1081
	4. <i>Assessment of Child Pornography, Controlled Substances, and Online Tobacco</i> .....	1083
	C. ONLINE COPYRIGHT INFRINGEMENT .....	1087
	1. <i>Legal Context for Intermediary Liability in Copyright Infringement</i> .....	1087
	2. <i>Payment System Complaint Program</i> .....	1090
	3. <i>Allofmp3.com</i> .....	1092
	4. <i>Assessment of Payment System Actions on Online Copyright Infringement</i> .....	1095
IV.	<b>INTERNET GOVERNANCE</b> .....	1098
	A. INTERNET EXCEPTIONALISM.....	1099
	1. <i>The Original Version</i> .....	1099
	2. <i>Critique of Internet Exceptionalism</i> .....	1103

---

© 2010 Mark MacCarthy

<sup>†</sup> Adjunct Professor in the Communications Culture and Technology Program at Georgetown University. Formerly a Senior Vice President for Public Policy at Visa Inc.

B.	PAYMENT SYSTEMS AND THE BORDERED INTERNET .....	1108
C.	INTERNATIONALISM.....	1115
V.	CONCLUSION .....	1119

## I. INTRODUCTION

In the mid-1990s, commentators began debating the best way for governments to react to the development of the Internet as a global communications medium. Internet exceptionalists argued that the borderless nature of this new medium meant that the application of local law to online activities would create insoluble conflicts of law. The exceptionalists believed that as the Internet grew, reliance on local governments to set rules for the new online world would not scale well. Their alternative was the notion of cyberspace as a separate place that should be ruled by norms developed by self-governing communities of users.<sup>1</sup>

Critics of the exceptionalist view responded with a vision of a bordered Internet where local governments could apply local law.<sup>2</sup> In this view, cyberspace is not a separate place. It is simply a communications network that links real people in real communities with other people in different jurisdictions. Governments can regulate activity on this new communications network in many different ways, including by relying on the local operations of global intermediaries. Global intermediaries are the internet service providers (ISPs), payment systems, search engines, auction sites, and other platform and application providers that provide the infrastructure necessary for internet activity. Although they are often global in character, they also have local operations subject to local government control. According to critics of the exceptionalist view, governments have the right and the obligation to use this regulatory power over intermediaries to protect their citizens from harm.<sup>3</sup> Conflicts that might arise from this regulatory activity can be resolved through the normal mechanisms governments use to resolve conflict of law questions.<sup>4</sup> Governments generally followed the advice of the proponents of regulation, not the regulatory skeptics.<sup>5</sup> And despite some set-

---

1. See, e.g., David R. Johnson & David Post, *Law and Borders—The Rise of Law in Cyberspace*, 48 STAN. L. REV. 1367, 1387–92 (1996).

2. E.g., Jack L. Goldsmith, *Against Cyberanarchy*, 65 U. CHI. L. REV. 1199 (1998).

3. See *id.* at 1238–39.

4. *Id.* at 1200–01 (arguing that “regulation of cyberspace is feasible and legitimate from the perspective of jurisdiction and choice of law”).

5. The U.S. exception is § 230 of the Telecommunications Act of 1996 which immunizes many internet actors from liability in many contexts for the illegal activity of their users. 47 U.S.C. § 230(c) (2006).

backs in First Amendment cases,<sup>6</sup> regulators have continued a steady march toward controlling the Internet by regulating intermediaries.<sup>7</sup> Some legal scholars argue that government reliance on intermediaries to control unlawful behavior on the Internet is justified because putting the enforcement burden on intermediaries is the least expensive way for governments to effectively assert jurisdiction.<sup>8</sup> The key rationale is that governments cannot easily find wrongdoers on the Internet, but intermediaries can. They are in the best position to monitor their own systems. As Mann and Belzley put it, they are the “least-cost avoider.”<sup>9</sup>

The defenders of local government jurisdiction over the Internet often rely on historical analogies to buttress their case that local control is inevitable and desirable. Debra Spar developed the thesis that society’s reaction to new technologies follows a sequence of innovation, commercial exploitation, creative anarchy, and then government rules.<sup>10</sup> The four stages progress predictably: in the innovative stage, a new technology is developed; in the second stage, it is used in commercial ventures; in the third stage, there is a tension between the anarchist impulse and the need for commercial order and stability; and in the final stage, society reaches out to regulate the now mature technology to create and maintain the needed stability.<sup>11</sup> The development of radio is the standard example of this pattern. Radio’s initial pioneers thought its ability to wirelessly broadcast information from one point to many made government control difficult and unnecessary.<sup>12</sup> But later commercial enterprises actively sought out government regulation in order to end the chaos on the airwaves that prevented broadcasters from

6. *See, e.g.*, *Reno v. ACLU*, 521 U.S. 844, 885 (1997) (“The interest in encouraging freedom of expression in a democratic society outweighs any theoretical but unproven benefit of censorship.”); *Ctr. for Democracy & Tech. v. Pappert*, 337 F. Supp. 2d 606, 665 (E.D. Pa. 2004) (finding that a statute requiring ISPs to block access to websites displaying child pornography violated the First Amendment).

7. *See generally* JACK GOLDSMITH & TIM WU, *WHO CONTROLS THE INTERNET?: ILLUSIONS OF A BORDERLESS WORLD* (2006) (citing many examples of this trend). This Article documents further examples in which payment systems were induced by laws, regulations, pressure, and notions of corporate responsibility to take actions to control the illegal online behavior of people using their systems.

8. *See, e.g.*, Ronald J. Mann & Seth R. Belzley, *The Promise of Internet Intermediary Liability*, 47 WM. & MARY L. REV. 239, 249–50 (2005).

9. *Id.* at 249.

10. DEBORA L. SPAR, *RULING THE WAVES: CYCLES OF DISCOVERY, CHAOS, AND WEALTH FROM THE COMPASS TO THE INTERNET* 11–22 (2001).

11. *Id.*; *see also* Mann & Belzley, *supra* note 8, at 243–44; GOLDSMITH & WU, *supra* note 7, at 124 (relying on Spar’s work).

12. *See generally* SPAR, *supra* note 10, at 124–90 (describing the history of radio technology development).

reaching their intended audience.<sup>13</sup> Applying Spar's analysis here, the Internet is somewhere between stage three and stage four, where we can expect further regulation of internet activity under the watchful eye of government. The historical example demonstrates that although every new technology is thought to be outside the jurisdiction of government, this belief usually gives way in time to the realities of government control.

In the case of the Internet, the advent of government control prompted many observers to think the internet exceptionalists had been routed.<sup>14</sup> However, internet exceptionalism is still a widely held viewpoint,<sup>15</sup> and the notion that government control of cyberspace is both impossible and illegitimate still motivates much discussion of internet policy.<sup>16</sup> Moreover, the initial legislative expression of internet exceptionalism—§ 230 of the 1996 Telecommunications Act—is still on the books. This section provides a safe harbor from indirect liability for what might be called pure internet intermediaries—those entities providing internet access service or online services.<sup>17</sup> Despite a growing call to revisit this immunity,<sup>18</sup> it has been

---

13. *Id.* at 171–72.

14. See GOLDSMITH & WU, *supra* note 7, at 14 (asserting that “notions of a self-governing cyberspace are largely discredited”).

15. See generally DAVID G. POST, IN SEARCH OF JEFFERSON'S MOOSE (David Kairys ed., 2009) [hereinafter POST, IN SEARCH OF JEFFERSON'S MOOSE] (demonstrating an elegant take on internet exceptionalism). The heart of the response to Goldsmith is that scale matters and that while it is physically possible and permissible under current “settled” law of cross-border jurisprudence, it is not “workable” to subject all websites to perhaps hundreds of different and possibly conflicting jurisdictions. See David G. Post, *Against “Against Cyberanarchy,”* 17 BERKELEY TECH. L.J. 1365, 1384 (2002) [hereinafter Post, *Against “Against Cyberanarchy”*].

16. See H. Brian Holland, *In Defense of Online Intermediary Immunity: Facilitating Communities of Modified Exceptionalism*, 56 U. KAN. L. REV. 369, 397 (2007). Holland's version of modified exceptionalism is closely connected with the legal principle that online intermediaries are not liable for third party conduct. He asserts that the immunity from liability created by § 230 of the Communications Decency Act

helps to effectuate a modified form of exceptionalism by moderating the imposition of external legal norms so as to permit a limited range of choices—bounded, at least, by criminal law, intellectual property law and contract law—in which the online community is free to create its own norms and rules of conduct.

*Id.* at 397.

17. 47 U.S.C. § 230(c)(1) (2006) (“No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.”). The interpretation of this provision is quite broad. See, e.g., *Zeran v. Am. Online, Inc.*, 129 F.3d 327, 330–31 (4th Cir. 1997) (finding that plaintiff's tort claims of defamation were preempted by § 230). The immunity does not extend to criminal law, contract law, or intellectual property law. 47 U.S.C. § 230(e)(1)–(4) (2006).

18. See, e.g., Doug Lichtman & Eric Posner, *Holding Internet Service Providers Accountable*, 14 U. CHI. SUP. CT. ECON. REV. 221 (2006).

extended several times. The internet gambling law, which creates liability for traditional intermediaries such as payment systems, contains a limitation on liability for pure internet intermediaries.<sup>19</sup> Similarly, the recently passed online pharmacy law exempts pure internet intermediaries from a general duty to avoid aiding or abetting unauthorized internet sales of controlled substances.<sup>20</sup> The adoption of these provisions in recent laws might be merely § 230 on automatic pilot, but more likely, some version of internet exceptionalism is at work in these legislative distinctions.

Given that parties on both sides of the internet exceptionalism debate can point to legislative manifestations of their arguments, it is clear that the question is far from fully settled. For these reasons, it is worth revisiting the intermediary liability debate in light of the experience that internet intermediaries have had in controlling internet content.

If the internet exceptionalists rested their case on the literal impossibility of extending local law to cyberspace, then there is not much left to their argument. A “bordered Internet” where intermediaries try to control behavior prohibited by local law is becoming a reality. Most internet intermediaries have explicit policies generally prohibiting them from aiding illegal activities.<sup>21</sup> These general policies are supplemented with specific policies and procedures designed to prevent the use of these systems for specific illegal activities. This Article focuses on the traditional payment intermediaries—payment card companies such as Visa, MasterCard, and American Express—as an instructive category of intermediary platforms.

19. 31 U.S.C. § 5365(c) (2006).

20. Ryan Haight Online Pharmacy Consumer Protection Act of 2008, Pub. L. No. 110-425, § (h)(3)(A)(iii), 122 Stat. 4829–30.

21. Participants in Google’s advertising programs “shall not, and shall not authorize any party to . . . advertise anything illegal or engage in any illegal or fraudulent business practice.” Google Inc. Advertising Program Terms 4 (Aug. 22, 2006), *available at* <https://adwords.google.com/select/tsandcsfinder>. MasterCard has rules for both merchants and their acquiring banks: “A Merchant must not submit for payment into interchange . . . and an Acquirer must not accept from a Merchant for submission into interchange, any Transaction that is illegal.” MASTERCARD, MASTERCARD RULES 5.9.7 (2008), *available at* [http://www.merchantcouncil.org/merchant-account/downloads/mastercard/MasterCard\\_Rules\\_5\\_08.pdf](http://www.merchantcouncil.org/merchant-account/downloads/mastercard/MasterCard_Rules_5_08.pdf). MasterCard prohibits its issuing banks from engaging in illegal transactions. *Id.* at 3.8.4. Visa has similar rules. For example: “A Merchant Agreement must specify that a Merchant must not knowingly submit, and an Acquirer must not knowingly accept from a Merchant, for submission into the Visa payment system, any Transaction that is illegal or that the Merchant should have known was illegal.” VISA, VISA INTERNATIONAL OPERATING REGULATIONS 4.1.B.1.c (2008), *available at* <http://usa.visa.com/download/merchants/visa-international-operating-regulations.pdf>. Visa’s regulations also specify acquirer penalties for merchants engaging in illegal cross-border transactions. *Id.* at 1.6.D.16.

Developments over the last several years conclusively demonstrate that these payment intermediaries can control specific illegal activities on the Internet.

Thus, the debate over internet exceptionalism has rapidly shifted from the “nature” of the Internet as something intrinsically beyond the control of governments to a problem of choice.<sup>22</sup> Intermediaries can control illegal behavior on the Internet and governments can control intermediaries, but should they? And if governments should exert control over intermediaries, how should the global legal order be structured to accommodate their role?

The experiences of traditional payment intermediaries in acting to limit internet gambling, child pornography, controlled substances, online tobacco, and online copyright infringement provide a useful lens to address these questions. These payment intermediary practices suggest several lessons. First, regardless of the precise legal liabilities, intermediaries have a general responsibility to keep their systems free of illegal transactions and they are taking steps to satisfy that obligation. Second, the decision to impose legal responsibilities on intermediaries should not be based on the least cost avoider principle. Assessments of intermediary liability must take into account market failures, as well as an analysis of costs, benefits, and equities. Third, exceptionalism is not the right framework for internet governance because intermediaries should not defer to the judgments of self-governing communities of internet users when these judgments conflict with local law. Fourth, the exceptionalists are right that a “bordered Internet” will not scale well. The experience of traditional payment systems points towards international harmonization. If governments are going to use intermediaries to regulate the Internet, they need to coordinate their laws to make that role possible.

Part II of this Article outlines a framework for analyzing intermediary liability. This framework calls for a thorough analysis, including an assessment of market failure and an examination of the costs, benefits, and equities involved in imposing intermediary liability. Part III applies this framework to the policies and practices of payment intermediaries in the areas of internet gambling, child pornography, controlled substances, online tobacco, and online copyright infringement. Revisiting the internet governance question, Part IV rejects the vision of a bordered Internet on the exceptionalist ground that it will not scale well. But despite the exceptionalist view, some form of internationalism is the best way forward.

---

22. See Holland, *supra* note 16, at 376–77 (“In this context, exceptionalism became an objective to be pursued and protected as a matter of choice, rather than a natural state.”).

## II. INDIRECT INTERMEDIARY LIABILITY

### A. INDIRECT LIABILITY REGIMES

This Section distinguishes indirect liability regimes from other ways of imposing obligations on intermediaries. The basic contrast is between legal schemes that hold intermediaries responsible solely for their own bad behavior versus those that also make them responsible for the illegal behavior of people using their platforms. Since these indirect liability regimes are often controversial both in principle and in practice, this Section also develops a framework for analyzing them. It calls for a multipart analysis of market failure, costs, benefits, and equities before the imposition of intermediary liability.

Most legal regimes impose direct liability. In contrast, an indirect liability regime holds a person responsible for the wrongs committed by another. There are usually several parties involved in an indirect liability regime: the bad actor, the wronged party, and a third party. The bad actor is the person directly involved in causing the harm to the wronged party. A third party, neither the bad actor nor the wronged party, is assigned responsibility to prevent the harmful conduct of the bad actor or to compensate the wronged party for the harm. In the case of copyright infringement, for example, the bad actor would be the infringer, the wronged party would be the record company that owned the music copyrights, and the third party would be an ISP or a payment system that facilitates the infringement.<sup>23</sup>

Indirect liability can be imposed through a variety of legal mechanisms.<sup>24</sup> In a tort damages regime, a third party must pay for harms caused by others either on a strict liability or negligence basis. Employer liability for the harms

---

23. Indirect liability is not the same as holding a person responsible for the external negative effects of his own actions, but it bears a resemblance. With a negative externality, a person engages in some action, such as cattle-raising or industrial production, and the spillover effects of that action harm some other party who is not directly involved in the activity. Cattle-raising might hurt the neighboring farmers and industrial pollution might harm innocent parties far and near. In this case, the responsible person's actions are directly causing the harm. He is the bad actor. In the indirect liability case, the responsible person is in some fashion involved in the creation or maintenance of the harm and is also in a position to reduce the harm, either by detecting and deterring it or by reducing his own activity that contributes to it. But he is not the bad actor who is directly bringing about the harm. In a case of indirect copyright infringement, for example, the bad actor is the infringer, while the third party would be some intermediary, an ISP or a payment system, whose activity or service allows the bad actor to commit the infringement.

24. See, e.g., Douglas Lichtman,  *Holding Internet Service Providers Accountable*, 27 REG. 54, 59 (2004) (proposing that ISP liability for cyber security issues could be established in a regime of "negligence or strict liability, whether it is best implemented by statute or via gradual common law development"); Mann & Belzley, *supra* note 8, at 269–72 (suggesting three possible regimes: traditional tort regime, a takedown requirement, and a hot list).

caused by employees is a standard example. Statutes or court decisions can impose liability for monetary damages for specific types of harms. Additionally, statutes can require third parties to take certain specific steps to prevent harms to others. A wide variety of legal structures can be usefully viewed as indirect liability regimes, including data security and notification requirements,<sup>25</sup> some privacy requirements,<sup>26</sup> and some consumer protection requirements imposed on financial service companies.<sup>27</sup>

---

25. Data security and notification statutes can be conceptualized as third party liability regimes which impose preventive and mitigation duties. The duty for a data controller to secure personal information under his control is designed to protect the data subject from potential wrongs perpetrated by data thieves. The duty to notify a data subject of a security breach when there is a reasonable likelihood of identity theft or other harm is intended to provide the data subject with information that he can use to protect himself from these harms. A further example of an indirect liability scheme in the data security area is the Minnesota cost recovery statute that holds merchants liable for the costs associated with a breach when they failed to take specific precautions that are part of an industry data security standard. MINN. STAT. § 325E.64 (2009). Minnesota's law specifies that

[n]o person or entity conducting business in Minnesota that accepts an access device in connection with a transaction shall retain the card security code data, the PIN verification code number, or the full contents of any track of magnetic stripe data, subsequent to the authorization of the transaction or in the case of a PIN debit transaction, subsequent to 48 hours after authorization of the transaction.

MINN. STAT. § 325E.64(2). This precaution of not saving authentication codes is based on the PCI DSS industry standard, Requirement 3.2. PCI SECURITY STANDARDS COUNCIL, PAYMENT CARD INDUSTRY (PCI) DATA SECURITY STANDARD: REQUIREMENTS AND SECURITY ASSESSMENT PROCEDURES VERSION 1.2.1 22 (July 2009), [https://www.pcisecuritystandards.org/security\\_standards/pci\\_dss\\_download.html](https://www.pcisecuritystandards.org/security_standards/pci_dss_download.html) ("Do not store sensitive authentication data after authentication . . ."). The law goes on to state,

Whenever there is a breach of the security of the system of a person or entity that has violated this section, or that person's or entity's service provider, that person or entity shall reimburse the financial institution that issued any access devices affected by the breach for the costs of reasonable actions undertaken by the financial institution as a result of the breach in order to protect the information of its cardholders or to continue to provide services to cardholders . . . .

MINN. STAT. § 325E.64(3).

26. Some privacy requirements can also be thought of as third party liability regimes. Data controllers have a duty to protect the accuracy and integrity of the personal information under their control (e.g., by making sure that it is up to date and current, and by responding to data subject complaints of inaccuracy). This duty protects data subjects from harm by third parties who obtain this information from data controllers and use it for eligibility decisions (such as employment, credit or insurance).

27. Some consumer protection requirements in the financial services industry are also usefully viewed as third party liability regimes. Financial institutions participating in the provision of payment card services are required under federal law to protect cardholders in various ways: they must investigate and promptly correct billing errors that consumers allege have occurred in connection with their accounts; consumers are eligible to maintain against a

The current United States regime for intermediaries depends on the nature of the intermediary and the legal context. ISPs and some others enjoy substantial immunity.<sup>28</sup> The Communications Decency Act of 1996 exempted “interactive computer service providers” from certain kinds of third party liability by determining that they are not “the publisher or speaker of any information provided by another information content provider.”<sup>29</sup> The Digital Millennium Copyright Act of 1998 (DMCA) bars indirect copyright liability for ISPs who are acting only as a conduit and limits liability for web hosting and other service providers if they follow a prescribed notice-and-takedown procedure.<sup>30</sup> The recent internet gambling and online pharmacy laws continue this tradition of immunity.<sup>31</sup>

Payment intermediaries are subject to an indirect liability regime by the provisions of the Unlawful Internet Gambling Enforcement Act (UIGEA).<sup>32</sup> Similarly, the new online pharmacy law might subject them to specific aiding

---

creditor many of the same claims that they might assert against a merchant in connection with the purchase of defective or otherwise unsatisfactory goods and services; and the law limits a consumer’s liability for unauthorized use of payment cards. *See* 15 U.S.C. § 1601 (2006) (regulating credit cards); 15 U.S.C. § 1693 (2006) (regulating debit cards). These regulations oblige financial institutions to step in to protect cardholders against harms such as improper billing, fraud or non-delivery of goods by merchants who are linked together with cardholders in a payment system, and oblige financial institutions to protect cardholders from financial harms by fraudsters in connection with the fraudulent uses of payment cards. *See* 15 U.S.C. § 1601; 15 U.S.C. § 1693. The Truth in Lending Act (TILA), Pub. L. No. 90-321, 82 Stat. 146 (codified as amended at 15 U.S.C. § 1601 (2006)), was originally passed by Congress in 1968. Major amendments to the TILA were made by the Fair Credit Billing Act of 1974, Pub. L. No. 93-495, 88 Stat. 1511 (codified at 15 U.S.C. § 1666 (2006)), the Consumer Leasing Act of 1976, Pub. L. No. 94-240, 90 Stat. 257 (codified as amended at 15 U.S.C. § 1667 (2006)), and the Truth in Lending Simplification and Reform Act of 1980, Pub. L. No. 96-221, 94 Stat. 132. The Board of Governors of the Federal Reserve System implemented these requirements through Regulation Z. The implementation through Regulation Z is found in 12 C.F.R. § 226.12 (2009). The Electronic Fund Transfer Act, Pub. L. No. 95-630, 92 Stat. 3728 (codified at 15 U.S.C. § 1693 (2006)), was passed by Congress in 1978. The Board of Governors of the Federal Reserve System implemented these protections through Regulation E. 12 C.F.R. § 205 (2009). Regulation E does not provide redress to a consumer who has purchased allegedly defective goods or services using a debit card. *Id.*

28. 47 U.S.C. § 230(c)(1) (2006); *see* Holland, *supra* note 16, at 373–76 (discussing the extension of this immunity).

29. 47 U.S.C. § 230(c)(1).

30. 17 U.S.C. § 512(a), (c)–(d) (2006).

31. Unlawful Internet Gambling Enforcement Act of 2006, Pub. L. No. 109-347, 120 Stat. 1884 (codified at 31 U.S.C. §§ 5361–5367 (2006)); Ryan Haight Online Pharmacy Consumer Protection Act of 2008, Pub. L. No. 110-425 §§ (h)(1)(B), (h)(2)(C), 122 Stat. 4820 (codified in 21 U.S.C. §§ 829, 802). This potential liability is discussed *infra* Sections III.A–III.B.

32. Unlawful Internet Gambling Enforcement Act of 2006, Pub. L. No. 109-347, 120 Stat. 1884 (codified at 31 U.S.C. §§ 5361–5367 (2006)).

or abetting liability for online sales of controlled substances.<sup>33</sup> But the Ninth Circuit has held that payment intermediaries are not liable for copyright infringement on their systems.<sup>34</sup>

Online markets appear to be subject to some degree of indirect liability for the sale of counterfeit goods. The district court in *Tiffany (NJ) Inc. v. eBay, Inc.* held that eBay has some responsibilities under trademark law to avoid providing its services to sellers when it knows or has reason to know of trademark infringement by those sellers.<sup>35</sup> The online auction service satisfied that responsibility by implementing a series of measures including an effective voluntary notice-and-takedown system.<sup>36</sup> That responsibility did not extend, however, to a positive duty to monitor its auction site and preemptively remove possibly infringing listings.<sup>37</sup> The e-Fencing legislation, under consideration in the House of Representatives in 2009, would extend indirect liability for the sale of stolen goods to online markets.<sup>38</sup>

#### B. A FRAMEWORK FOR ANALYSIS

Indirect liability holds a party responsible for wrongs committed by another person. Why should there be any such rule? Why not simply hold the bad actor responsible? The economic analysis of indirect liability attempts to answer this question using some standard economic tools and concepts.<sup>39</sup> A

---

33. Ryan Haight Online Pharmacy Consumer Protection Act of 2008, Pub. L. No. 110-425 §§ (h)(1)(B), (h)(2)(C), 122 Stat. 4820; *see infra* Section III.B.2.

34. *Perfect 10, Inc. v. Visa Int'l Serv. Ass'n*, 494 F.3d 788, 795 (9th Cir. 2007).

35. *Tiffany (NJ) Inc. v. eBay, Inc.*, 576 F. Supp. 2d 463, 469–70 (S.D.N.Y. 2008).

36. *Id.* at 478.

37. The court held that eBay exerted sufficient control to be subjected to contributory liability, and then found that eBay's procedures satisfied the requirements of taking action when they knew or should have known of specific acts of infringement:

Nevertheless, under the law as it currently stands, it does not matter whether eBay or Tiffany could more efficiently bear the burden of policing the eBay website for Tiffany counterfeits—an open question left unresolved by this trial. Instead, the issue is whether eBay continued to provide its website to sellers when eBay knew or had reason to know that those sellers were using the website to traffic in counterfeit Tiffany jewelry. The Court finds that when eBay possessed the requisite knowledge, it took appropriate steps to remove listings and suspend service. Under these circumstances, the Court declines to impose liability for contributory trademark infringement.

*Id.* at 470.

38. E-Fencing Enforcement Act of 2009, H.R. 1166, 111th Cong. (2009) (requiring an online market provider to deny high volume sellers access to the marketplace if he has good reason to believe that the sellers acquired their goods unlawfully).

39. *See generally* Lichtman & Posner, *supra* note 18 (summarizing this perspective); Douglas Lichtman & William Landes, *Indirect Liability for Copyright Infringement: An Economic Perspective*, 16 HARV. J.L. & TECH. 395, 396–99 (2003) (same as above).

standard economic framework considers issues of market failure, costs and benefits, and equity to assess the need for an indirect liability regime in specific cases.<sup>40</sup>

### 1. *Market Failure Analysis*

Before imposing an indirect liability regime, economic analysis asks whether there is really any market failure. If there is no market failure then, there is no need for an indirect liability rule. In particular, there need not be an indirect liability rule when the law or the wronged party can effectively reach the bad actor directly<sup>41</sup> and transaction costs are not significant.

Even if the wronged party cannot easily reach a bad actor that a third party can reach, it is still not necessary to impose liability on the third party. When the wronged party and the intermediary can easily negotiate an arrangement, efficiency will guide the third party to undertake enforcement efforts on behalf of the wronged party. This is a key aspect of a market failure analysis. Unless transaction costs interfere with contracting, affected parties can allocate liability efficiently through contractual design.<sup>42</sup>

The presence of transaction costs deserves more emphasis. When a wronged party can directly deal with an involved third party to mitigate measurable financial harm, it is difficult to see why a third party liability rule matters from an efficiency point of view. No matter which party is liable, efforts to stop the harm should continue until further efforts are not worth

40. See Lichtman & Posner, *supra* note 18, at 228–33.

41. The effective reach condition is evaluated prior to an assessment of the ability of a third party to effectively control the bad activity. See *id.* at 230–31. If the law or the wronged party can easily reach the bad actor, then why even consider whether to impose a duty on a third party? Of course, the bad actors are never totally out of reach of the law or wronged parties. With some finite expenditure of resources, perhaps very large, the direct bad actors could be brought to justice or harms could be prevented. The real economic question is whether those costs are larger than the costs of assigning that enforcement role to a third party. And this means that the effective reach condition collapses into the control factor, discussed *infra*. Landes and Lichtman put the comparative point accurately, applied to the specific case of contributory copyright liability: “Holding all else equal, contributory liability is more attractive . . . the greater the extent to which indirect liability reduces the costs of copyright enforcement—as compared to a system that allows only direct liability.” Lichtman & Landes, *supra* note 39, at 398.

42. Lichtman & Posner, *supra* note 18, at 235. Lichtman and Posner also focus on what the parties might do: “The right thought experiment is to imagine that all the relevant entities and all the victims and all the bad actors can efficiently contract one to another and then to ask how the parties would in that situation allocate responsibility for detecting and deterring bad acts.” *Id.* at 257. But there is no need to conduct this thought experiment in the abstract. Free, equal, and rational parties can bargain to allocate responsibility and so we can answer the question of what the parties would do in this thought experiment by looking at what they actually do. The relevant inquiry is whether the bargaining situation is free of significant transaction costs or other obstacles to reaching an agreement.

it—until mitigation efforts cost more than they save. If liability is assigned to the intermediary, then he will take mitigation steps himself or pay the wronged party to do it, whichever is less expensive. If liability is assigned to the wronged party, then he will take the mitigation steps himself, or pay the intermediary to do it, whichever is less expensive. Liability assignments do not change the level of mitigation effort, but change the burden of distribution. In the one case, the intermediary pays, in the other the wronged party pays. As Coase noted, the allocation of resources is the same, but the equities are different.<sup>43</sup>

In some cases, wronged parties bear the costs of these harms under today's legal regime.<sup>44</sup> It is possible that mitigation efforts by intermediaries could reduce these harms. If so, efficient markets should lead wronged parties to create arrangements with intermediaries to take these steps, at least up to the point where further payments to intermediaries do not produce an equivalent reduction in damages.

The processes created by payment intermediaries to respond to external private party complaints reflect the beginnings of these arrangements and might develop into more productive relationships. The extensive processes used by eBay to police trademark violations are also the kind of measures that could be expanded to create efficient enforcement efforts.<sup>45</sup> However, mutually satisfactory enforcement arrangements involving third parties have not emerged to any large degree.<sup>46</sup>

This transaction cost perspective illuminates the issues at stake in *Tiffany (NJ) Inc. v. eBay, Inc.*<sup>47</sup> eBay has a notice-and-takedown program, described by a senior eBay official as follows:

---

43. See generally Ronald H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1 (1960) (presenting the Coase theorem).

44. This is simply the other side of the point that there is no universal rule of indirect liability for all harms.

45. See *supra* notes 35–37 and accompanying text (discussing *Tiffany (NJ) Inc. v. eBay, Inc.*).

46. See, e.g., Lichtman & Posner, *supra* note 18, at 235–37 (expressing puzzlement as to why the parties have not worked out liability arrangements in their discussion of ISP liability for security flaws). This fact could mean that there are no mitigation efforts that intermediaries can undertake that would effectively avoid damages at a price that the wronged parties are willing to pay. It could mean that transaction costs are so high that intermediaries and wronged parties cannot reach efficient arrangements. It could mean that perceptions of equities prevent the parties from reaching a rational accommodation, in the same fashion that parties to the “ultimatum” game in behavioral economics reject advantageous but unfair low-ball offers. See JAMES SUROWIECKI, *THE WISDOM OF CROWDS* 112–13 (2004) (describing the ultimatum game). Or it might mean that the wronged parties are counting on changes in legal liability that would require intermediaries to take enforcement efforts at their own expense.

47. 576 F. Supp. 2d 463 (S.D.N.Y. 2008).

When we are notified that a particular item is counterfeit, we are notified by the intellectual property owner, someone that actually has knowledge of that product because it is their product, and often they are able to tell just by looking at the item on our site that it is counterfeit.

When they certify to us, under penalty of perjury, that that item is counterfeit, we immediately remove the item from our Web site.<sup>48</sup>

Tiffany thought this program was ineffective because the sales of counterfeited products could be completed before this notice-and-takedown program had a chance to operate.<sup>49</sup> Instead, Tiffany wanted eBay to screen its customers, especially large volume customers, to check on whether their sale items were counterfeit. When these arrangements could not be worked out, Tiffany sued eBay for secondary trademark infringement.<sup>50</sup> But Tiffany could have offered to compensate eBay for the costs involved in the extra enforcement efforts it requested. There is no indication that transaction costs prevented negotiations.<sup>51</sup> The fact that eBay and Tiffany were unable to come to an agreement on compensation, despite years of negotiations and discussions, suggests that the full costs of these enforcement efforts exceeded what Tiffany was willing to pay. If Tiffany is a rational actor, willing to pay up to the amount that it would cost it to take its own enforcement actions, the failure to reach an enforcement agreement suggests that eBay is not the least cost enforcer after all.<sup>52</sup>

---

48. *Organized Retail Theft Prevention: Fostering a Comprehensive Public-Private Response: Hearing Before the Subcomm. on Crime, Terrorism, and Homeland Security of the H. Comm. on the Judiciary*, 110th Cong. 26 (2007) (testimony of Robert Chesnut, Senior Vice President, Rules, Trust and Safety, eBay, Inc.).

49. *eBay*, 576 F. Supp. 2d at 482 n.15.

50. *Id.* at 469, 481–83.

51. See Mann & Belzley, *supra* note 8, at 279 n.122. Mann and Belzley state, In a perfect world, the baseline would be irrelevant because the trademark owner would negotiate to purchase a takedown from eBay if that were an efficient outcome. Some reason exists to think that might happen where, as in this case, transaction costs between two large companies are low when compared to the value of the rights being negotiated.

*Id.*

52. The *eBay* court wrote,

In effect, Tiffany's contributory trademark infringement argument rests on the notion that because eBay was able to screen out potentially counterfeit Tiffany listings more cheaply, quickly, and effectively than Tiffany, the burden to police the Tiffany trademark should have shifted to eBay. Certainly, the evidence adduced at trial failed to prove that eBay was a cheaper cost avoider than Tiffany with respect to policing its marks.

576 F. Supp. 2d at 518. But if Tiffany thought that eBay could take enforcement action "more cheaply, quickly, and effectively than Tiffany," *id.*, why didn't they negotiate arrangements with eBay to do that? The fact that they didn't should be at least relevant

A similar point applies to efforts to require eBay to take steps to prevent the sale of stolen merchandise on its website. Congress is considering legislation in this area, and has held several hearings on the topic.<sup>53</sup> eBay has an enforcement program called Partnering with Retailers Offensively Against Crime and Theft (PROACT)<sup>54</sup> and a mechanism to work with aggrieved merchants in connection with stolen merchandise.<sup>55</sup> When an aggrieved party comes to eBay alleging that an item for sale on eBay is stolen, eBay asks for evidence and then conducts an investigation.<sup>56</sup> If it seems that the item is stolen, eBay takes the item down, suspends the seller, and notifies law enforcement.<sup>57</sup> The contentious issue appears to be who should do the serious investigative work required in these cases. As a witness for the retailer community put it at a Congressional hearing, “PROACT for eBay is a good first step, but it doesn’t go nearly far enough to . . . put an affirmative responsibility on eBay to do the work.”<sup>58</sup> If all that is at stake is efficiency and not equity, and if the costs of this enforcement program to eBay are worth the benefits in loss prevention for retailers, then the companies should be able to work out a voluntary program in which eBay undertakes these efforts on behalf of retailers.

A broader framework that accounts for equity and long-term intangible costs and benefits might provide a reason to impose third party liability even

---

evidence that their own efficiency argument is mistaken and that they are really relying on an equity argument: that eBay should be forced to pay for enforcement actions that are not commensurate with Tiffany’s gains because it is their responsibility to do so.

53. The House Judiciary Committee held a hearing on an earlier version of this legislation. To get a sense of how the proposed legislation might affect an online marketplace like eBay, see *E-fencing Enforcement of Act of 2008, the Organized Retail Crime Act of 2008, and the Combating Organized Retail Crime Act of 2008: Hearing on H.R. 6713, H.R. 6491 and S. 3434 Before the Subcomm. on Crime, Terrorism, and Homeland Security of the H. Comm. on the Judiciary*, 110th Cong. 32 (2008) [hereinafter *E-fencing Enforcement Hearing*] (statement of Edward Torpoco, Senior Regulatory Counsel, eBay, Inc.) (expressing concern that the bill would violate a fundamental legal principle that ISPs like eBay “should not be held liable for content posted by third parties”).

54. PROACT allows retailers to submit reports to eBay’s fraud investigators concerning suspected sales of stolen goods on eBay.

55. *E-fencing Enforcement Hearing*, *supra* note 53, at 27 (testimony of Edward Torpoco, Senior Regulatory Counsel, eBay, Inc.).

56. *Id.*

57. *Id.*

58. *Id.* at 46 (testimony of Joseph J. LaRocca, Vice President, Loss Prevention, National Retail Federation). It is hard to avoid the conclusion, voiced by Steve DelBianco on behalf of the online marketplaces at the same hearing, that “retailers would understandably say, we are not ready to sign up for a voluntary program if someone is dangling in front of us legislation that creates a club . . . in the form of being able to demand the interrogation of customers without any law enforcement being involved.” *Id.* at 42 (testimony of Steve DelBianco, Executive Director, NetCHOICE).

when the parties can negotiate enforcement arrangements themselves. But the absence of such agreements in cases where parties can freely negotiate suggests that the wronged party does not think third party liability is worth paying for. This evaluation implied by the wronged party's actions should be a relevant factor taken into account in considering whether to impose third party liability.

The informal arrangements between law enforcement agencies and intermediaries with respect to specific illegal activities reflect the idea that parties can agree about who is responsible for enforcement even without any explicit indirect liability rules. Third parties have frequently worked closely with law enforcement to deter or prevent violations of the law. Intermediaries work with federal, state, and local law enforcement on a variety of issues, including fraud, child pornography, online tobacco sales, and controlled substances.<sup>59</sup> Market failure analysis should look to the presence or absence of these arrangements with law enforcement when assessing whether the direct bad actors are beyond the effective reach of the law.

## 2. *Cost–Benefit Analysis*

Indirect liability regimes can be evaluated using traditional economic tools. Lichtman and Posner emphasize two factors as relevant to deciding whether to impose an indirect liability rule.<sup>60</sup> The first is the extent to which the third party is in a good position to detect or deter the illegal activity.<sup>61</sup> This is the “control” factor.<sup>62</sup> The second comes into play when the third party cannot do anything to detect or deter the illegal activity, but imposing liability can reduce the harm by reducing all the third party's activity, legal and illegal alike.<sup>63</sup> This is the “activity” factor.<sup>64</sup> Both of these ideas apply traditional economic thinking to the special case of indirect liability rules.

The first factor—“control”—does not simply focus on whether the third party is in a good position to detect or deter the illegal activity. It also looks to whether the third party is in a better position than the wronged party, or other potentially liable third parties.<sup>65</sup> Through this analysis, the control factor reflects the “least cost” concept of third party liability.<sup>66</sup>

---

59. *See infra* Section III.B.

60. Lichtman & Posner, *supra* note 18, at 230–31.

61. *Id.* at 230.

62. *Id.*

63. *Id.* at 231–32.

64. *Id.*

65. Several commentators seem to stop with the “good position” analysis. *See id.* at 223 (“Our argument in favor of service provider liability is primarily based on the notion that ISPs are in a good position to reduce the number and severity of bad acts online.”); *see also*

A “least cost” perspective puts the burden of enforcing the law on the party that can stop the illegal transactions at the lowest cost. In the case of internet intermediaries, this line of thought can be summarized as follows: aggrieved parties and enforcement officials face prohibitively high enforcement costs—often because the perpetrators of these illegal acts are individuals or small enterprises, widely dispersed in offshore jurisdictions. In contrast, internet intermediaries possess substantial information regarding activities on their system, they can detect the illegal transactions easily, they are already global in character and they can stop all or most of the illegal transactions using simple methods. It thus seems efficient to adopt a legal rule assigning intermediaries the responsibility for stopping illegal internet transactions. As the least cost avoiders, they should be the internet police.<sup>67</sup>

Focusing on costs is desirable in order to create an efficient enforcement regime. In a “least cost” framework, the cost to the intermediary itself and to the direct customers of the intermediary must be taken into account. If ISPs or payment systems have to incur costs to monitor their system for illegal content, those costs will be passed down to their direct customers. With the price increase, some customers stop using the service or reduce their usage of it. If the service provided is a network service, then the external network effects on other users of the service from an overall reduction in use also have to be counted.<sup>68</sup> According to the least cost idea, when these costs are less than the cost of enforcement activity by the wronged party or by enforcement officials, then liability rests with the intermediary.

---

Lichtman & Landes, *supra* note 39, at 409 (arguing regarding internet intermediaries that “although these parties are only indirectly responsible, they are typically in a good position to either prevent copyright infringement or pay for the harm it causes”).

66. There is an extensive law and economics literature in this area. *See, e.g.*, Mann & Belzley, *supra* note 8, at 265 (discussing the least-cost avoider principle); WARD FARNSWORTH, *THE LEGAL ANALYST: A TOOLKIT FOR THINKING ABOUT THE LAW* 47–56 (2007) (same as above).

67. For the clearest argument in favor of intermediary liability based on this least cost avoider perspective, see Mann & Belzley, *supra* note 8, at 265.

68. If there are fewer internet subscribers, then the service is less valuable to e-commerce merchants as well since there are fewer potential customers. *See* Matthew Schruers, *The History and Economics of ISP Liability for Third Party Content*, 88 VA. L. REV. 205, 250–52 (2002); *see also* Lichtman & Posner, *supra* note 18, at 241–43 (seeming to minimize the importance of these external, network effects in assessing liability regimes: “Immunizing ISPs from liability is not the correct mechanism for encouraging them to provide positive externalities.”). However, the loss of ISP-generated external benefits is a potential cost of assigning liability that has to be taken into account when assessing whether to assign liability. Mann and Belzley’s article gets the overall point right, noting: “To the extent the regulation affects conduct with positive social value, as is likely in at least some of the contexts this Article discusses, the direct and indirect effects on that conduct must be counted as costs of any regulatory initiative.” Mann & Belzley, *supra* note 8, at 274.

This least cost avoider proposal is not advanced as an interpretation of current law.<sup>69</sup> In fact, some courts have explicitly repudiated this principle as an interpretation of current law.<sup>70</sup> Rather, the idea is that regulators and legislators should assess in some fashion whether third parties might be in a better position than the wronged parties to take enforcement action and impose liability on the third parties if the answer is affirmative.

But cost-based analysis is limited because it ignores the possibility that the benefits of enforcement efforts are less than the costs of these enforcement efforts. The mistake is to think that if efforts by third parties provide more enforcement than efforts by the wronged parties, then it must be worthwhile for the third parties to take these enforcement steps. Similarly, it is sometimes thought that if third parties can more easily reach bad actors than the wronged parties, then they should be required to do so. But this is wrong. It is almost always possible to spend more on enforcement and obtain some return. From an economic point of view, the question is whether that extra spending provides commensurate reductions in damages. Therefore, the least cost rule is not the right decision rule, even in a strictly economic analysis. Instead, a full cost-benefit analysis is more appropriate.<sup>71</sup>

This discussion of benefits is the appropriate context in which to analyze the “activity” factor, described *infra*. Effort to reduce the harm facilitated by the third party imposes an external harm on someone else in the form of increased costs. Assigning liability to the third party for monetary damages for this harm means that the price of the activity it makes possible will rise, reducing all of the activity. This in effect internalizes the negative externality that these activities inflict on the wronged parties. Rather than just providing

69. See, e.g., Mann & Belzley, *supra* note 8, at 272. Mann and Belzley state: The liability schemes that this Article envisions are plainly not the type of thing readily adopted through the development of the common law. This Article’s frame-work is intended to provide fodder for legislators and regulators, not for judges. Hopefully, this Article’s analysis can lead to well-specified statutory schemes or regulatory initiatives.

*Id.*

70. See, e.g., Tiffany (NJ) Inc. v. eBay, Inc., 576 F. Supp. 2d 463, 518 (S.D.N.Y. 2008) (“[E]ven if it were true that eBay is best situated to staunch the tide of trademark infringement to which Tiffany and countless other rights owners are subjected, that is not the law.”).

71. The least-cost analysis seems to function like a cost effectiveness analysis, where a given level of enforcement is assumed and the question is how that goal can be reached at the lowest cost. See Mann & Belzley, *supra* note 8, at 250 (adopting that perspective as “a mature scheme of regulation that limits the social costs of illegal Internet conduct in the most cost-effective manner”). But a full cost-benefit analysis gives up the assumption of a fixed benefit goal and takes the value of benefits into account as well.

the third party with an incentive to stop the harmful activity, the regulation of the intermediary discourages all of the activity it makes possible.<sup>72</sup>

The harm that intermediaries and the users of intermediary products and services would suffer from the imposition of an “activity” tax is an essential effect that deserves to be emphasized in these analyses. The implications of this approach need to be made explicit. Assigning liability to intermediaries, even when they cannot reasonably take enforcement steps to prevent harmful activity, is justified only when the overall activity itself causes more harm than good. Following this approach in an effort to reduce online copyright infringement would require finding, in effect, that there is too much use of the Internet given the amount of copyright infringement it allows, and so we have to reduce internet usage. This analysis helps assess the full cost of copyright enforcement: the price of additional copyright protection is that people will use internet services and applications less.<sup>73</sup>

A cost-benefit approach is often not considered in the case of general law enforcement because policymakers are reluctant to put a value on enforcement benefits. For example, policy makers may be unwilling to quantify the benefits of reduced child pornography or reduced sale of controlled substances. But in cases of more tangible harms, such as damages from copyright infringement or counterfeiting, a traditional cost-benefit analysis seems more feasible.

Cost-benefit analysis must take into account long-term considerations and dynamic conditions.<sup>74</sup> A version of the infant industry argument is also

---

72. Lichtman & Posner, *supra* note 18, at 231; Lichtman & Landes, *supra* note 39, at 404–05 (illustrating how this “activity” factor works in discussing, “an instance where it would be prohibitively expensive to distinguish legal from illegal copyright activity,” concluding that “Internet service providers are a good example in this category”). But then Lichtman and Landes note that perhaps they should still be liable in this case:

After all, instead of trying in vain to distinguish lawful from unlawful activity, a firm in this situation would simply increase its price and use that extra revenue to pay any ultimate damage claims. Legal liability, then, would function like a tax. In many instances such a tax would be welfare-reducing in that higher prices discourage legal as well as illegal uses. But in some settings, discouraging both legal and illegal activity would yield a net welfare gain. This would be true where illegal behavior is sufficiently more harmful than legal behavior is beneficial; it would be true where the harms and benefits are comparable but illegal behavior is more sensitive to price; and it would be true where the benefits in terms of increased copyright incentives outweigh the harms associated with discouraging legitimate use.

Lichtman & Landes, *supra* note 39, at 405.

73. It should be noted that Lichtman and Posner reject the activity factor rationale for imposing cyber-security liability on ISPs. Lichtman & Posner, *supra* note 18, at 238–40.

74. See Lichtman & Landes, *supra* note 39, at 408. (“[L]ike any legal issue, these questions about the relative virtues of indirect liability have to be evaluated dynamically.”).

relevant in this long-term context. Often industries need special help and consideration from government when they are just beginning. This help can take the form of public subsidies or immunities from normal legal liabilities. For example, the immunity from liability set out in § 230 of the Communications Decency Act and in the DMCA eased the online community's uncertainty over the extent of their liability and helped spur dramatic new investment in internet infrastructure and services.<sup>75</sup>

A long-term analysis might consider what the world would look like if the same indirect liability burden was placed on third parties by other jurisdictions. It might also consider the likelihood that imposing liability in one case would cause other jurisdictions to also impose liability in other cases. The political implications are one aspect of this analysis. Unilateral attempts to use intermediaries to enforce local laws might create substantial international discord and ramifications in other areas of political or economic life. The cost analysis also needs to account for consequences for the third party, which might potentially be burdened with costs from many jurisdictions—burdens that might be individually rational, but collectively unworkable. If the foreseeable result of imposing third party liability is a race to the bottom, for other cases and other jurisdictions, this has to be part of the cost analysis.

Finally, there is a difference between the costs and benefits to private parties involved and the costs and benefits to society. The costs and benefits of third party enforcement efforts fall on different parties. A wronged party benefits from third party enforcement efforts and the third party pays the costs. The wronged party has a natural incentive to have the third party do as much as possible in the way of enforcement—even past the point where there is a corresponding reduction in damages—because the wronged party appropriates the damage reduction but pays no costs. From an economic efficiency point of view, enforcement efforts that do not yield a commensurate reduction in damages are wasted. Private benefits may not be worth it from a social point of view when balanced against the costs to other parties.

### 3. *Equity Analysis*

The economic framework described above lacks a normative dimension. It does not take into account questions of fairness, rights, and justice. And it

---

75. *See id.* at 406 (referencing this infant industry argument). *But see* Mann & Belzley, *supra* note 8, at 261 (appearing to be critical of the infant industry argument, arguing that exemptions from liability for pure internet actors derive from “the reflexive and unreflective fear that recognition of liability for intermediaries might be catastrophic to Internet commerce”).

does not consider who deserves the benefit of protection from harm or who is at fault, or blameworthy, for failing to take preventive measures. This Section points to the need to take these normative considerations into account.

Mann and Belzley take a strong position on this question and state that “a focus on traditional tort law notions of fault necessarily diverts attention to subjective normative questions of blame and responsibility . . . .”<sup>76</sup> The worry is that these notions will inevitably tangle up policy makers in difficult causation and responsibility questions and will divert attention from the key issue of who can fix the problem. The crucial factor for Mann and Belzley is not who created and maintains the problem, but who can fix it at the least cost.<sup>77</sup>

The view that an economic efficiency standard, by itself, is sufficient to create indirect liability is too strong. The focus on parties who had no part in creating the problem and who are not responsible for the illegal activity puts a burden on people who are innocent of any wrongdoing. Burdening innocent people seems unfair, and arguments that justify this approach on the grounds that it is good for society as a whole violate widely accepted moral principles and are unlikely to withstand public scrutiny.<sup>78</sup>

We should require a person to right the wrongs committed by others only if we think that person is somehow responsible for those wrongs. Determining who is responsible for righting wrongs committed by others is controversial in both moral and political philosophy.<sup>79</sup> Libertarians generally maintain that people need to fix only the problems that they themselves directly created.<sup>80</sup> Without this limitation, it is difficult not to slide into a

---

76. Mann & Belzley, *supra* note 8, at 249. Mann and Belzley propose their idea, “liability without fault,” to be that “intermediaries, without regard to their blame-worthiness, might be the most effective sources of regulatory enforcement.” *Id.* at 262.

77. Mann and Belzley also treat the least cost standard as a legal litmus test. *Id.* at 250–51. Being the least cost avoider is necessary and sufficient for indirect liability. No other standard or consideration intervenes to affect the analysis. As discussed, that standard leaves out the benefits part of the equation and is too limited.

78. See, e.g., JONATHAN WOLFF, AN INTRODUCTION TO POLITICAL PHILOSOPHY 57 (1996) (stating that “utilitarianism will permit enormous injustice in the pursuit of the general happiness”). A more sophisticated indirect or rule utilitarian approach can attempt to meet this difficulty, but that approach is subject to difficulties of its own. See generally JOHN RAWLS, A THEORY OF JUSTICE (1971) (critiquing utilitarianism). The underlying intuition behind this alternative account of social justice is that “[e]ach person possesses an inviolability founded on justice that even the welfare of society as a whole cannot override.” *Id.* at 3.

79. See *infra* notes 80–83 and accompanying text.

80. See Jim Harper, *Against ISP Liability*, 28 REG. 30, 30–31 (2005) (arguing that ISPs should be liable for harms to third parties only if they have a duty to these parties and that “efficiency” considerations do not override the lack of a such duty founded on justice). Libertarians generally reject the idea that we have positive duties to ameliorate harms we did

doctrine that requires all actors to stop misconduct whenever they can.<sup>81</sup> Others think that one has a duty to correct injustices to the extent that one participates in an institutional framework which produces injustice.<sup>82</sup> Still others believe in general positive duties to eliminate harms even when one has no direct role in causing them.<sup>83</sup>

Ultimately, the analysis of indirect liability cannot avoid considerations of fairness, rights, and justice. The key factors in this assessment will be those that have been used traditionally: directness of the involvement by third parties, an assessment of the degree of harm, the knowledge that third parties have or should have about the specific harm, third parties' intentions, whether the third parties are consciously acting in furtherance of a crime or other illegal act, and other similar considerations.<sup>84</sup> These complicated normative and empirical questions cannot be avoided by a single principle that purports to look at costs and benefits alone.<sup>85</sup>

---

not cause. *E.g., id.*

81. Mann & Belzley, *supra* note 8, at 272 (noting that the principle that liability should be assigned regardless of blameworthiness "easily could shade into judicial doctrines that would obligate all actors to stop all misconduct whenever possible" and finding that this "unbounded principle" is "unduly disruptive"). But it is hard to see how their proposal to implement indirect liability through regulation whenever it would be less expansive than leaving liability with the wronged party would be less disruptive.

82. *See, e.g.,* THOMAS W. POGGE, *WORLD POVERTY AND HUMAN RIGHTS* 172 (2002) (arguing that those involved in an institutional order that authorizes and upholds slavery have a duty to protect slaves or to promote institutional reform, even if they do not own slaves themselves).

83. *See, e.g.,* David Luban, *Just War and Human Rights*, in *INTERNATIONAL ETHICS* 195, 209 (Charles R. Beitz et al. eds., 1985) (stating that "all humans in a position to effect" a human right have an obligation to do so).

84. Mann and Belzley criticize the "myopic focus on the idea that the inherent passivity of Internet intermediaries makes it normatively inappropriate to impose responsibility on them for the conduct of primary malfeasors." Mann & Belzley, *supra* note 8, at 261–62. But passivity is relevant to the knowledge and control factors needed to assess liability from an equity point of view. Lichtman and Landes seem to criticize the focus of current law on "knowledge, control, the extent of any non-infringing uses, and other factors" because they are not "particularly clear as to why those issues are central." Lichtman & Landes, *supra* note 39, at 405. But these factors are crucial because they relate to the way in which the equity issues can be resolved.

85. These equity considerations can interact with the cost analysis. Consider the following: suppose transaction costs make it impossible for the wronged parties to negotiate enforcement deals with a third party—they are too numerous or lack the resources to compensate the third party. Suppose further it is possible that the cost savings involved in assigning liability to a third party are substantial. And finally stipulate that the third party's involvement in the harm is so remote that assigning blame is a mistake. We might in that circumstance nevertheless assign liability to the third party. The gains to the rest of us are just too great. However, should we not compensate the third party for taking the enforcement steps he is required to take? Assigning indirect liability when there is not this level of control or fault to justify blameworthiness might be so efficient under a cost analysis

There is another reason to consider equity. Even when transaction costs are low and parties can negotiate enforcement arrangements that make economic sense to them, it is still desirable to know whether it is fair, just, and equitable for one party to bear the cost. Distribution matters, not just efficiency. In fact, when courts decide cases where there are no apparent barriers to an efficient enforcement arrangement,<sup>86</sup> the major issues left are questions of equity.

Many who analyze indirect liability questions from an economic point of view recognize that non-economic factors also require consideration. Lichtman and Posner state that “[t]hese factors—call them ‘control’ and ‘activity level’—help to identify cases where liability might be attractive. The actual question of whether liability should be imposed, however, typically turns on other, often setting-specific considerations.”<sup>87</sup> They note for instance that a rule imposing liability on telephone companies for crank calls would raise separate privacy concerns that might override the control that telephone companies have in that area.<sup>88</sup>

Because the economic argument is not sufficient on its own to justify indirect liability, it is usually supplemented with the language of blame. Judge Kozinski’s dissent in *Perfect 10* is an excellent example of this reasoning:

The weak link in the pirates’ nefarious scheme is their need to get paid; for this they must use the services of legitimate financial institutions. If plaintiff’s allegations are to be believed, the financial institutions (defendants here) collect billions for sellers of stolen merchandise; in a very real sense, they profit from making piracy possible. I can see no reason they should not be held responsible.<sup>89</sup>

According to this argument, the intermediaries should be responsible for stopping the illegal activity, not simply because they are the least cost avoider, but because they profit from the illegal activity. There is a normative

---

that it is worth considering, but in that case the use of compensation mechanisms should also be considered.

86. *See, e.g.*, *Tiffany (NJ) Inc. v. eBay, Inc.*, 576 F. Supp. 2d 463 (S.D.N.Y. 2008).

87. Lichtman & Posner, *supra* note 18, at 231.

88. *Id.* at 231–32.

89. *Perfect 10, Inc. v. Visa Int’l Serv. Ass’n*, 494 F.3d 788, 823 (9th Cir. 2007) (Kozinski, J., dissenting). Kozinski also states in his dissent that “the complaint alleges that defendants are not merely passive providers of services available on equal terms to legal and illegal businesses alike; they are actually in cahoots with the pirates to prop up their illegal businesses and share their ill-gotten gains.” *Id.* at 820. This allegation turns on what I think is a factual mistake—that the higher prices that adult content merchants pay for accepting cards is an attempt to share their ill-gotten gains rather than an attempt to compensate for the extra credit risk these merchants impose on the system. But the normative tone is unmistakable.

judgment here: people should not profit from theft. The blame is usually established by reference to the large volume of revenue that the intermediary is making from the illegal transactions or the cost savings involved in not stopping the transactions. Simultaneously, any efforts that the intermediary takes to mitigate the illegal activity are often downplayed or ignored, making it appear that the third party is willfully refusing to do his part.

This example's focus is not that normative considerations are inappropriate and so it would be better to limit the analysis to neutral economic analysis. Rather, the point is that these normative judgments are an essential ingredient in determining third party liability, and that it is better to accept that fact than to purport to reduce the question to neutral cost-benefit analysis.

### III. APPLYING THE FRAMEWORK TO PAYMENT INTERMEDIARIES

Payment intermediaries have developed and refined policies and practices to deal with illegal internet transactions in their payment networks. This Part of the Article discusses several examples where these intermediaries took action to control illegal activity on their systems and applies the framework developed in Part II to these specific cases. This discussion illuminates some answers to the general question of the appropriate role of intermediaries in controlling illegal internet activity. For example, the internet gambling example illustrates the general point that intermediaries are better than others at monitoring their systems for activity of a certain type, but not for determining that these activities are illegal. Two general conclusions from this analysis are presented.

One conclusion that can be drawn from these examples is that payment intermediary action has been effective. As the following discussions demonstrate, internet gambling websites have been denied access to the U.S. market, and their current and projected revenues are in decline.<sup>90</sup> Websites for pornography, controlled substance, and tobacco have been substantially eliminated from the traditional payment systems.<sup>91</sup> Ongoing monitoring and enforcement actions in these areas continue to keep their presence in traditional payment networks at a minimal level.<sup>92</sup> For example, as a result of the payment system action in the Allofmp3.com copyright infringement case,

---

90. *See infra* Section III.A.

91. *See infra* Section III.B.

92. *See infra* Section III.B.

Allofmp3.com was confined to a domestic market and experienced a dramatic reduction in the volume of activity at its website.<sup>93</sup>

The second conclusion from these examples is that the widespread assumption that payment system action in this area is simple and almost cost-free deserves more careful consideration.<sup>94</sup> The discussion of payment intermediaries' activities in the five case studies, *infra*, reveals substantial costs that should give policy makers pause before moving ahead with the imposition of an indirect liability scheme for payment providers. These include:

- The cost to maintain and enforce an internet gambling coding and blocking scheme that is entirely manual and cannot be automated;
- The cost from over-blocking legal transactions;
- The cost to screen and check the business activity of merchants participating in the payment systems;
- The cost to monitor the use of payment systems for specific illegal activity, where the payment systems are in no better position than anyone else to conduct this monitoring activity;
- The cost to assess complaints of illegality, where the intermediary has no special expertise and is often less familiar with the legal and factual issues than the wronged party and the allegedly bad actor;
- The cost to defend against legal challenges to enforcement actions, where the challenge typically comes in an off-shore jurisdiction; and
- Long-term costs to the United States from taking unilateral action in this area, including the encouragement of copycat regimes in other areas of law and in other jurisdictions.

The reasonableness of these costs in light of the benefits achieved has not yet been seriously studied. Instead, it seems to be assumed that small compliance costs are justified by large enforcement benefits. Although precision in the estimates of costs and benefits is unlikely in this area, a more disciplined qualitative analysis is required.

Cost-benefit analysis can be labeled the enemy of regulation, but, in principle, this is not the case. It is simply one tool for analysis. It can be turned into a procedural obstacle to regulation if a requirement for extensive

---

93. *See infra* Section III.C.3.

94. *See, e.g., Perfect 10*, 494 F.3d at 824 (Kozinski, J., dissenting).

analysis is imposed even in cases where the costs and benefits seem obvious. But the costs and benefits of indirect liability for intermediaries are not obvious. A careful market analysis, assessment of costs and benefits, and equity analysis must be undertaken before these indirect liability regimes are imposed on payment intermediaries.

The payment intermediaries considered here are the traditional payment networks, such as Visa, MasterCard, American Express, and Discover. These enterprises are private, contractual systems that provide a platform linking merchants who accept cards for payment and cardholders who use cards to pay for goods and services. Payment systems include unitary enterprises such as American Express and Discover, and independent companies such as Visa and MasterCard who link separate financial institutions into an electronic payment network.<sup>95</sup>

These financial intermediaries are different from pure internet intermediaries. Pure internet intermediaries such as internet access providers, search engines, and online auction sites do not have a separate existence apart from the internet service or application that they provide.<sup>96</sup> Conversely, the traditional payment intermediaries evolved offline. The primary application of their services is face-to-face retail transactions, though they have extended this primary service to other channels such as mail-order and telephone service. The Internet is just another channel of commerce for them.<sup>97</sup> Despite this difference, their experiences controlling illegal online activity illuminate the general issue of intermediary liability for illegal online activity.

---

95. In the Visa and MasterCard systems, a payment card transaction involves an authorization message sent from the merchant where the card is being used to the financial institution that provides processing services for the merchant. The message is routed through the network's communications and computer systems to the bank that issued the card to the customer. The issuing bank authenticates the card information submitted in the message and authorizes the transaction after ascertaining that the cardholder has sufficient funds or credit in his or her account. Sometime after the initial authorization of the transaction, a second process routed through the network system clears and settles the transaction, transferring funds from the cardholder's financial institution to the merchant's account at his payment card bank. *See* DAVID S. EVANS & RICHARD SCHMALENSSEE, *PAYING WITH PLASTIC: THE DIGITAL REVOLUTION IN BUYING AND BORROWING* 9-14 (2d ed. 2005) (further describing these payment systems).

96. ISPs are often part of companies that also provide wireline, wireless telephone service, or video programming. But in so far as they provide internet access, they are intrinsically linked to the Internet as an essential part of their business.

97. Payment intermediaries are similar to delivery services such as UPS or FedEx in this respect. They provide a service that is essential to the proper functioning of electronic commerce, but that service is not intrinsically tied to the online channel.

## A. INTERNET GAMBLING LEGISLATION

The early efforts of government to use financial intermediaries to restrain internet gambling have been noted by other commentators.<sup>98</sup> These efforts came from state and federal law enforcement officials, who used their existing resources to pressure financial institutions to take steps against internet gambling merchants.<sup>99</sup> For example, Elliott Spitzer, then Attorney General for New York, pressured Citigroup and other financial institutions to agree to block internet gambling transactions in 2002.<sup>100</sup>

In 2006, Congress passed the Unlawful Internet Gambling Enforcement Act (UIGEA), which imposed a system of indirect liability on financial institutions for the purpose of preventing illegal internet gambling transactions.<sup>101</sup> At the time, many states made internet gambling illegal, and federal law outlawed at least some versions of it in interstate commerce.<sup>102</sup> But customers could evade these local laws by visiting internet gambling merchants located outside of the United States' jurisdiction. UIGEA was an attempt to enforce these local laws through the policies and practices of payment intermediaries.

Prior to the passage of UIGEA, payment card networks devised a coding and blocking system in order to manage the risks of internet gambling.<sup>103</sup> Each merchant in the payment system is normally required to identify its major line of business and to include a four digit "merchant category code" in each authorization message.<sup>104</sup> For gambling, this merchant category code was 7995.<sup>105</sup> In addition, merchants were required to use an electronic commerce indicator when an internet transaction was involved.<sup>106</sup> Together,

---

98. See, e.g., Mann & Belzley, *supra* note 8, at 288–90.

99. *Id.*

100. GOLDSMITH & WU, *supra* note 7, at 82.

101. Unlawful Internet Gambling Enforcement Act of 2006, Pub. L. No. 109-347, 120 Stat. 1884 (codified at 31 U.S.C. §§ 5361–5367 (2006)).

102. The Wire Act is used to prosecute internet gambling activities across state lines. 18 U.S.C. § 1084 (2006); see also U.S. GEN. ACCOUNTING OFFICE, INTERNET GAMBLING: AN OVERVIEW OF THE ISSUES 3–5 (2002) (describing the way these issues were perceived in 2002 when Congress was moving forward with internet gambling legislation).

103. *Financial Aspects of Internet Gaming: Good Gamble or Bad Bet?: Hearing Before the Subcomm. on Oversight and Investigations of the H. Comm. on Financial Servs.*, 107th Cong. 25–27, 34–35 (2001) [hereinafter *Financial Aspects of Internet Gaming Hearing*] (statement and testimony of Mark MacCarthy, Senior Vice President, Public Policy, Visa, U.S.A., Inc.) (describing this system of coding and blocking internet gambling transactions); U.S. GEN. ACCOUNTING OFFICE, *supra* note 102, at 20–25 (same as above).

104. U.S. GEN. ACCOUNTING OFFICE, *supra* note 102, at 22.

105. VISA MERCHANT CATEGORY CLASSIFICATION (MCC) CODES DIRECTORY, available at [http://www.da.usda.gov/procurement/card/card\\_x/mcc.pdf](http://www.da.usda.gov/procurement/card/card_x/mcc.pdf).

106. U.S. GEN. ACCOUNTING OFFICE, *supra* note 102, at 22.

these two pieces of information in the authorization message allowed payment networks or issuing banks to identify transactions involving internet gambling merchants.<sup>107</sup>

Given this system, it was entirely feasible for the issuing bank or the payment network to block internet gambling transactions. The system could accommodate conflicting laws in different jurisdictions in the following way: if it was illegal in one country, such as the United States, for cardholders to engage in internet gambling, then the issuing banks based in that country could decline authorization requests for all properly coded internet gambling transactions. This would effectively block these transactions. However, the banks in other countries who permit internet gambling, such as the United Kingdom, could allow the use of their cards for internet gambling by not declining properly coded internet gambling transactions.

This system was a reasonable accommodation of the conflicting laws in different jurisdictions, but it was not perfect. For one thing, banks were able to issue cards outside of their own jurisdiction. A British citizen could obtain both a card issued by a U.S. bank and a card issued by a British bank. The British bank-issued card would work for internet gambling whereas the U.S. bank-issued card would not. People residing in countries where internet gambling was illegal were able to evade payment system blocking by obtaining cards issued by banks from jurisdictions where internet gambling was legal. This loophole was likely small and applied mostly to expatriates who did not give up their local cards when they moved to a new jurisdiction.

A second issue with the blocking system was that it could not accommodate legal gambling transactions made overseas by U.S. cardholders, who would find their cards declined for such activity. These “in-transit” transactions, however, were likely to be few, and this seemed to be a relatively small price to pay for a system that largely mapped the major contours of the internet gambling problem.

The third way in which the system was limited was in detecting nuances in illegal versus legal internet gambling. If a jurisdiction recognized some internet gambling transactions as legal and others as illegal, the system would not detect it.<sup>108</sup> The merchant category code described a type of business, not the legal status of the transaction involved.<sup>109</sup> If a particular jurisdiction allowed casino gambling, but not sports betting, both transactions would nevertheless be labeled 7995. And if the system was set up to block these

---

107. *Id.*

108. U.S. GEN. ACCOUNTING OFFICE, *supra* note 102, at 22.

109. *See* VISA MERCHANT CATEGORY CLASSIFICATION (MCC) CODES DIRECTORY, *supra* note 105 (listing all the MCC codes by “merchant type”).

coded transactions, then both transactions, legal and illegal, would be blocked.<sup>110</sup>

Another weakness in the system was enforcement. If an internet gambling merchant realized that his transactions would be blocked in a large jurisdiction such as the United States, then he would have every incentive to hide.<sup>111</sup> Instead of describing itself as a gambling operation, the merchant would just code itself as a T-shirt sales site or some other legal entity. Without the proper merchant category code, the system was blind and could not effectively block the merchant's transactions.<sup>112</sup>

The payment networks addressed this enforcement issue with a special program to verify that internet gambling merchants coded their transactions correctly.<sup>113</sup> Payment network personnel would test transactions at popular internet gambling sites.<sup>114</sup> They would enter a transaction at the website and track the transaction through the payment system.<sup>115</sup> They would be able to tell whether the transaction was coded properly or not after they identified the transaction in the system.<sup>116</sup> If the transaction was not properly coded, the network would contact the bank that worked with the merchant and tell the bank that its merchant was out of compliance with the coding rule.<sup>117</sup> The payment network would ask the bank to take steps to bring the merchant into compliance.<sup>118</sup> Finally, the network would retest the site for proper coding.<sup>119</sup>

There was nothing automatic about the coding enforcement program. Staff manually entered the transactions, tracked them in the processing and authorization system, found the appropriate financial institution, and contacted that institution. It was time consuming and resource intensive. And it was not a temporary measure. Unless the coding enforcement program was maintained indefinitely, merchants would simply return to their

---

110. U.S. GEN. ACCOUNTING OFFICE, *supra* note 102, at 22.

111. *Id.* at 26.

112. *Id.*

113. *Id.* at 31–32. The fines for incorrectly identifying authorization requests for online gambling transactions are set out at 1.6.D.7 of the Visa International Operating Regulations. VISA, *supra* note 21. In addition, Visa requires online gambling merchants to post certain notices: “a Website for an Online Gambling Merchant must contain . . . [t]he statement ‘Internet Gambling may be illegal in the jurisdiction in which you are located; if so, you are not authorized to use your payment card to complete this transaction.’” *Id.* at 5.4.C.2.

114. U.S. GEN. ACCOUNTING OFFICE, *supra* note 102, at 32.

115. *Id.*

116. *Id.*

117. *Id.*

118. *Id.*

119. *Id.*

previous practice of miscoding their transactions, thus undercutting the blocking scheme's effectiveness.

An additional way in which a payment system could react to the concern about internet gambling was to avoid signing up gambling merchants. American Express, for instance, refused to sign up these merchants, partly because of the substantial risk of non-payment for gambling debts, partly because of legal risk, and partly because of the reputational damage involved in accepting transactions that many viewed as sinful or harmful.<sup>120</sup>

The UIGEA required payment systems to have policies and procedures reasonably designed to stop illegal internet gambling transactions.<sup>121</sup> The implementing regulations required the affected parties to “establish and implement written policies and procedures reasonably designed to identify and block or otherwise prevent or prohibit” illegal internet gambling transactions.<sup>122</sup>

The statute creates a safe harbor for payment systems that adopt a coding and blocking scheme.<sup>123</sup> The Federal Reserve Board and the Department of the Treasury implemented this safe harbor with a non-exclusive description of one way in which a payment system can demonstrate that its policies and practices are reasonably designed to stop illegal internet gambling transactions.<sup>124</sup> This description tracked the existing industry practices.

To take advantage of this safe harbor, a payment system must, in effect, maintain the coding and blocking scheme that was in place around 2006. There are other ways to satisfy the general duty, but the presence of an approved safe harbor written into both the statute and the implementing regulations means that any replacement mechanism has to demonstrate that

120. *Id.* at 20–21.

121. Unlawful Internet Gambling Enforcement Act of 2006, Pub. L. No. 109-347, 120 Stat. 1884 (codified at 31 U.S.C. §§ 5361–5367 (2006)).

122. 12 C.F.R. § 233.5(a) (2009).

123. 12 C.F.R. § 233.6(d)(1)(ii) (2009).

124. The code's relevant section reads:

(ii) Implementation of a code system, such as transaction codes and merchant/business category codes, that are required to accompany the authorization request for a transaction, including—

(A) The operational functionality to enable the card system operator or the card issuer to reasonably identify and deny authorization for a transaction that the coding procedure indicates may be a restricted transaction; and

(B) Procedures for ongoing monitoring or testing by the card system operator to detect potential restricted transactions, including—

(1) Conducting testing to ascertain whether transaction authorization requests are coded correctly; and

(2) Monitoring and analyzing payment patterns to detect suspicious payment volumes from a merchant customer . . . .

*Id.*

it is at least as effective as the approved safe harbor. The practical effect is to extend indefinitely the current coding and blocking system.

1. *Implementation Challenges with the Internet Gambling Act*

A key issue faced by payment intermediaries in implementing the new law was the need to clarify the status of certain gambling operations as legal or illegal. This issue illustrates an important point that will emerge in other areas as well: providing for private sector enforcement of legally ambiguous laws creates significant problems for the intermediary and for other market participants who are affected by the legal ambiguity.

The regulations implementing the UIGEA generally define “unlawful Internet gambling” as

placing, receiving, or otherwise knowingly transmitting a bet or wager by any means which involves the use, at least in part, of the Internet where such bet or wager is unlawful under any applicable Federal or State law in the State or Tribal lands in which the bet or wager is initiated, received, or otherwise made.<sup>125</sup>

However, the regulations “should not be construed to alter, limit, or extend any Federal or State law or Tribal-State compact prohibiting, permitting, or regulating gambling within the United States.”<sup>126</sup> The regulations “do[] not spell out which activities are legal and which are illegal, but rather relies on the underlying substantive Federal and State laws.”<sup>127</sup>

The Federal Reserve Board and the Department of the Treasury (the Agencies) reasonably refused to try to define the unlawful internet gambling. Given the states’ varying approaches to the regulation of gambling within their jurisdictions, it found that “the underlying patchwork legal framework does not lend itself to a single regulatory definition of ‘unlawful Internet gambling.’”<sup>128</sup>

The difficulty of the issue can be seen in the context of horse racing. The existing statute appears to authorize internet betting on horse racing. The Interstate Horseracing Act (IHA) authorizes interstate off-track wagers and

---

125. Prohibition on Funding of Unlawful Internet Gambling, 72 Fed. Reg. 56,680, 56,681 (Oct. 4, 2007) (to be codified at 12 C.F.R. § 233, 31 C.F.R. § 132).

126. *Id.* at 56,681–82.

127. *Id.* at 56,682.

128. Prohibition on Funding of Unlawful Internet Gambling, 73 Fed. Reg. 69,382, 69,384 (Nov. 18, 2008) (to be codified at 12 C.F.R. pt. 233 & 31 C.F.R. pt. 132). For similar reasons, the Agencies declined to develop, publish and update a list of merchants who were in violation of the Act. *Id.*

then defines this term as “a legal wager placed or accepted in one State with respect to the outcome of a horserace taking place in another State.”<sup>129</sup>

The facial meaning of this definition seems to allow bets on horseracing to take place through electronic media, such as the Internet, provided that the transaction is legal in each State. Of course, only domestic providers of these betting services can participate in this type of internet gambling because only they are located in one of the States. This was at the heart of a World Trade Organization (WTO) ruling finding that the IHA violated U.S. trade commitments.<sup>130</sup> In effect, U.S. law discriminated against offshore commercial establishments by allowing domestic companies to provide betting on horse racing while prohibiting offshore websites from doing so.<sup>131</sup> The U.S. Department of Justice maintains that internet horse race betting is illegal.<sup>132</sup>

There is no good resolution of this issue for the financial institutions involved. If they block internet horse racing bets, they appear to be siding with the Department of Justice’s interpretation of the law, against the views of the horse racing industry and the WTO. If they do not, they appear to be defying the agency charged with enforcing that law.<sup>133</sup>

129. Interstate Horseracing Act of 1978, 15 U.S.C. §§ 3001–3007 (2006). Interstate horseracing “includes pari-mutuel wagers, where lawful in each State involved, placed or transmitted by an individual in one State via telephone or other electronic media and accepted by an off-track betting system in the same or another State, as well as the combination of any pari-mutuel wagering pools.” *Id.* § 3002(3).

130. Appellate Body Report, *United States—Measures Affecting the Cross-Border Supply of Gambling and Betting Services*, ¶ 358–64, WT/DS285/AB/R (Apr. 7, 2005).

131. See GOLDSMITH & WU, *supra* note 7, at 172–73 (discussing the WTO case).

132. See *Internet Gambling Prohibition Act of 2006: Hearing on H.R. 4777 Before the Subcomm. on Crime, Terrorism, and Homeland Security of the H. Comm. on the Judiciary*, 109th Cong. 80 (2006) (testimony of Rep. John Conyers, Jr., Ranking Member, H. Comm. on the Judiciary) (“The Department of Justice views the existing criminal statutes as prohibiting the interstate transmission of bets or wagers including wagers on horse races.” (quoting the Department of Justice’s earlier statement)).

133. One attempt to resolve the issue was written into the statute. The Agencies were required to “ensure that transactions in connection with any activity excluded from the [Act’s] definition of unlawful internet gambling . . . are not blocked or otherwise prevented or prohibited by the prescribed regulations.” 31 U.S.C. § 5364(b)(4) (2006). This appeared to exempt the horse racing industry, Indian gaming, and intrastate gambling from the purview of the statute. The Agencies could have interpreted this to mean that the payment systems were required to process transactions that were not prohibited. They might have required the card systems to use special codes for transactions that are not prohibited by UIGEA. But the Agencies declined to do these things. They determined that they did not have the authority to require card systems to process certain transactions and they left the creation of special merchant category codes to the business judgment of the card systems. See *Prohibition on Funding of Unlawful Internet Gambling*, 73 Fed. Reg. 69,382, 69,391 (Nov. 18, 2008) (to be codified at 12 C.F.R. pt. 233 & 31 C.F.R. pt. 132).

This left the determination of the legality of horse betting to the payment systems. Card systems were permitted to continue using the industry gambling code as a way to implement the safe harbor, which could result in the blocking of some transactions that industry participants feel are perfectly legal.<sup>134</sup> Financial institutions were generally instructed to resolve all questions of ambiguous legality by following a process of “due diligence” where they would be required to consult the various state and federal statutes to determine if a particular business was engaged in unlawful internet gambling.<sup>135</sup>

The legislation imposed indirect liability obligations on other parties in addition to financial institutions. The law allows the U.S. Attorney General (AG) or a state attorney general to ask a district court to enter a restraining order or injunction against any person in order to prevent or restrain an unlawful internet gambling transaction.<sup>136</sup> But these actions are limited in the case of interactive computer services.<sup>137</sup> Under this remedy, the AG can only ask the computer service to remove or disable access to a site, or link to a site, involved in illegal internet gambling, and it has to specifically identify the location of the offending site or link.<sup>138</sup> The request must be directed to a specific service, not to all interactive computer services generally, and the remedy cannot require the service to monitor its system for illegal internet gambling sites.<sup>139</sup> UIGEA reflects how pure internet intermediaries are treated differently from the traditional payment systems.

## 2. *Internet Gambling Assessment*

The experience of payment intermediaries in controlling illegal internet gambling can be assessed through the framework developed in Part II. This assessment consists of an equity analysis, a market analysis, and a cost-benefit analysis. The specific questions are: whether the online gambling legislation imposed an unfair burden (equity analysis); whether the market was adequately addressing the issue without it (market analysis); and whether it imposed burdens that exceeded their benefits and whether there was a feasible alternative that would achieve the same benefit at lower cost (cost-benefit analysis).

---

134. They were given liability protection for this possible over-blocking. *See id.*

135. *Id.* at 69,391–92.

136. 31 U.S.C. § 5365(a)–(b) (2006).

137. *Id.* § 5365(c).

138. *Id.*

139. *Id.* An interactive computer service has the same meaning as in § 230 of the Communications Decency Act of 1996, *id.*, indicating that this provision has its roots in the same internet exceptionalist thinking that generated that statute.

On equity grounds, the payment system connection to internet gambling is too passive to justify imposing legal responsibility for blocking illegal internet gambling. Payment intermediaries are not to blame when others use their system for internet gambling because these intermediaries have no specific connection to the activity other than operating a general purpose payment system. They do not reap extra profits through special arrangements with the internet gambling merchants. Internet gambling transactions are no different from any other payment card transaction. On pure equity grounds alone, then, there is no reason to single out these transactions and impose special legal responsibilities.

A market analysis indicates that there are still some feasible enforcement arrangements that were not established prior to the passage of the internet gambling law. Although intermediaries may not be responsible for their customers' gambling, many of them are concerned about the social ills connected with the activity and want to reduce its prevalence.<sup>140</sup> U.S. financial intermediaries had already refused to sign up domestic internet merchants because these merchants were not authorized to act legally in the United States.<sup>141</sup> Some state attorneys general requested that the intermediaries block offshore gambling activities, and many cooperated.<sup>142</sup> These agreements did not extend to all financial institutions and did not cover all states, but they could have been extended without imposing a legislative requirement.

A cost-benefit analysis of UIGEA starts with an estimate of its effect on the amount of illegal internet gambling activity. Shortly after the bill was signed into law in 2006, analysts estimated that the value of British internet gambling stocks declined by \$7.6 billion.<sup>143</sup> A reduction in internet gambling activity in the United States resulted from the voluntary agreements just described. This reduction would continue after the legislation's compliance date. It is likely that the end result would be the substantial elimination of internet gambling by customers of U.S. financial institutions.

The costs associated with the payment systems' compliance with the legislation include the costs of maintaining and enforcing an internet gambling coding and blocking scheme, which is entirely manual and cannot

---

140. See *Financial Aspects of Internet Gaming Hearing*, *supra* note 103, at 25–26 (statement of Mark MacCarthy, Senior Vice President, Public Policy, Visa U.S.A., Inc.).

141. *Id.* at 26; U.S. GEN. ACCOUNTING OFFICE, *supra* note 102, at 20.

142. See, e.g., GOLDSMITH & WU, *supra* note 7, at 82 (discussing Spitzer's efforts "to convince every major American credit card provider and online payment system to stop honoring web gambling transactions").

143. Eric Pfanner & Heather Timmons, U.K. *Seeks Global Rules for Online Gambling*, INT'L HERALD TRIB., Nov. 2, 2006, at 14.

be automated. This Section highlights the additional costs associated with increases in international tensions, the copycat effect, and over-blocking.

There are substantial costs associated with increased tension in international relations. The United States' passage of the internet gambling law was one of the first unilateral legislative actions by a major jurisdiction using intermediaries to enforce local law on the Internet. It had substantial international relations repercussions. Reactions to the U.S. internet gambling legislation from abroad have not been favorable. The British culture minister, noting that the industry "has been very hard hit by the U.S. ban" and that the Internet is a "global marketplace," urged "action at the global level."<sup>144</sup> While Britain was seeking to develop a consensus on a global standard for legalizing and regulating internet gambling,<sup>145</sup> the U.S. law went in the opposite direction of taking unilateral action to close off the U.S. market.

The WTO reaction, concerning the legality of the United State's action in light of its trade commitments, continues to stir international trade and political issues.<sup>146</sup> Among the costs of using intermediaries to bar internet gambling transactions are these continuing conflicts. For example, in June 2009, the European Union released a report criticizing the U.S. internet gambling laws, asserting that they violated WTO agreements, and urging negotiation to resolve the issues.<sup>147</sup>

International concerns about the U.S. gambling law may also prompt costs stemming from the copycat effect, which uses the assignment of indirect liability in one case to argue for its legitimacy in another unrelated case. The copycat effect of the internet gambling liability rule cannot be overestimated as one of the costs of the legislation deserving consideration.

Judge Kozinski, dissenting in *Perfect 10*, demonstrates the force of the copycat argument: "Requiring defendants to abide by their own rules, which strictly prohibit members from servicing illegal businesses, will hardly impair the operation of a vibrant and competitive free market, any more than did the recent law prohibiting the use of credit cards for Internet gambling."<sup>148</sup> Simply put, the argument is that since payment system contracts bar all illegal activity, payment systems should be responsible for enforcing all laws everywhere. But the key point here is the effect of the internet gambling precedent. Kozinski seems to be saying: "If it worked in the case of internet

---

144. *Id.* (quoting Tessa Jowell).

145. *Id.*

146. *See supra* note 130 and accompanying text.

147. Press Release, European Comm'n, EU Commission Publishes Report on US Internet Gambling Laws (June 10, 2009), *available at* <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/09/902>.

148. *Perfect 10, Inc. v. Visa Int'l Serv. Ass'n*, 494 F.3d 788, 824 (9th Cir. 2007) (Kozinski, J., dissenting) (internal citations omitted).

gambling, why not elsewhere?" Indeed, why not everywhere? And once the precedent is set, it is hard to return to a detailed analysis of the facts to see if indirect liability makes sense in a specific case.

The copycat effect could very well extend globally. First, it might encourage other countries to use payment systems to implement their own internet gambling regimes. The copycat effect easily extends to other areas of law. It is worth considering how a French court approaching the Yahoo! case and influenced by the copycat effect might exert authority over French financial institutions. In this case, a French court attempted to require Yahoo!, a U.S. company, to prevent the online sale of Nazi paraphernalia to French citizens, which was prohibited under French law.<sup>149</sup> The new U.S. internet gambling law reveals an additional strategy that could be made available to French enforcement entities: deputizing French banks to prevent any transactions involving French citizens and Nazi paraphernalia. French banks are clearly subject to French law. There is no direct extraterritoriality involved. The obligation would be similar to the U.S. internet gambling obligation by establishing reasonable policies and procedures that prevent the use of the French banking facilities for payment of transactions involving Nazi paraphernalia.

Such an application of U.S. gambling law would have extremely costly results, not just for the international payment brands but for all financial institutions globally. To avoid this obligation, French financial institutions would have to drop out of global payment networks, something they would be reluctant to do. If they chose to remain part of the global payment system, the global payment system would have to modify itself in order to accommodate this new obligation. For example, a coding and blocking scheme would no longer be feasible because the content of retail transactions is not coded in payment system transaction messages. Furthermore, a code for Nazi paraphernalia would be much too specific for any business needs. To implement the rule, the global payment systems would have to require that all banks in their system ensure that their merchants do not submit

---

149. *L'Union des Etudiants Juifs de France (UEJF), la Ligue Contre le Racisme et L'Antisemitisme (LICRA) v. Yahoo!, Inc. et Yahoo France*, T.G.I. Paris, Nov. 20, 2000, (Reporter) 05308, Gaz. Pal. 2000, somm. jurispr. 1307. The ultimate French enforcement power was over assets that Yahoo! had in France. *Id.* Yahoo! could have avoided this enforcement power by simply withdrawing from France. In this particular example, other countries probably would not need to use financial institutions, since most online entities, including Yahoo!, have taken steps to prevent the shipment of Nazi paraphernalia to countries that ban them. Similar issues arise in conjunction with a suit against Wikipedia by a convicted murderer who is invoking Germany's privacy laws in a bid to remove references to his killing of a Bavarian actor in 1990. See David Kravets, *Convicted Murderer Sues Wikipedia, Demands Removal of His Name*, WIRED, Nov. 11, 2009, [http://www.wired.com/threatlevel/2009/11/wikipedia\\_murder/#ixzz0el12tNDD](http://www.wired.com/threatlevel/2009/11/wikipedia_murder/#ixzz0el12tNDD).

transactions involving Nazi paraphernalia and customers of French banks. Enforcing this effectively would require locating the internet sites involved in these sales, perhaps through their advertisements or through searching the web for appropriate keywords and symbols. Further steps might involve doing test transactions using a card issued by a French bank, and then taking enforcement steps against merchants submitting transactions for authorization in violation of this rule.

Another cost, as introduced above, is the over-blocking problem created by the way in which payment intermediaries comply with UIGEA. Perfectly legal transactions will likely be blocked because payment intermediaries cannot distinguish them from illegal transactions. This example illustrates that intermediaries are usually better than others at monitoring their own systems for business activity of a certain type, but not at detecting the illegality of activity on their systems.<sup>150</sup> The point arises in internet gambling because the codes used by financial institutions reflect the business activity of gambling, not its status as legal or illegal. As a result, the payment systems' policies and procedures, which were adopted to comply with the Act and which have been accepted by the implementing regulations, over block and prevent perfectly legal activity from taking place.<sup>151</sup>

---

150. See Mann & Belzley, *supra* note 8, at 278 (“Surely eBay is more adept at searching and monitoring its marketplace than Tiffany & Co., while eBay probably is not as effective as Tiffany & Co. in distinguishing bona fide Tiffany products from counterfeits.”); see also Schruers, *supra* note 68, at 252 (“[T]he ISP is not the least-cost avoider when it comes to discovering [illegal] content; it is only well suited for cost avoidance after it is apprized of the problem.”). Schruers adds that in this case, the wronged party may be better suited to the task of locating the offending content. *Id.* at 252.

151. Mann & Belzley, *supra* note 8, at 294. Mann and Belzley present a useful discussion of this over-blocking issue:

[A] risk always exists that imposing additional burdens on intermediaries will chill the provision of valuable goods and services. That will be especially problematic in cases where considerable risk of chilling legal conduct that is adjacent to the targeted conduct exists. As discussed below, that might tend to make the use of intermediaries less plausible in file-sharing contexts where determining whether any particular act of file-sharing is illegal is difficult, and much more plausible in the gambling context where in many cases substantially all traffic to a particular site likely involves illegal conduct. Requiring intermediaries to make those kind of subjective decisions imposes costs not only on the intermediaries that must make those decisions, but also on the underlying actors whose conduct might be filtered incorrectly.

*Id.* at 274. The internet gambling case illustrates that determining when a website is engaged in illegal gambling is not a simple task. It is fraught with the kind of “subjective decisions” that Mann and Belzley are properly concerned about. Payment systems faced with this difficulty do not to make these subjective decisions, but block all gambling activity, including legal gambling transactions.

In light of this difficulty, there might be more effective ways of assigning liability. The new law creates unnecessary confusion by failing to define the term “unlawful internet gambling.” As Congressman Barney Frank wrote to former Secretary of the Treasury Henry Paulson, “The proposed regulations, like the underlying UIGEA statute, fail to define the term ‘unlawful internet gambling,’ leaving it to each financial institution to reconcile conflicting state and federal laws, court decisions and inconsistent Department of Justice interpretations when determining whether to process a transaction.”<sup>152</sup> He has introduced legislation to regulate internet gambling merchants.<sup>153</sup> It would require the Secretary of the Treasury to license internet gambling establishments, and would provide immunity for financial service companies who process transactions to licensed entities.<sup>154</sup> The licensing process would be the exclusive way for internet gambling sites to operate legally.<sup>155</sup> The obligations to block transactions from other, unlicensed internet gambling merchants would remain.<sup>156</sup> The lack of clarity about which merchants are legal would be resolved through the licensing process. At best, the system would rely on a list of approved gambling entities that the payment networks could check before approving gambling transactions from particular internet merchants.

To be effective, however, this list would have to be coordinated with other jurisdictions. Payment systems might respond to such a regime by updating their coding and blocking system or by requiring banks in other jurisdictions to take other steps. Either system might be manageable for the

---

152. Press Release, House Comm. on Fin. Servs., Frank Calls on Bush Administration to Delay Internet Gambling Regulations (Nov. 10, 2008), *available at* [http://www.house.gov/apps/list/press/financialsvcs\\_dem/11102008.shtml](http://www.house.gov/apps/list/press/financialsvcs_dem/11102008.shtml). Chairman Frank wrote that he “introduced legislation (HR 5767, later HR 6870) that would prohibit the implementation of these flawed rules and replace them with a formal rulemaking process that would define the term ‘unlawful internet gambling,’ something the proposed rules fail to do.” *Id.* House Bill 6870 was passed by the Financial Services Committee on September 16, 2008. *Id.* But the Agencies issued their final rules in November 2008. Press Release, Bd. of Governors of the Fed. Reserve Sys., Agencies Issue Final Rule to Implement Unlawful Internet Gambling Enforcement Act (Nov. 12, 2008), *available at* <http://www.federalreserve.gov/newsevents/press/bcreg/20081112b.htm>. In November 2009, the compliance date for the internet gambling rules was postponed for six months. *See* Press Release, Bd. of Governors of the Fed. Reserve Sys., Agencies Extend Compliance Date for Final Rule to Implement Unlawful Internet Gambling Enforcement Act (Nov. 27, 2009), *available at* <http://www.federalreserve.gov/newsevents/press/bcreg/20091127a.htm> [hereinafter Press Release, Agencies Extend Compliance Date for Final Rule].

153. Internet Gambling Regulation, Consumer Protection, and Enforcement Act, H.R. 2267, 111th Cong. (2009).

154. *Id.* §§ 5383, 5385.

155. *Id.* § 5383.

156. *See id.* § 5385 (implying blocking by implication).

United States in the short run, but the long term ramifications could be very complex. For instance, websites authorized to engage in internet gambling by the U.S. Secretary of the Treasury might not be authorized by other countries. These other countries could seek to enforce Congressman Frank's proposed law by limiting the ability of their banks to accept any internet gambling transactions. Moreover, other countries might have their own legal requirements for registering and regulating internet gambling establishments. If these countries decide to make financial intermediaries responsible for enforcing their restrictions, then the payment network systems would have to be updated to deal with their policies. Countries could seek to build on the payment systems in the hopes of using it to enforce their local rules regarding internet gambling. But a system that works well for a few countries in the short term could easily collapse if each country attempts to use it for the purpose of enforcing local rules.

None of these issues are imminent, however, and the new licensing regime proposed in Congressman Frank's legislation would be an improvement over the existing system in the short term. But over time the only way payment systems can operate is through a reduction in the diversity of the laws they must accommodate. This suggests that the government either seeks other ways to enforce its local laws or begins the process of harmonizing its laws. In the case of internet gambling, one solution would be an international agreement recognizing licensing arrangements in different countries as long as they satisfy certain agreed upon minimum standards.<sup>157</sup>

## B. CHILD PORNOGRAPHY, CONTROLLED SUBSTANCES, AND ONLINE TOBACCO

### 1. *Child Pornography*<sup>158</sup>

Payment intermediaries and most other internet intermediaries have explicit policies against child pornography.<sup>159</sup> This section examines more

---

157. The postponed compliance date, Press Release, Agencies Extend Compliance Date for Final Rule, *supra* note 152, might provide time for such an agreement to be drafted and adopted.

158. Portions of this Section are taken directly from the author's congressional testimony, as cited.

159. Google bans child pornography. Making the Internet Safe for Kids: The Role of ISP's and Social Networking Sites: Hearing Before the Subcomm. on Oversight and Investigations of the H. Comm. on Energy and Commerce, 109th Cong. 137 (2006) (testimony of Nicole Wong, Associate General Counsel and Chief Privacy Officer, Google Inc.). Nicole Wong stated,

As a company, Google is deeply committed to protecting children on the Internet in our actions and in our guiding principles. Child pornography is a horrific and vicious crime and has no place in a civilized society. Google has a zero-tolerance policy for child pornography and those who would

specifically how payment intermediaries take steps to prevent commercial child pornography transactions.<sup>160</sup>

Card payment systems work closely with law enforcement to remove any merchants in their systems that are involved in child pornography. In addition, they actively screen merchants for this illegal activity without a legal compulsion to do so and without waiting for complaints from law enforcement or other third parties.<sup>161</sup>

Specifically, card payment systems enforce their prohibition against child pornography with a two-part program.<sup>162</sup> The first part is a set of due diligence requirements designed to prevent child pornography merchants from entering payment systems. The second part is a monitoring program to detect and expel any child pornography merchants that manage to fraudulently enter the systems.<sup>163</sup>

promote it. When we become aware of child pornography anywhere in our search index or on our site, we remove it immediately and report it to the appropriate authorities. We do not accept any advertising related to it. We cooperate assiduously with law enforcement to help track down online criminals and child predators.

*Id.* Visa explicitly bans child pornography from its payment system, stating in its International Operating Regulations,

An Acquirer must both: Ensure that a Merchant, Internet Payment Service Provider (IPSP), or Sponsored Merchant that displays a Visa-Owned Mark on its Website does *not* accept Cards for the purchase or trade of child pornography[;] and [t]erminate a Merchant, IPSP, or Sponsored Merchant within 7 calendar days of Notification from Visa if the Merchant is identified as engaging in the purchase or trade of child pornography[.]

VISA, *supra* note 21, at 4.1.C.5.b (emphasis in original). MasterCard also bans child pornography as its general rule against illegal transactions applies to “[t]he sale of a product or service, including an image, which is patently offensive and lacks serious artistic value (such as, by way of example and not limitation, images of . . . sexual exploitation of a minor . . .). . . .” MASTERCARD, *supra* note 21, at 5.9.7.

160. This account of Visa’s policies and procedures on child pornography is based on my Congressional testimony. *Deleting Commercial Pornography Sites from the Internet: The U.S. Financial Industry’s Efforts to Combat This Problem: Hearing Before the Subcomm. on Oversight and Investigations of the H. Comm. on Energy and Commerce*, 109th Cong. 70–72 (2006) [hereinafter *Deleting Commercial Pornography Sites From the Internet Hearing*] (statement and testimony of Mark MacCarthy, Senior Vice President, Public Policy, Visa, U.S.A., Inc.). American Express, MasterCard, and PayPal have similar policies and procedures. *See id.* at 51–58 (statement of Arne L. Christenson, Senior Vice President, Federal Governmental Affairs, American Express Company); *id.* at 60–63 (statement of Jodi Golinsky, Vice President & Senior Regulatory Counsel, MasterCard International, Inc.); *id.* at 66–67 (statement of Joe Sullivan, Associate General Counsel, PayPal, Inc.).

161. *Id.* at 71 (statement and testimony of Mark MacCarthy, Senior Vice President, Public Policy, Visa, U.S.A., Inc.).

162. *Id.* at 70.

163. *Id.* at 71.

The card systems require financial institutions that are part of their payment network to ensure that all merchants are properly qualified to accept payment cards.<sup>164</sup> This normally involves a determination that a prospective merchant is financially responsible, and will abide by system requirements and applicable law.<sup>165</sup>

By taking these precautions, financial institutions can provide a line of defense against child pornography merchants entering the payment system. These due diligence requirements are closely observed by financial institutions, but they are not a panacea for addressing the problem. Child pornography merchants do not present themselves as such to financial institutions. They often appear to be legitimate merchants. They use a variety of techniques to fool financial institutions, and thereby gain access to the payment systems, despite the best efforts of these financial institutions to screen them out.<sup>166</sup>

Accordingly, the payment systems supplement these due diligence requirements with an active monitoring system. Traditional payment networks maintain separate monitoring campaigns to identify and eliminate transactions from child pornography merchants. Visa's program, for example, began in 2002.<sup>167</sup> Visa has retained the services of an outside firm to search the Internet for child pornography websites that appear to be accepting Visa payment cards. This firm uses advanced web crawling and filtering technology to detect these websites. It looks for websites that display the Visa logo, and satisfy one or more indicators that they are engaged in the sale of child pornography or are marketing themselves as engaged in that business. The sweeps are ongoing; they are conducted daily and search hundreds of millions of webpages each month.<sup>168</sup>

It is important to emphasize that the payments systems are in no better position than anyone else to detect child pornography sites that appear to use their payment systems. Any party could do this. Nothing internal to the system identifies a transaction as pertaining to child pornography. This creates the possibility, to be discussed later, of achieving greater efficiency by having a common search program rather than duplicative individual detection efforts.

---

164. *Id.*

165. There are a variety of methods that financial institutions may use to determine these qualifications, such as reviewing credit reports, business financial statements, and income tax returns, conducting physical inspections of the business premises of prospective brick and mortar merchants, and obtaining a detailed business description of electronic commerce merchants and examining merchants' websites. *Id.*

166. *Id.*

167. *Id.*

168. *Id.* at 72.

The payment systems do the monitoring themselves because it helps them take the next step: determining which of the apparently active child pornography websites actually accept the cards for transactions. When the outside search firm detects one of these problematic sites, it conducts test transactions to see whether the site is actually accepting the brand's cards or whether they are merely using the trademark illegally on their site. The search firm tells the card network immediately if it finds a site that is accepting the network's cards for these transactions. Unless requested by law enforcement to leave these sites open,<sup>169</sup> the network then contacts any financial institution found to be processing these child pornography transactions and directs it to stop processing these transactions immediately. If the financial institutions have not done so within seven calendar days, they are fined.<sup>170</sup>

If these identified sites are not actually accepting the payment cards, but are merely using the trademarks on their site, the networks use their best efforts to locate the web hosting companies to direct them to remove the logo.<sup>171</sup> In addition, the payment networks, such as Visa, provide information regarding all these sites to U.S. and international law enforcement officials.

These individual efforts against child pornography have been supplemented by collective action. In July 2005, Senator Richard Shelby, then Chairman of the Senate Banking Committee, convened a meeting involving the National Center for Missing & Exploited Children (NCMEC), the International Centre for Missing & Exploited Children (ICMEC), and key financial industry leaders to encourage the private sector to work together to attack the problem of commercial child pornography.<sup>172</sup> In March 2006, payment systems and financial institutions joined with the NCMEC to form the Financial Coalition Against Child Pornography and announced the formation of this group at a press conference attended by Senator Shelby.<sup>173</sup> The group shares information and takes collective action against child pornography merchants identified by complaints to the NCMEC hotline or from internet searches. This effort has produced some positive results in disrupting the activities of child pornographers and pushing bad actors away

---

169. If requested, the networks do allow problematic sites to remain operational as part of an ongoing law enforcement investigation. *Id.*

170. *Id.*

171. *Id.*

172. *Combating Child Pornography by Eliminating Pornographers' Access to the Financial Payment System: Hearing Before S. Comm. on Banking, Housing, and Urban Affairs*, 109th Cong. 51 (2006) [hereinafter *Combating Child Pornography Hearing*] (statement of Ernie Allen, President and Chief Executive Officer, National Center for Missing & Exploited Children).

173. *Id.* at 97 (statement of Mark MacCarthy, Senior Vice President, Public Policy, Visa USA, Inc.).

from recognized payment brands and toward less traditional payment mechanisms.<sup>174</sup>

The effectiveness of the payment card campaign is reflected in industry statistics. For instance, in August 2006, the search firm working for Visa examined over eleven million internet sites a day and found two child pornography sites that accepted Visa cards.<sup>175</sup> Since the beginning of 2006, nine such sites had been identified.<sup>176</sup> All of these sites were quickly expelled from the Visa system.<sup>177</sup>

In summary, the process developed by the payment systems to control child pornography transactions consists of pro-active monitoring followed by aggressive efforts to expel child porn merchants from the system. No legal compulsion exists to take this action, but it is done in close cooperation with law enforcement and other private bodies.

## 2. *Controlled Substances*<sup>178</sup>

Internet intermediaries also take extra precautions to protect against the use of their systems for traffic of controlled substances.<sup>179</sup> Payment systems treat controlled substances similar to child pornography. They have a proactive policy of screening merchants and monitoring transactions to prevent the use of their systems for transactions involving controlled substances. They distinguish controlled substances from other prescription drugs and take special precautions to ensure that websites selling controlled substances without proper authorization are expelled from their systems, while allowing legitimate internet pharmacies to function normally.<sup>180</sup>

---

174. See *Deleting Commercial Pornography Sites From the Internet Hearing*, *supra* note 160, at 30 (statement of Ernie Allen, President and Chief Executive Officer, National Center for Missing & Exploited Children) (“We are seeing indications of a trend toward directing buyers away from credit cards and toward alternative payment methods to make the actual transaction.”). Allen also stated that “we are seeing that the credit card logos we are finding on these sites in most cases do not lead you to an actual account.” *Id.* at 27.

175. *Id.* at 69 (testimony of Mark MacCarthy, Senior Vice President, Public Policy, VISA U.S.A., Inc.).

176. *Id.*

177. *Id.*

178. Portions of this Section are taken directly from the author’s congressional testimony, as cited.

179. For example, Microsoft, Yahoo!, and Google all require U.S. based pharmaceutical advertisers to be registered with PharmacyChecker. PharmacyChecker.com Verification Program, <https://www.pharmacychecker.com/sealprogram/choose.asp> (last visited Feb. 4, 2010) (“Google, Yahoo! and Microsoft adCenter *require* all advertisers and their affiliates who sell prescription drugs (as well as advertisers who refer visitors to prescription drug selling sites) to be approved through the PharmacyChecker Verification Program. The advertised pharmacy must also be based in the U.S. or Canada.”).

180. This summary of the policies and procedures used to combat controlled substances

Payment systems' efforts to prevent the use of their networks for transactions involving controlled substances were developed in conjunction with U.S. law enforcement agencies.<sup>181</sup> In 2004 and 2005, payment system representatives met with the Food and Drug Administration (FDA) and the Drug Enforcement Administration (DEA) to develop strategies to deal with the problem of controlled substances sales on the Internet.<sup>182</sup> The health and safety risks associated with these transactions and the likelihood that these dangerous pharmaceuticals could be obtained by minors led the networks to take active steps.<sup>183</sup>

Coding the nature of the transaction so that financial institutions could block as necessary, which is effective for gambling transactions, does not work in the instance of controlled substances. Available codes can only identify the business of the website, not the nature of the pharmaceuticals dispensed. Many legitimate websites sell pharmaceuticals under appropriate government regulation and private sector oversight. If the business code were used as a trigger to block these transactions, it would eliminate a substantial and desirable internet business as well.

The only effective approach is a program of due diligence supplemented by an active monitoring program. Payment systems refined these programs in cooperation with the FDA and DEA. For example, as part of its due diligence program in 2004 and 2005, Visa reminded affiliated financial institutions of their responsibility to ensure that only legal transactions enter the Visa Payment System and directed the affiliates' attention to the lists of controlled substances and problematic drugs on the FDA and DEA websites.<sup>184</sup> Visa also directed its members to the public safety bulletins on the FDA website about buying medicines online.<sup>185</sup> Visa noted that a safe website should be licensed by the state board of pharmacy where the website is operating, have a licensed pharmacist available to answer questions, require a prescription from a U.S. licensed doctor or other healthcare professional

---

is based on my congressional testimony. *Safety of Imported Pharmaceuticals: Strengthening Efforts to Combat the Sales of Controlled Substances over the Internet: Hearing Before the Subcomm. on Oversight and Investigations of the H. Comm. on Energy and Commerce*, 109th Cong. 174–76 (2005) [hereinafter *Safety of Imported Pharmaceuticals Hearing*] (statement of Mark MacCarthy, Senior Vice President, Public Policy, VISA, U.S.A., Inc.). MasterCard has a similar program in place. *Id.* at 178–81 (statement of Michael McEneny, Partner, Sidley, Austin, Brown, & Wood, LLP).

181. *Id.* at 174 (statement of Mark MacCarthy, Senior Vice President, Public Policy, VISA, U.S.A., Inc.).

182. *Id.*

183. *Id.* at 214–15 (testimony of Andrew McLaughlin, Senior Policy Counsel, Google, Inc.).

184. *Id.* at 175 (statement of Mark MacCarthy, Senior Vice President, Public Policy, VISA, U.S.A., Inc.).

185. *Id.*

licensed in the United States to write prescriptions, and provide a way to speak to a person about problems.<sup>186</sup> Visa also advised its members to consider relying on a reputable seal program, such as the Verified Internet Pharmacy Practices Site Program operated by the National Association of Boards of Pharmacy, as a means of identifying reputable internet pharmacies.<sup>187</sup> When alerted that specific internet pharmacies may be accepting Visa cards for illicit transactions, Visa worked to investigate these pharmacies and to terminate the acceptance of Visa cards for illicit activity.<sup>188</sup>

As in the case of child pornography, these due diligence efforts are necessary but not sufficient. The payment networks each retain the services of an outside firm to search the Internet for websites selling controlled substances and accepting the networks' payment cards.<sup>189</sup> This program builds on the efforts to monitor the Internet for child pornography, using the same web crawling and filtering technology and the same outside search firm to conduct sweeps. This vendor looks for websites that display the card's brand logo, that sell Schedule II controlled substances or other prescription drugs that the FDA or DEA have indicated are especially dangerous, and that do not require a prescription or an exam. The sweeps are ongoing; they are conducted daily and search hundreds of millions of webpages each month. The networks then take steps to remove merchants who appear to be selling controlled substances.<sup>190</sup>

These efforts have produced positive results. In 2005, MasterCard indicated that they had located and shut down 500 websites selling illegal controlled substances.<sup>191</sup> The National Center on Addiction and Substance Abuse (CASA) at Columbia University has studied the nation's problem of controlled prescription drug abuse and has documented the internet availability of these drugs. Its early reports described what looked like a steady increase in the availability of controlled substances.<sup>192</sup> In 2008, however, it documented a decline in the number of websites advertising or

---

186. *Id.*

187. *Id.*

188. *Id.*

189. *Id.*

190. *Id.*

191. Visa's program shut down forty-nine similar sites. For statistics from both payment networks, see *id.* at 208 (testimony of Mark MacCarthy, Senior Vice President, Public Policy, VISA, U.S.A., Inc.).

192. See *Rogue Online Pharmacies: The Growing Problem of Internet Drug Trafficking: Hearing Before the S. Comm. on the Judiciary*, 110th Cong. 7–9 (2007) [hereinafter *Rogue Online Pharmacies Hearing*] (statement of Joseph A. Califano, Jr., Chairman and President, The National Center on Addiction and Substance Abuse at Columbia University) (summarizing the early studies' findings that “[o]ver the 4-year course of our analysis, the number of selling sites has climbed from 154 [in 2004 and 2005] to 187 [in 2007]”).

selling controlled substances, and suggested that this decline could be attributed to financial service company efforts to block controlled substance transactions.<sup>193</sup>

### 3. *Online Tobacco*

Sales of tobacco products online are controversial for many reasons, including the inability to control sales to minors. For this reason, many internet intermediaries restrict the way their systems can be used in connection with online tobacco sales.<sup>194</sup>

A series of coordinated steps by the Bureau of Alcohol, Tobacco and Firearms (BATF) and the state attorneys general in 2004 and 2005 led to changes in the way payment intermediaries processed online tobacco sales.<sup>195</sup> These efforts by state and federal law enforcement officials were not couched as demands to end illegal action by the intermediaries themselves. Rather, the idea was that the underlying activity was illegal, and once this was

193. Press Release, Nat'l Ctr. on Addiction & Substance Abuse at Columbia Univ., New CASA Report Finds: Most Web Sites Selling Prescription Opioids, Stimulants and Depressants Require No Prescription; Some Sites Now Sell Prescriptions and Online "Medical Consultations" to Get Controlled Drugs 1 (July 9, 2008), *available at* <http://www.casacolumbia.org/absolutenm/templates/PressReleases.aspx?articleid=531&zoid=66> ("The new White Paper reports that CASA researchers found a total of 365 Web sites advertising or selling controlled prescription drugs during 210 hours of research in the first quarter of 2008, compared to 581 sites during the same period in 2007."). The report noted that the "decline in the number of Web sites advertising or selling controlled prescription drugs may reflect efforts of federal and state agencies and financial institutions to crack down on Internet drug trafficking." *Id.* (quoting Joseph A. Califano, Jr., CASA's Chairman and President). The large number of websites offering to sell controlled substances does not indicate that all of them actually engage in that activity. CASA offered to Visa, MasterCard, American Express, and PayPal a sample of forty-five anchor sites from their analysis that offered to sell controlled prescription drugs and indicated that they accepted payment from one or more of these payment systems. NAT'L CTR. ON ADDICTION AND SUBSTANCE ABUSE AT COLUMBIA UNIV., "YOU'VE GOT DRUGS!" V: PRESCRIPTION DRUG PUSHERS ON THE INTERNET 13 (2008), *available at* <http://www.casacolumbia.org/articlefiles/531-2008%20You%27ve%20Got%20Drugs%20V.pdf>. Test transactions by the payment systems revealed that only four of these sites actually attempted to process a transaction. *Id.* The report acknowledges that this "could be the result of efforts made by these financial service providers to shut down use of their systems of payment for Internet trafficking." *Id.*

194. For instance, Google does not accept online advertisements for tobacco products. *See* Google AdWords, Advertising Policies: Tobacco and Cigarettes, <http://adwords.google.com/support/aw/bin/static.py?page=guidelines.cs&topic=all&answer=47228&country=US&adtype=text> (last visited Mar. 25, 2010) ("Advertising is not permitted for the promotion of tobacco or tobacco-related products, including cigarettes, cigars, tobacco pipes, rolling papers, electronic cigarettes, and e-cartridge cigarettes.>").

195. Consumeraffairs.com, Credit Card Companies Snuff Online Tobacco Sales (Mar. 17, 2005), [http://www.consumeraffairs.com/news04/2005/tobacco\\_ag.html](http://www.consumeraffairs.com/news04/2005/tobacco_ag.html).

brought to the attention of the intermediaries, they would respond by taking voluntary measures to stop the illegal transactions. Because all payment intermediaries have a general rule against allowing their systems to be used for illegal activity, it seemed reasonable that they would take this step. In fact, almost all of them did.<sup>196</sup>

In January 2005, forty-two state attorneys general wrote to the payment card networks informing them that virtually all online tobacco retailers engage in illegal sales.<sup>197</sup> They listed the laws they believed were violated and requested that the payment networks not allow their cards to be used for online tobacco product purchases until retailers proved that their sales were not in violation of state or federal laws.<sup>198</sup> They also asked that the networks take appropriate steps to ensure that their credit cards were not used to facilitate violations of state or federal laws.<sup>199</sup>

In March 2005, after several joint meetings, the card companies reached an agreement with the state attorneys general and with BATF to take steps against the online sale of tobacco products, including adopting policies prohibiting use of their cards for the illegal online sale of cigarettes and taking action against any such sellers identified by law enforcement.<sup>200</sup>

MasterCard reacted by sending a notice to their member banks requiring them to cease card acceptance for internet tobacco, or prove to the satisfaction of BATF and the relevant state attorneys general that they were in compliance with relevant laws.<sup>201</sup> Unlike the case of child pornography and

---

196. *Id.*

197. *Id.*

198. *Id.*

199. *Id.*

200. There was only a release; the agreement itself was verbal. Press Release, Office of the Att’y Gen., State of N.Y., State AGs and ATF Announce Initiative with Credit Card Companies to Prevent Illegal Internet Cigarette Sales (Mar. 17, 2005), *available at* [http://www.ag.ny.gov/media\\_center/2005/mar/mar17b\\_05.html](http://www.ag.ny.gov/media_center/2005/mar/mar17b_05.html). The release announcing this agreement stated,

Among the many actions the credit card companies have adopted to stop illegal online sales are:

- 1.) Adopting policies to prohibit the use of credit cards for the illegal sale of cigarettes over the Internet; and,
- 2.) Agreeing to investigate and take action with respect to any internet sellers identified by law enforcement agencies as using credit cards for illegal online cigarette sales.

*Id.*

201. *MasterCard Urges Merchant Compliance with Rules Governing the Internet Sale of Tobacco*, BUS. WIRE, Mar. 8, 2005, *available at* WestLaw, 3/8/05 BWIRE 15:00:00 (“MasterCard said that financial institutions can continue to provide MasterCard acceptance for internet tobacco sales if they have documented evidence to substantiate that the merchant is in compliance with all applicable federal, state, and local laws to the satisfaction of ATF and

controlled substances, however, the payment networks relied on complaints from law enforcement, instead of conducting their own investigations, to direct their efforts to stop the illegal activity. As a result of these steps, the number of online tobacco merchants declined dramatically. MasterCard estimated that they had cut off the 100 largest online sites.<sup>202</sup>

In addition, the state attorneys general sought and obtained the cooperation of the carriers that delivered cigarettes from online tobacco stores to purchasers. In July 2005, DHL agreed to stop delivering cigarettes for illegal internet sellers.<sup>203</sup> In October 2005, UPS agreed to stop shipping cigarettes to consumers throughout the United States.<sup>204</sup> In February 2006, FedEx agreed to stop shipping cigarettes from online stores.<sup>205</sup> The U.S. Postal Service rejected the request to stop delivering cigarettes from online tobacco stores, citing federal requirements that it deliver the mail.

#### 4. *Assessment of Child Pornography, Controlled Substances, and Online Tobacco*

Child pornography, controlled substances, and online tobacco can be discussed together because they raise common issues. They all involve extensive cooperation between payment systems and federal government agencies policing criminal activity on the Internet. As discussed in the previous sections, payment networks monitor their systems for transactions involving child pornography and controlled substances, and share the results with law enforcement. In the case of online tobacco, they react to complaints brought to them by law enforcement. These enforcement arrangements emerged in the absence of any legal compulsion.

The assessment of public policies in this area follows the framework outlined in Part II. The analysis of equities suggests that payment

any applicable State Attorney General.”).

202. Bob Tedeschi, *E-Commerce Report; Now that Credit Card Companies Won't Handle Online Tobacco Sales, Many Merchants Are Calling It Quits*, N.Y. TIMES, Apr. 4, 2005, at C5; GOLDSMITH & WU, *supra* note 7, at 76–77; Mann & Belzley, *supra* note 8, at 247; *see also* COMM. ON REDUCING TOBACCO USE: STRATEGIES, BARRIERS & CONSEQUENCES, INST. OF MED. OF THE NAT'L ACADS., ENDING THE TOBACCO PROBLEM: A BLUEPRINT FOR THE NATION 670 (Richard J. Bonnie et al. eds., 2007), *available at* <http://www.nap.edu/openbook.php?isbn=0309103827>.

203. Press Release, Office of the Att'y Gen., State of N.Y., *Leading Package Delivery Company Agrees to Stop Shipping Cigarettes to Individual Consumers* (July 5, 2005), *available at* [http://www.ag.ny.gov/media\\_center/2005/jul/jul05a\\_05.html](http://www.ag.ny.gov/media_center/2005/jul/jul05a_05.html).

204. Press Release, Office of the Att'y Gen., State of N.Y., *UPS Joins Effort to Reduce Youth Smoking* (Oct. 24, 2005), *available at* [http://www.ag.ny.gov/media\\_center/2005/oct/oct24a\\_05.html](http://www.ag.ny.gov/media_center/2005/oct/oct24a_05.html).

205. Press Release, Office of the Att'y Gen., State of N.Y., *FedEx to Strengthen Policies Restricting Cigarette Shipments* (Feb. 7, 2006), *available at* [http://www.ag.ny.gov/media\\_center/2006/feb/feb07a\\_06.html](http://www.ag.ny.gov/media_center/2006/feb/feb07a_06.html).

intermediaries have a general responsibility to take affirmative action in this area, but the market analysis suggests that they are already living up to these responsibilities. There is no need for legal compulsion. The costs of additional legal burdens are likely to be higher than any benefits from the increase in their efforts to control these illegal activities.

The degree of control exercised by payment intermediaries in these cases is no greater than in their provision of services to any merchant. If all that is relevant is the extent of actual control that payment intermediaries have against online purveyors of child porn and controlled substances, then payment intermediaries would not have affirmative responsibilities to act in these areas. But the degree of harm imposed by a merchant's activity is also relevant. The public health harm created by child pornography and controlled substances sales is substantial.<sup>206</sup> While this is undoubtedly a normative view, it is one that is widely, if not universally, shared. As a result, payment systems should take positive steps in this area to prevent the use of their systems for these purposes. System monitoring is needed because it is the most effective way to catch these bad actors if they manage to slip into the system.

Because of the public health issues involved, payment systems also owe a duty to respond in the area of online tobacco sales. In this case, however, the payment networks have calibrated their response to the needs of law enforcement. Law enforcement is able to efficiently investigate cases and bring them to the attention of the payment systems. Responsiveness to these complaints, rather than pro-active system monitoring, is acceptable.

Intermediaries are cooperating fully with the relevant government agencies to satisfy these general obligations. As discussed above, the steps that the payment card industry has taken to cooperate with law enforcement have successfully controlled commercial online child pornography, illegal sales of controlled substances, and online tobacco sales. There is no need for additional regulatory requirements.<sup>207</sup>

---

206. For harms caused by the internet sales of controlled substances, see *Rogue Online Pharmacies Hearing*, *supra* note 192, at 10–11 (statement of Philip B. Heymann, James Barr Ames Professor of Law, Harvard Law School). For the extent of child pornography, see *Combating Child Pornography Hearing*, *supra* note 172, at 25–26 (testimony of Ernie Allen, President and Chief Executive Officer, National Center for Missing and Exploited Children). See also National Center for Missing and Exploited Children, What is Child Pornography?, available at [http://www.missingkids.com/missingkids/servlet/PageServlet?LanguageCountry=en\\_US&PageId=1504](http://www.missingkids.com/missingkids/servlet/PageServlet?LanguageCountry=en_US&PageId=1504) (last visited Mar. 5, 2010).

207. Mann and Belzley note that regulators in a variety of contexts have reached informal agreements with intermediaries in which intermediaries voluntarily agree to cooperate. Most of those agreements seemingly do not reflect the view of the intermediaries that they could be forced in litigation to provide that

Despite the success of these voluntary efforts, Congress has proposed legal requirements on payment systems to control child pornography and controlled substances. Proposed legislation mandates further efforts by intermediaries to prevent the use of their systems for child pornography.<sup>208</sup> For financial intermediaries, the Bill would impose fines and prison sentences on anyone who “knowingly conducts, or attempts or conspires to conduct, a financial transaction . . . knowing that such transaction will facilitate access to, or the possession of, child pornography . . . .”<sup>209</sup> For other intermediaries, the Bill would impose fines and prison terms on any web hosting company or email provider who “knowingly engages in any conduct the provider knows or has reason to believe facilitates access to, or the possession of, child pornography.”<sup>210</sup> The Bill also imposes a two year record keeping requirement on ISPs and others to facilitate child pornography investigations.<sup>211</sup> In light of the extensive efforts already undertaken by intermediaries in this area, it is unlikely that these requirements will produce any additional net benefit.

Congress passed legislation regulating online pharmacies in 2008.<sup>212</sup> This law imposes obligations on U.S. based internet pharmacies regarding controlled substances and prohibits the sale of controlled substances except as specifically authorized.<sup>213</sup> The law requires that certain intermediaries must not “knowingly or intentionally . . . aid or abet” any unauthorized sale of controlled substances.<sup>214</sup> The statute defines such aiding or abetting as “serving as an agent, intermediary, or other entity that causes the Internet to be used to bring together a buyer and seller to engage in the [unauthorized]

---

cooperation, but rather the view that a failure to cooperate would result in formal legislative regulation: the settlements proceed not in the shadow of existing law, but in the shadow of potential law.

Mann & Belzley, *supra* note 8, at 260 n.59. But if the pressure to cooperate “in the shadow of potential law” has worked, why actually legislate?

208. Internet Stopping Adults Facilitating the Exploitation of Today’s Youth (SAFETY) Act of 2009, H.R. 1076, 111th Cong. (2009).

209. *Id.* § 2.

210. *Id.* § 3. The Center for Democracy and Technology (CDT) describes the problem with this approach. “The problem is that any major national service provides knows—as almost a statistical certainty—that *someone* is using their services to ‘facilitate’ access to child pornography.” Memorandum from John Morris & Gregory T. Nojeim, Ctr. for Democracy & Tech. to Interested Persons 1 (June 18, 2009) (on file with author) (emphasis in original).

211. *See* H.R. 1076, § 5. CDT’s analysis of this provision notes the difficulties record retention would create for ISPs and the tension with privacy concerns. Memorandum from John Morris & Gregory T. Nojeim, *supra* note 210, at 2–3.

212. Ryan Haight Online Pharmacy Consumer Protection Act of 2008, Pub. L. No. 110-425, 122 Stat. 4820 (codified in 21 U.S.C. §§ 829, 802).

213. *Id.* § 3(f)(1)(A).

214. *Id.* § 3(f)(1)(B).

dispensing of a controlled substance . . . .”<sup>215</sup> It is not clear that these requirements apply to payment systems.<sup>216</sup> Nonetheless, it is very likely that current financial intermediary processes would easily satisfy these general requirements if they applied.

As in the case of internet gambling, the new online pharmacy bill carried over the exemptions from liability first developed in § 230 of the Communications Decency Act for pure internet intermediaries.<sup>217</sup> The duties imposed on third parties not to aid or abet illegal controlled substance transactions do not apply to

the provision of a telecommunications service, or of an Internet access service or Internet information location tool . . . or . . . the transmission, storage, retrieval, hosting, formatting, or translation (or any combination thereof) of a communication, without selection or alteration of the content of the communication.<sup>218</sup>

This exemption from liability appears justified for these parties because they are far removed from the actual causation of the illicit transactions, their involvement is passive, and they undertake substantial efforts to control the sale of controlled substances online. This exemption should also be extended to payment systems, given their substantial role in fighting these illegal activities in conjunction with law enforcement.

On cost grounds, however, the current arrangements could be improved. Payment systems have no particular expertise in monitoring their own systems for child pornography and controlled substances. The networks have outsourced their system monitoring efforts, and some efficiency could be achieved by combining these efforts, in coordination with appropriate law enforcement.<sup>219</sup>

---

215. *Id.* § 3(f)(2)(C).

216. See Sarah Rubenstein, *New Bill Targets Rogue Druggists on the Internet*, WALL ST. J., Oct. 9, 2008, at D1 (“Finally, the bill does not create new requirements for Internet search engines, credit-card companies or package-delivery concerns whose services are used in online pharmacy transactions.”).

217. Ryan Haight Online Pharmacy Consumer Protection Act of 2008, Pub. L. No. 110-425, § 3(f)(3)(A)(iii), 122 Stat. 4820 (codified in 21 U.S.C. §§ 829, 802).

218. *Id.* § 3(f)(3)(A)(iii).

219. In the area of controlled substances, some groups have called for mechanisms to supplement the efforts of third parties with a special government-funded monitoring entity. See, e.g., *Rogue Online Pharmacies Hearing*, *supra* note 192, at 11 (statement of Philip B. Heymann, James Barr Ames Professor of Law, Harvard Law School). This monitoring entity would send information to payment card companies and other intermediaries, which would then trigger an automatic legal requirement to investigate and block. *Id.* This is a system of indirect liability with the obligation to act triggered by the activity of a non-governmental private party. It is not necessary and would not improve the efficiency of the existing monitoring system. Centralizing the monitoring function in cooperation with law

A different mechanism for improving current enforcement efforts would be a list of companies licensed to sell controlled substances on the Internet. Monitoring by payment card companies would be greatly simplified if the U.S. government maintained a list of websites, domestic and international, that are properly licensed to sell controlled substances.<sup>220</sup> This improvement, however, would require substantial international coordination to be effective.

### C. ONLINE COPYRIGHT INFRINGEMENT

When payment intermediaries take steps against internet gambling merchants, child pornographers, purveyors of controlled substances, and online tobacco merchants, they are acting together with law enforcement officials. The information they gather in the searches of their systems are provided to law enforcement, and they react to complaints brought to them by law enforcement agencies. In contrast, with copyright cases, the complaining party is a private party alleging that some other party has harmed them by infringing on their copyright.

On its face, this is a very awkward place for intermediaries to be. In copyright infringement cases, private parties dispute their respective rights under the law. No adjudication has been made on the merits of the case. In some instances, no legal assertion of rights is made at all to the allegedly infringing party. Why should another private party simply take one side of the dispute, and use whatever relationship they have with one of the parties to enforce the other party's rights? The third party may not know who is right. And if it does take action, but chooses the wrong side, the aggrieved party may pursue the third party for taking wrongful action.

The ideal would be for copyright owners to sue direct infringers. But direct infringers are sometimes too ubiquitous, too small, or too difficult to find. The result is well-developed notions of secondary liability for copyright infringement that involve intermediaries.<sup>221</sup> These doctrines of secondary liability have evolved substantially over the past decades.

#### 1. *Legal Context for Intermediary Liability in Copyright Infringement*

Court cases and federal statute define some indirect responsibilities of intermediaries regarding copyright. The 1984 Supreme Court decision in *Sony Corp. of America v. Universal City Studios, Inc.*<sup>222</sup> established a standard for assessing third party liability. Providers of a technology that can be used for

---

enforcement, however, would create some savings.

220. See *Safety of Imported Pharmaceuticals Hearing*, supra note 180, at 206 (statement of Michael McEnaney, Partner, Sidley, Austin, Brown, & Wood, LLP).

221. For a brief discussion of vicarious and contributory liability in copyright law, see Lichtman & Landes, supra note 39.

222. 464 U.S. 417 (1984).

infringing activities are not liable when there are “substantial non-infringing uses” of the technology.<sup>223</sup> The DMCA enabled copyright owners to enforce their existing rights in the Internet context by enlisting the help of internet intermediaries.<sup>224</sup> The key mechanism for gaining the cooperation of intermediaries is a safe harbor from secondary liability. ISPs are given an exemption from secondary liability so long as they act as a pure conduit, providing only transitory communications and system caching.<sup>225</sup> Web hosts and search engines also receive a safe harbor, provided they comply with a specific notice-and-takedown procedure.<sup>226</sup> Upon receiving notification of claimed infringement, the provider must expeditiously take down or block access to the material.<sup>227</sup>

Successful litigation against peer-to-peer networks in the digital music area also increased the ability of copyright owners to use third parties to combat copyright infringement where the third party is affirmatively involved in fostering the infringement. In an early file-sharing case, the Ninth Circuit found that the peer-to-peer service Napster was liable for secondary infringement based on its control and facilitation of its users’ infringement of music copyrights.<sup>228</sup> The company subsequently went out of business in its original form.<sup>229</sup> More recently, the Supreme Court found that another peer-to-peer service, Grokster, violated federal copyright law when it took “affirmative steps . . . to foster infringement . . . by third parties,” such as advertising an infringing use or instructing how to engage in an infringing use.<sup>230</sup>

Against this background arose a question regarding payment systems: are they liable for secondary infringement when their payment systems are used for direct infringement? In *Perfect 10 v. Visa International Service Ass’n*,<sup>231</sup> a

223. *Id.* at 442.

224. 17 U.S.C. § 512 (2006).

225. 17 U.S.C. § 512(a).

226. 17 U.S.C. § 512(b).

227. *Id.*

228. *A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004 (9th Cir. 2001).

229. Benny Evangelista, *Napster Runs Out of Lives—Judge Rules Against Sale*, S.F. CHRON., Sept. 4, 2002, at B1.

230. *MGM Studios, Inc. v. Grokster, Ltd.*, 545 U.S. 913, 919 (2005).

231. 494 F.3d 788 (9th Cir. 2007); see Jonathan Band, *The Perfect 10 Trilogy*, 5 COMPUTER L. REV. INT’L 142 (2007) (discussing *Perfect 10 v. Visa International Service Ass’n* and its relationship to similar secondary liability cases). Band summarizes the Visa case:

Here the Ninth Circuit rejected what would have represented a significant expansion of secondary liability to actors far removed from the infringing activity. However, unlike the other cases, this case provoked a strong dissent by respected jurist Alex Kozinski. This dissent suggests that the outer edges of secondary liability remain to be defined.

*Id.* at 146. Judge Kozinski’s dissent is indeed stinging, but it also underestimates the burden

subscription adult-content website alleged that numerous websites based in several countries had stolen its proprietary images, altered them, and illegally offered them for sale online.<sup>232</sup> Visa did not deny payment services to the allegedly infringing sites in response to the complaints, and Perfect 10 brought a contributory and vicarious infringement action against Visa.<sup>233</sup> The Ninth Circuit affirmed the district court to reject liability for Visa.<sup>234</sup>

In *Perfect 10*, the Ninth Circuit dismissed the charge of contributory infringement by focusing on whether the card companies “materially contributed” to the infringement.<sup>235</sup> The court said the credit card companies did not materially contribute to the infringement because they had no “direct connection” to the infringement.<sup>236</sup> To have direct connection to the infringement they would have had to reproduce, display, or distribute the allegedly infringing works, which they did not do.<sup>237</sup> Payment services might make it more profitable to infringe, but they are too far removed in the causal chain that leads to the actual infringing acts for them to be described as making a material contribution.<sup>238</sup>

The court made a similar point about vicarious liability, finding that the card companies had no practical ability or right to prevent the infringing activity.<sup>239</sup> While credit card services can exert financial pressure on the infringing websites, they cannot stop the actual reproduction or distribution of the infringing images.<sup>240</sup>

In his dissent, Judge Kozinski rejected both arguments.<sup>241</sup> According to Judge Kozinski, the card companies were directly connected to the infringement because they provided payment services.<sup>242</sup> Without these payment services there would be no infringement.<sup>243</sup> The card companies had the contractual right to terminate illegal activity on their systems, as well as the practical ability to exert financial pressure to stop or limit the infringing activity.<sup>244</sup>

---

that secondary liability would place on intermediaries. *Id.*

232. *Perfect 10*, 494 F.3d at 793.

233. *Id.*

234. *Id.*

235. *Id.* at 796.

236. *Id.*

237. *Id.*

238. *Id.* at 797.

239. *Id.* at 803.

240. *Id.* at 804.

241. *Id.* at 810–11 (Kozinski, J., dissenting).

242. *Id.* at 811–12.

243. *Id.*

244. *Id.* at 816–17.

## 2. *Payment System Complaint Program*<sup>245</sup>

Even though payment intermediaries may not be required to take steps against online copyright infringement, they have chosen to do so.<sup>246</sup> This Section describes the policy behind this activity and one example of such action. The payment intermediaries go beyond their legal duty for a variety of business reasons. Being associated with illegal activity is harmful to their brands. Responding to complaints from reputable businesses about losses from illegal activities strengthens a payment intermediaries' reputation as a responsible business partner. It also lends credence to the intermediaries' oft-repeated assertions that their payment systems should not be used for illegal activities. By keeping it free of illegal activity, the payment networks promote trust in electronic commerce, a channel of commerce in which they have a competitive advantage over traditional payment mechanisms like cash and check.<sup>247</sup>

How did payment systems take on the challenge of responding to complaints of copyright infringement? Payment systems cannot monitor their networks for copyright law violations. They do not have the factual basis to conclude that a particular sale of a product is a violation of someone's copyright.<sup>248</sup> Many music downloads are perfectly legal transactions, but some are not. Distinguishing the two is often a complex factual and legal question which payment intermediaries do not have the expertise or ability to resolve.

For this reason a coding and blocking system like the one used to address internet gambling will not work.<sup>249</sup> Merchants' transactions are coded by business category, not legal status.<sup>250</sup> If financial institutions blocked transactions based on the business code for a music download, they would block a substantial number of legal transactions. Since copyright owners benefit from these legal transactions, they would not want an overly broad coding and blocking scheme.

---

245. Portions of this Section are taken directly from the author's congressional testimony, as cited.

246. See generally *International Piracy: The Challenges of Protecting Intellectual Property in the 21st Century: Hearing Before the Subcomm. on Courts, the Internet, and Intellectual Property of the H. Comm. on the Judiciary*, 110th Cong. 73–82 (2007) [hereinafter *International Piracy Hearing*] (statement of Mark MacCarthy, Senior Vice President for Global Public Policy, Visa Inc.) (providing this account of payment intermediaries and intellectual property).

247. *Id.* at 75.

248. *Id.* at 76.

249. See *supra* Section III.A (discussing the internet gambling coding and blocking scheme).

250. VISA MERCHANT CATEGORY CLASSIFICATION (MCC) CODES DIRECTORY, *supra* note 105.

A coding and blocking scheme for copyright would also be inappropriate because the international legal status of copyright differs from the international legal status of internet gambling. Internet gambling is illegal in some jurisdictions. For that reason, one could require merchants to properly code themselves and to allow financial institutions in countries where internet gambling is illegal to block these labeled transactions. But internet merchants involved in illegal intellectual property infringement typically violate the laws of most countries.<sup>251</sup> It would be highly inefficient for payment systems to allow a merchant to introduce a transaction into the system when the vast majority of financial institutions in the network would have to program their payment processing operations to reject these transactions.

Therefore, the best way to respond to complaints about infringing activity is not to require coding by the infringing merchant, but to prevent the merchant from entering the illegal transaction into the system or to restrict it to those few jurisdictions where it might be legal.

The payment systems have to react to complaints of copyright violations because they cannot monitor for them on their own. They cannot take proactive steps as they could in the case of child pornography or controlled substances.<sup>252</sup> The payment networks have thus developed policies and procedures to handle these complaints.<sup>253</sup> These complaints do not involve health and safety, but they pose a business problem for these companies, and the payment networks attempt to respond, especially in large magnitude cases.<sup>254</sup>

The complaint process starts when a business entity approaches a payment system with clear, documented evidence of illegal activity and adequately identifies the infringing internet merchant.<sup>255</sup> The business entity must provide substantiation that the activity is illegal and documentation that payment cards are actually being used for this illegal activity.<sup>256</sup>

The next step is to assess legality. This is easier if there is regulatory or judicial precedent establishing the illegality, but that is rare. If the buyer and the seller are in the same jurisdiction, this legal assessment can be relatively

251. *International Piracy Hearing, supra* note 246, at 77 (statement of Mark MacCarthy, Senior Vice President for Global Public Policy, Visa Inc.).

252. *See supra* Sections III.B.1, III.B.2.

253. *International Piracy Hearing, supra* note 246, at 77 (statement of Mark MacCarthy, Senior Vice President for Global Public Policy, Visa Inc.).

254. *Id.*

255. This Section describes the process at Visa, but other payment networks use a similar process. *See id.* at 85.

256. *Id.*

straightforward.<sup>257</sup> But these cases are not typical because companies tend to pursue domestic remedies for domestic cases. They usually come to payment networks only in cross-border cases, involving a merchant in one location and the customer in another. If the legal situation in both countries is the same, the legal assessment can be relatively uncomplicated. But when the merchant is in one jurisdiction, the customer is in another, and the laws are not the same or the legal situation in one country is not as clear as the legal situation in the other, the assessment is far more complex.<sup>258</sup>

After wrestling with these issues, the payment networks developed a policy for cross-border transactions: if a transaction would be illegal in either the jurisdiction of the merchant or the jurisdiction of the cardholder, the transactions should not be in the payment system.<sup>259</sup> In cases like copyright infringement, this means that merchants are responsible for making sure that the transactions they submitted to the payment system are legal in both their operating jurisdiction and the jurisdiction in which their customer is located.

The assessment of legality requires the payment network to determine whether the type of transaction would be illegal in either jurisdiction.<sup>260</sup> Since the facts and law involved are often complex, the payment networks are willing to take on only the clearest cases of copyright violation. Once they determine illegality, the payment providers do what they reasonably can to assist the complaining party. Since payment networks do not work directly with merchants, they typically try to locate the bank that has the merchant account, and providing the complaint to the bank involved usually resolves the issue.<sup>261</sup> In most cases, either the bank does not want the business and terminates the merchant, or it takes other action to bring the merchant into compliance.<sup>262</sup> If the bank does not take action, the payment networks can take further enforcement action against the bank.<sup>263</sup>

### 3. *Allofmp3.com*

In some instances, the merchant resists the enforcement efforts of payment systems, and insisting on the legality of the underlying activity, the merchant goes to a local court to vindicate its perceived rights under local law. This is what occurred in the *Allofmp3.com* case.

As part of its general policy on cross-border transactions, Visa concluded that it did not want this type of transaction—illegal downloads of music—in

---

257. *Id.* at 77.

258. *Id.* at 77–78.

259. *Id.* at 78.

260. *Id.*

261. *Id.*

262. *Id.*

263. *Id.*

its payment system.<sup>264</sup> This decision enabled it to extend enforcement actions against this one site to different sites or to the same site processed by a different bank.

In 2005, Visa received a documented complaint from International Federation of the Phonographic Industry (IFPI), which represents copyright owners based in more than seventy countries.<sup>265</sup> The complaint alleged that Allofmp3.com, a website located in Russia, was infringing on the copyrights of IFPI's members by allowing unauthorized downloads of music.<sup>266</sup> Visa assessed the legal situation, in part by obtaining a review by outside counsel, and concluded that the transactions were illegal under local Russian law.<sup>267</sup> They were also illegal under the laws of the vast majority of the merchant's customers who were located primarily in the United Kingdom and the United States.<sup>268</sup> In October 2005, the Italian authorities shut down a portal to Allofmp3.com, allofmp3.it, and began a criminal investigation of the Italian site.<sup>269</sup> In addition, the United States Trade Representative intervened with the Russian government to urge them to shut down Allofmp3.com.<sup>270</sup>

At the beginning of September 2006, after appropriate notice, the Russian bank working with Allofmp3.com stopped processing Visa transactions for Allofmp3.com.<sup>271</sup> At the end of September 2006, the bank also stopped processing transactions from an affiliated site called allTunes.<sup>272</sup> After these Visa transactions ended, further confirmation of the site's illegality was forthcoming; a Danish court ordered the internet provider Tele2 to block its subscribers' access to allofmp3.com, thereby making it harder for potential customers to access the site.<sup>273</sup> MasterCard also cut off

264. *Id.* at 79.

265. *Id.*

266. *Id.* (discussing IFPI's role); Nate Anderson, *Music Industry Encouraged Visa to Pull the Plug on AllofMP3.com*, ARSTECHNICA, Oct. 19, 2006, <http://arstechnica.com/business/news/2006/10/8029.ars>.

267. *International Piracy Hearing*, *supra* note 246, at 79 (statement of Mark MacCarthy, Senior Vice President for Global Public Policy, Visa Inc.).

268. *Id.*

269. Press Release, IFPI, Allofmp3.com: Setting the Record Straight (June 2, 2006), *available at* [http://www.ifpi.org/content/section\\_news/20060601.html](http://www.ifpi.org/content/section_news/20060601.html).

270. *See International Piracy Hearing*, *supra* note 246, at 26 (testimony of Victoria A. Espinel, Assistant U.S. Rep. for Intellectual Property and Innovation, Office of the U.S. Trade Rep.) ("We will continue to press Russia to shut down and prosecute the operators of illegal Web sites operating in Russia, including the successors to the infamous AllOfMP3.com.").

271. *Id.* at 79 (statement of Mark MacCarthy, Senior Vice President for Global Public Policy, Visa Inc.).

272. *Id.*

273. Press Release, IFPI, New Court Setback for Allofmp3.com (Oct. 26, 2006), *available at* [http://www.ifpi.org/content/section\\_news/20061026.html](http://www.ifpi.org/content/section_news/20061026.html).

payment services to allofmp3.com.<sup>274</sup> By May of 2007, the site's popularity had plummeted.<sup>275</sup>

The company was all but out of business, but the legal process was just starting. The owner of allTunes sued the bank that had stopped processing its Visa transactions in a Russian court.<sup>276</sup> Visa was a party to that litigation on the side of the bank.<sup>277</sup> In June 2007, the owner won a judgment that the bank had violated its contract with the merchant, and the judgment required the bank to continue to provide processing services.<sup>278</sup> In response to the bank's claim that the merchant was acting illegally, the court determined that there were no rulings in Russia establishing that allTunes was making illegal use of exclusive rights belonging to rights holders.<sup>279</sup>

In August 2007, another Russian court issued a ruling in a different case, relating to criminal copyright infringement initiated by IFPI against the owner of Allofmp3.com.<sup>280</sup> This ruling stated that there had not been sufficient confirmation of any illegal activity by the site's owner.<sup>281</sup> Even though the copyright owners had not given permission to distribute their recorded material, a Russian collective rights society (the Russian Multimedia and Internet Society, or ROMS by its initials in Russian) was deemed to be operating legitimately under Russian law.<sup>282</sup> The court implied that Allofmp3.com and similar sites would be in compliance with Russian law to the extent that they paid for rights from this Russian collective rights society.<sup>283</sup>

---

274. *MP3 Site's Voucher System Closes*, BBC NEWS, May 21, 2007 available at <http://news.bbc.co.uk/2/hi/entertainment/6677265.stm>.

275. IFPI reported in May 2007 that Allofmp3 "rated outside the top 2000 websites." Press Release, IFPI, Police Dawn Raid Stops Allofmp3.com Pirate Vouchers Scheme (May 21, 2007), available at [http://www.ifpi.org/content/section\\_news/20070521.html](http://www.ifpi.org/content/section_news/20070521.html).

276. Arbitration Court of Moscow 2007, A40-70411/06-67-500.

277. *Id.* at 1.

278. *Id.* at 5.

279. *Id.* The court stated,

According to Article 49 of the Russian Federation Law "On Copyright and Allied Rights," it is only the Court that can execute actions in connection with illegal use of copyrights and allied rights, if there is a lawsuit filed by exclusive right holders, which the Defendants, VISA and IFPI are not, while in this case there are no court rulings with the force of *res judicata* establishing the Plaintiff's illegal use of exclusive rights belonging to some right holders.

*Id.* The Defendant was Rosbank, the Russian financial institution licensed by Visa to authorize merchants in Russia to accept Visa. *Id.*

280. Cheremushkinsky [District Court of Moscow], 2007, No. 1-151-07.

281. *Id.* at 4.

282. *Id.* at 5.

283. *Id.*; see also *International Piracy Hearing*, *supra* note 246, at 99 (testimony of Victoria A. Espinel, Assistant U.S. Rep. for Intellectual Property and Innovation, Office of the U.S.

These court cases created a challenge for Visa because the payment system had responded to a documented complaint of copyright infringement.<sup>284</sup> Despite an outside review that seemed to establish illegality in the local jurisdiction, a local court ordered a local bank to continue to provide payment services.<sup>285</sup> Yet these transactions would still be illegal in virtually every other country in the world. To preserve its cross-border policy, Visa decided to allow the local bank to provide only domestic service to the site involved in the court case.<sup>286</sup> Transactions from customers in other countries would not be allowed.<sup>287</sup>

#### 4. *Assessment of Payment System Actions on Online Copyright Infringement*

The actions of payment systems to limit use of their systems for copyright infringement can be evaluated using the framework set out in Part II by examining the equities, market, and costs and benefits that would be involved if public policy imposed intermediary liability.

First, *Perfect 10* properly rejected indirect liability for payment intermediaries.<sup>288</sup> The involvement of payment networks in copyright violations is attenuated and entirely passive. On control grounds, there is simply no way to draw a line between payment network involvement in allegedly infringing transactions and involvement in a wide range of other potentially illegal activities. If they are liable in this case, why wouldn't they be liable for all cases of illegal activity on their payment systems? Unintentionally, Judge Kozinski's dissent brought out this implication.<sup>289</sup>

---

Trade Rep.) ("My understanding of the case is that Media Services, the company that operated allTunes, was able to successfully argue in Russian court that it was not acting illegally because it was paying royalties to collecting societies, collecting societies that were not authorized by the rights holders.").

284. *Id.* at 80 (statement of Mark MacCarthy, Senior Vice President for Global Public Policy, Visa Inc.).

285. *Id.* at 80–81.

286. *Id.*

287. *Id.* at 81.

288. *Perfect 10, Inc. v. Visa Int'l Serv. Ass'n*, 494 F.3d 788, 798 (9th Cir. 2007). For analysis, see Band, *supra* note 231.

289. *See id.* at 824 (Kozinski, J., dissenting) ("Credit cards already have the tools to police the activities of their merchants, which is why we don't see credit card sales of illegal drugs or child pornography."). Of course, card companies use different tools in the case of illegal drugs and child pornography, namely, proactive monitoring, but it is hard to see on Kozinski's analysis why card companies shouldn't use whatever tools they can to stop illegal activity in all cases. *See id.* Kozinski argues that

[p]laintiff is not asking for a huge change in the way credit cards do business; they ask only that defendants abide by their own rules and stop doing business with crooks. Granting plaintiff the relief it seeks would not . . . be the end of Capitalism as we know it.

*Id.* But it might be the end of payment systems as we know them if indirect liability for them

Judge Kozinski also painted a clear picture of how, in his opinion, payment intermediaries might act if they were liable for copyright infringement occurring through their services:

[T]he cards have the authority, given to them by contract, to force the Stolen Content Websites to remove infringing images from their inventory as a condition for using defendants' payment systems. If the merchants comply, their websites stop peddling stolen content and so infringement is stopped or limited. If they don't comply, defendants have the right—and under copyright law the duty—to kick the pirates off their payment networks, forcing them to find other means of getting paid or go out of business. In that case, too, infringement is stopped or limited.<sup>290</sup>

Judge Kozinski contemplated that the U.S. based payment intermediaries could more easily take action against parties in other jurisdictions compared to other actors: “Here, plaintiff alleges that many direct infringers have no physical presence in the United States. They operate from far-off jurisdictions, where lawsuits are difficult to bring and remedies impossible to enforce because the infringers can easily move their operations to servers in other remote jurisdictions.”<sup>291</sup>

But the actual experience of payment intermediaries reveals that things would not be that simple.<sup>292</sup> Even in the case of a well documented complaint, payment intermediaries are never truly asked to remove “infringing” material.<sup>293</sup> At best, there is a well-documented assertion of infringement under the laws of a particular jurisdiction.<sup>294</sup> Judge Kozinski appears to favor a notice-and-takedown approach, so that payment intermediaries are not responsible for illegal conduct of which they are unaware.<sup>295</sup> But as Visa found in *Allofmp3.com*, payment card services and their associated financial service partners can be liable for wrongful

---

means an obligation to stop doing business with everyone who might be involved with illegality anywhere. Kozinski attempts to limit his analysis to those cases where there are special arrangements between bad actors and the payment system, but nothing in his analysis turns on these special arrangements. *Id.* at 819–20. These special arrangements turn out to be risk-based pricing for adult content websites. Would he really have voted with the majority if the price that adult content merchants face for accepting cards was the same as the price set for less risky merchants?

290. *Id.* at 817.

291. *Id.* at 823.

292. *See supra* Section III.C.3.

293. *See supra* Section III.C.3.

294. *See supra* Section III.C.3.

295. *Perfect 10*, 494 F.3d at 824 (Kozinski, J., dissenting).

termination of services in those jurisdictions if they react to an allegation of infringement by “kick[ing] the pirates off their payment networks.”<sup>296</sup>

Second, there is no market failure in this situation that would justify imposing intermediary liability on payment systems. There are available arrangements between payment intermediaries and copyright owners that can reduce the amount of copyright infringement on the Internet.<sup>297</sup> These arrangements are informal, but expanding. They rely on complaints by copyright owners, followed by investigation and action by intermediaries.<sup>298</sup> They seem to strike a cost-based balance by putting the burden of discovering infringement on the copyright owner and triggering action by the third party only after notification. The arrangements may involve compensating payment intermediaries for performing enforcement services, but if this enables copyright owners to reduce the harm of copyright infringement, they might very well pay. If there are extra efforts, above and beyond standard practices, that a particular copyright owner would like payment intermediaries to make, those efforts should be open to negotiation. There do not seem to be any transaction costs that would prevent the parties from negotiating adjustments to these arrangements over time. And there appears to be no market failure that would justify not relying on private sector enforcement arrangements.

Third, given the legal risks involved, copyright owners should be willing to indemnify payment intermediaries for damages resulting from enforcement actions against alleged infringers. *Allofmp3.com* indicates that these legal risks are not hypothetical.<sup>299</sup> If the copyright owner is persuaded of the legal soundness of his case, he should be prepared to assume the risk.

296. *Id.* at 817. Mann and Belzley’s argument on *Perfect 10* also seems mistaken: In terms of equity, Visa has clean hands and Cybernet does not. That might make sense in a legal system designed to force bad actors to provide redress to injured parties. The better question, albeit one not readily susceptible of judicial analysis, is whether either Visa or Cybernet is the party best situated to stop the copyright violations in question. On that point, Visa probably is better situated, because of the real world likelihood that none of the sites that fosters the infringement could survive as a profitable commercial enterprise without accepting Visa payments.

Mann & Belzley, *supra* note 8, at 264. Several points need to be made. The equity considerations cannot be ignored. If Visa has “clean hands” it is hard to see why they should be held responsible. Also, the fact that Visa is better positioned than some other party to take enforcement action does not imply that these enforcement costs are worth the benefits. And the existence of complaint procedures suggests that indirect liability is not as a practical matter required.

297. *See supra* Section III.C.3.

298. *See supra* Section III.C.3.

299. *See supra* Section III.C.3.

It might be one way to assure that only strong complaints are brought to the attention of the payment intermediary. An additional mechanism might be to require the presence of a court or governmental agency that holds that the activity involved is infringing.

A statute could potentially help provide legal immunity to payment intermediaries when they take good faith action against alleged infringers. But U.S. law cannot provide immunity in other jurisdictions, which is where the aid of global payment intermediaries is needed.<sup>300</sup>

Fourth, this case illustrates the need for greater clarity in the legal environment in which intermediaries operate. Intermediaries cannot be in the position of creating new global law through their own interpretation of current statutes. Again, *Allofmp3.com* suggests the need for even greater harmonization of local laws that intermediaries are expected to enforce.<sup>301</sup> The United States Trade Representative attempted this by working with the Russian government to bring about changes in Russian law that would bring it closer to the international norm.<sup>302</sup>

In sum, the experience of payment intermediaries indicates that some efforts on their part to respond to legitimate complaints would be justified. It is not appropriate to do nothing in response to allegations of copyright infringement. The current complaint procedure and case-by-case response is reasonable. It could be improved through further discussions among the parties, further recourse to court judgments of infringement, and harmonization of current international standards.

#### IV. INTERNET GOVERNANCE

Parts II and III laid out a framework for analyzing intermediary liability and applied it to the actions of payment intermediaries. That analysis is also relevant to fundamental questions of internet governance. This Part explores the extent to which the experience of payment systems in controlling the illegal online behavior of their users illuminates the debate between the internet exceptionalists, defenders of the bordered Internet, and the

---

300. *See supra* Section III.C.3.

301. *See supra* Section III.C.3.

302. OFFICE OF THE U.S. TRADE REPRESENTATIVE, EXECUTIVE OFFICE OF THE PRESIDENT, RESULTS OF BILATERAL NEGOTIATIONS ON RUSSIA'S ACCESSION TO THE WORLD TRADE ORGANIZATION (WTO): ACTION ON CRITICAL IPR ISSUES (2006) ("Russia will work to enact legislation by June 1, 2007, to stop collecting societies from acting without right holder consent, Russia will also work to enact legislation implementing the 1996 World Intellectual Property Organization (WIPO) Internet treaties."). These new measures might enable Russian courts to reverse their earlier decisions. *See International Piracy Hearing, supra* note 246, at 30 (statement of Victoria A. Espinel, Assistant U.S. Rep. for Intellectual Property and Innovation, Office of the U.S. Trade Rep.).

internationalists. It concludes that exceptionalism, in either its original or modified forms, is not the right framework for internet governance because intermediaries should not defer to the judgments of self-governing communities of internet users when the judgments conflict with local law. The exceptionalists are correct that a “bordered Internet” will not scale up, but the experience of traditional payment systems points towards international harmonization. If governments are going to use intermediaries to regulate the Internet, they need to coordinate their own laws to make that role possible.

This Part addresses each of the three main approaches to internet governance: exceptionalism, the bordered Internet, and internationalism. Section IV.A, on exceptionalism, begins with a discussion of the original internet exceptionalist perspective, which viewed government regulation of the Internet as infeasible and normatively less desirable than government deference to the rules developed by self-governing internet communities. This is followed by a discussion of Brian Holland’s revised version of exceptionalism. Under this approach, the various immunities from intermediary liability established by local jurisdictions enable the development of autonomous Internet norms. Both versions are shown to have significant limitations when viewed in light of payment system experiences. Section IV.B explores the “bordered Internet,” the idea that in certain cases local governments may properly and unilaterally extend their jurisdiction over internet activities through intermediaries. Payment intermediaries use standard measures to resolve conflicts of law and follow a practical rule that treats a transaction as illegal if it is illegal in the jurisdiction of either the merchant or the cardholder. Section IV.B then discusses limitations on this method of resolving cross-border jurisdictional conflicts. Section IV.C concludes with a discussion and endorsement of the internationalist perspective, according to which local governments should only exercise control over specific internet activities in a coordinated fashion.

#### A. INTERNET EXCEPTIONALISM

##### 1. *The Original Version*

In February 1996, John Perry Barlow identified internet exceptionalism when he declared cyberspace to be independent of national governments, roughly on the grounds that cyberspace “does not lie within your borders” and that it “is a world that is both everywhere and nowhere, but it is not where bodies live.”<sup>303</sup> Conflicts in cyberspace would be resolved not with the

---

303. Declaration of John P. Barlow, Cognitive Dissident, Co-Founder, Elec. Frontier Found., A Declaration of the Independence of Cyberspace (Feb. 8, 1996), *available at* [http://w2.eff.org/Censorship/Internet\\_censorship\\_bills/barlow\\_0296.declaration](http://w2.eff.org/Censorship/Internet_censorship_bills/barlow_0296.declaration).

territorially-based “legal concepts of property, expression, identity, movement, and context,” which “do not apply,” to cyberspace because they “are all based on matter, and there is no matter here.”<sup>304</sup> Rather, in cyberspace “governance will arise according to the conditions of our world, not yours.”<sup>305</sup> Cyberspace “is different.”<sup>306</sup>

Almost concurrently, legal scholars David Johnson and David Post made a similar case for internet exceptionalism.<sup>307</sup> In their view, the Internet destroys “the link between geographical location” and “the power of local governments to assert control over online behavior; [and] . . . the legitimacy of a local sovereign’s efforts to regulate global phenomena . . . .”<sup>308</sup> The Internet destroys the power of local governments because they cannot control the flow of electrons across their physical boundaries, and if they attempted to do so, determined users would just route around the barriers. Moreover, if one jurisdiction could assert control over internet transactions they all could, resulting in the impossibility that all “Web-based activity, in this view, must be subject simultaneously to the laws of all territorial sovereigns.”<sup>309</sup> The Internet destroys the legitimacy of local jurisdiction because legitimacy depends on the consent of the governed and “[t]here is no geographically localized set of constituents with a stronger and more legitimate claim to regulate it than any other local group. The strongest claim to control comes from the participants themselves, and they could be anywhere.”<sup>310</sup> Since “events on the Net occur everywhere but nowhere in particular . . . no physical jurisdiction has a more compelling claim than any other to subject these events exclusively to its laws.”<sup>311</sup>

Behind these arguments seemed to be an appealing political vision. The ideal envisaged self-organizing groups of people making the rules that applied to their conduct. These rules would not be imposed from the outside, but would be freely chosen by the active participation of the community members. The key was deliberation by free, rational agents in their communities, not imposition of rules by an arbitrary act of will by a distant sovereign. This ideal of participatory democracy was intended, in part, to offset the alienating effects of large-scale modern democracies, which in practice had long failed to provide their members with the sense of

---

304. *Id.*

305. *Id.*

306. *Id.*

307. *See generally* Johnson & Post, *supra* note 1.

308. *Id.* at 1370 (emphasis added).

309. *Id.* at 1374.

310. *Id.* at 1375.

311. *Id.* at 1376.

community participation that alone seemed to justify the imposition of collective rules.

The way this vision would be implemented on the Internet would be through the development of autonomous communities of internet users. These internet communities were largely isolated from “real world” communities. Since it took special care and effort to reach out to participate in them, only those people who really wanted to participate would, and the effects of activities in those communities would be limited to those who chose to participate. Given the structure of the Internet as a communications network, which moved almost all major decisions on content to the edges of the network, a diversity of law could arise in cyberspace as each community developed its own norms for regulating the conduct of its members. People would be free to participate in the communities they wanted, but could easily avoid those they did not like. Enforcement of the community rules would be accomplished through peer pressure, reputational systems, informal dispute resolution mechanisms, and ultimately, banishment. The system as a whole would evolve through a process analogous to biological evolution, where diverse and potentially competing rule sets as embodied in different communities would vie for acceptance in a free marketplace of rules.

Internet exceptionalism is thus the view that activity on the Internet should be regulated by internet community norms, not laws of territorial jurisdictions or globally harmonized laws.<sup>312</sup> It is hard to avoid the sense that the political vision predated the Internet—that the feasibility argument masked the underlying vision and the arrival of the Internet simply created the possibility of implementing the vision in a way that the “real” world did not. To see this, imagine the reaction of internet exceptionalists to the idea of a world government that would establish uniform global laws. This would eliminate the conflict of law problem. But exceptionalists are even more appalled with the idea of world government control over the Internet than with the idea of nation-state control over it. This suggests that the issue is

---

312. Mann and Belzley describe their view as “consciously exceptionalist” because “specific characteristics of the Internet make intermediary liability relatively more attractive than it has been in traditional offline contexts because of the ease of identifying intermediaries, the relative ease of intermediary monitoring of end users, and the relative difficulty of directly regulating the conduct of end users.” Mann & Belzley, *supra* note 8, at 250–51. But this is an odd way of framing the issue. Internet exceptionalism is not simply the view that the Internet should be treated differently from the offline world. The claim is more specifically that the Internet should be free of local jurisdictions. Mann and Belzley’s view, which implies that the Internet should be brought under local jurisdictions through the mechanism of intermediary liability, is thus the very opposite of exceptionalism. It is one version of internet non-exceptionalism.

not feasibility of control, but the value of participative community decision making and diversity.

This early cyber libertarian vision was immediately attacked by those who defended the feasibility and legitimacy of extending local laws to cover internet activity.<sup>313</sup> As they note, “[t]he mistake here is the belief that governments regulate only through direct sanctioning of individuals. . . . Governments can . . . impose liability on intermediaries like Internet service providers or credit card companies.”<sup>314</sup> Government action against these intermediaries “makes it harder for local users to obtain content from, or transact with, the law-evading content providers abroad. In this way, governments affect Internet flows within their borders even though they originate abroad and cannot easily be stopped at the border.”<sup>315</sup> And these efforts to bring order to the Internet through pressure on intermediaries are often legitimate because they provide “something invisible but essential: public goods like criminal law, property rights, and contract enforcement . . . that can usually be provided only by governments.”<sup>316</sup>

This attack was so effective that many believe that these notions of a “self-governing cyberspace are largely discredited.”<sup>317</sup> But modified versions accept the basic premise that the Internet should be free of local regulation and governed by its users. One version of the revived exceptionalism, defended by Brian Holland, focuses on Web 2.0 communities.<sup>318</sup> This view argues that together with the immunity provisions of § 230 of Communications Decency Act, these communities have the potential to allow internal community norms to take the place of external territorially based laws.<sup>319</sup>

---

313. *See generally* Goldsmith, *supra* note 2 (challenging the regulation skeptics).

314. *Id.* at 1238.

315. GOLDSMITH & WU, *supra* note 7, at 68.

316. *Id.* at 140.

317. *Id.* at 14.

318. Holland writes,

By mitigating the imposition of certain external legal norms in the online environment, § 230 helps to create the initial conditions necessary for the development of a modified form of exceptionalism. With the impact of external norms diminished, Web 2.0 communities, such as wikis and social networks, have emerged to facilitate a limited market in norms and values and to provide internal enforcement mechanisms that allow new communal norms to emerge.

Holland, *supra* note 16, at 369.

319. *Id.*

## 2. *Critique of Internet Exceptionalism*

The experience of global payment intermediaries described in Part II confirms the view that intermediaries can effectively control illegal activity in cyber space. This still leaves the question of whether intermediaries should resist such attempts to control the behavior of their users. As a general matter, they should not defer to the judgments of self-governing communities of internet users when these judgments conflict with local law. As corporate citizens, they have an obligation to obey the laws of the jurisdictions in which they operate, and they simply have no basis to excuse themselves from that duty in order to let online communities determine their own fate. But even when local law does not require them to take action against illegal behavior, as in the child pornography, controlled substances, and online tobacco cases described earlier, their responsibility to keep their systems free of illegal activity means that they often should take specific steps to stop these activities.

The fundamental objection, even to Holland's modified exceptionalism, is that the "law" of internet communities is not really the law of that community. It is a commercial contract enforceable under the rules of some local jurisdiction, and the terms of the contract are subject to the same kinds of legal and regulatory oversight that bind contracts between people in local jurisdictions. Deferring to these contracts does not usually mean democratic community self-government. Local regulations are needed to fully protect the members of these communities.<sup>320</sup> Moreover, in some cases the legal discretion granted to intermediaries to control the conduct of their members may be too broad and should be limited by replacing intermediary judgment with public authority decisions. The remainder of this Section develops these points.

Even if internet communities could substantially exclude a significant portion of external legal norms, it still does not imply that internal norms will emerge from the process of debate and deliberation that Holland envisages. As Holland notes, "external legal norms are excluded, but internal communal norms are often unable to coalesce to take their place" because enforcement is "concentrated in private commercial entities."<sup>321</sup> The modified internet exceptionalism hope is that the intermediaries who control the new Web 2.0 platforms will be driven by internal incentives to accommodate the wishes of the online communities they create, allowing users to establish norms for their own communities.<sup>322</sup>

---

320. This Section focuses on competition policy, privacy, and consumer protection as examples.

321. Holland, *supra* note 16, at 398.

322. These internal incentives include "the need for financial support from community

But it is not clear that Web 2.0 platforms are likely to grant this kind of democratic self-governance. For example, intermediaries can be subject to pressure. Craig Newmark, the operator of Craigslist, has insisted that he made his decision to remove ads for erotic services as a result of consultation with his online community.<sup>323</sup> But it is also true that Craigslist was under criminal investigation by a number of state attorneys general for violation of state laws against prostitution.<sup>324</sup> One could argue immunity in this case, but Craigslist did not.<sup>325</sup> It complied with a law enforcement request to remove certain postings and the decision to remove these ads will be subject to ongoing oversight by these law enforcement agencies.<sup>326</sup> However, the question remains whether or not Craigslist would take the legal risk if the community voted to keep these ads in place.

These communities are not typically governed by democratic voting procedures that guarantee the consent of the governed. They are governed by contractual terms of service. Often prospective members of these communities have a simple take-it-or-leave-it choice when they decide to join.<sup>327</sup>

---

donations, a communal desire for information integrity, or the need to build an audience for advertising.” *Id.* at 400; *see also* Schruers, *supra* note 68, at 261 (“ISPs respond to content-based complaints as a matter of good business practice for the purpose of maintaining customer goodwill and satisfaction.”).

323. Craigslist Founder Seeks Larger DC Role, NAT’L J., June 2, 2009, *available at* <http://techdailydose.nationaljournal.com/2009/06/craigslist-founder-seeks-large.php> (reporting Craig Newmark’s comments to the Computers Freedom and Privacy Conference).

324. *See* Brad Stone, *Craigslist to Remove ‘Erotic’ Ads*, N.Y. TIMES, May 14, 2009, at B1. Craigslist’s attorneys asserted immunity under § 230, but chose voluntarily to remove the ads to which various state attorneys general had objected. *Id.* State Attorneys General felt confident that they could bring a case under state criminal law despite the immunity granted by § 230. *Id.* The case was given national attention when a medical student was accused of killing a masseuse whom he met through Craigslist. *Id.*

325. *Id.*

326. *Id.*

327. *See* Johnson & Post, *supra* note 1, at 1380 (describing AOL or Compuserve terms of service as examples of law in cyberspace). Johnson and Post view the rules for an internet community to be “a matter for principled discussion, not an act of will by whoever has control of the power switch.” *Id.* But it is hard to see how terms of service for a typical internet service or application is anything other than an act of will by the person who controls the service or application. It might satisfy certain legal standards for informed consent, but it is not the product of principled discussion. And this might be the way consumers want it. Online communities might not offer to determine their online laws through a political process because the members of the community cannot be bothered. People visit many different websites and use many different web services. It is hard to believe that they want full democratic participation rights to set up the rules for each of these services. And it is implausible that they would actually spend the time, if they were offered the opportunity. The example of privacy policies makes the point. A recent study concluded that if all U.S. consumers read all the privacy policies for all the websites they visited just

If consumers do not like the terms of service, then protest can be effective, as in the recent case of users objecting to the change in terms of service unilaterally offered by Facebook. By threatening the privacy rights of the community, the platform stirred up substantial community unrest, and ultimately the new terms of service were withdrawn.<sup>328</sup> But this exit right is not the same as democratic self-governance, and it is not always effective. What if Facebook had not responded to community objections? Would people actually have left, and where would they have gone? Lock-in is a real phenomenon in social networks.

The exemption from liability based on § 230 does not mean that online entities are exempt from local law. Often, local law is needed to protect consumers from the actions of internet intermediaries. Regulation of online communities by governments seems especially timely and urgent in three areas—competition policy, privacy, and consumer protection.

With respect to competition, concentration in particular sectors of the online world should be examined because it can so significantly reduce consumer choice. The Department of Justice has indicated, for example, that it is going to take a more active approach in this area.<sup>329</sup> Along with the FTC, they have initiated inquiries focused on the search engine market.<sup>330</sup>

Privacy and security rules need to be defined as well. The FTC has taken major action in this area, and is stepping up their enforcement.<sup>331</sup> They are

once a year, the total amount of time spent on just reading the policies would be 53.8 billion hours per year and the cost to the economy of the time spent doing this would be \$781 billion per year. Alecia M. McDonald & Lorrie F. Cranor, *The Cost of Reading Privacy Policies*, 4 I/S: J.L. & POL'Y FOR INFO. SOC'Y 543, 565 (2008).

328. N.Y. Times, Facebook, Inc., [http://topics.nytimes.com/top/news/business/companies/facebook\\_inc/index.html](http://topics.nytimes.com/top/news/business/companies/facebook_inc/index.html) (last updated May 27, 2009). In 2007, the company had created a community backlash when it introduced an advertising service that allowed a user's online activities to be distributed to other community members. Epic.org: Electronic Privacy Information Center, Social Networking Privacy, <http://epic.org/privacy/socialnet/default.html> (last visited Feb. 3, 2009). In the face of this protest, it provided a simple way for users to decline to participate. *Id.* In February 2009, it proposed new privacy rules according to which users will own and control their own information, and in April it allowed a vote of its users on these new principles. Over 75% of those voting endorse them, and on July 1, 2009 it adopted them. *Id.*

329. Press Release, U.S. Dep't of Justice, Justice Department Withdraws Report on Antitrust Monopoly Law: Antitrust Division to Apply More Rigorous Standard with Focus on the Impact of Exclusionary Conduct on Consumers (May 11, 2009), *available at* [http://www.justice.gov/atr/public/press\\_releases/2009/245710.pdf](http://www.justice.gov/atr/public/press_releases/2009/245710.pdf).

330. *See, e.g.*, Miguel Helft, *U.S. Inquiry Is Confirmed into Google Books Deal*, N.Y. TIMES, July 3, 2009, at B3; Miguel Helft & Brad Stone, *Board Ties at Apple and Google Scrutinized*, N.Y. TIMES, May 5, 2009, at B1; Peter Whoriskey, *Google Ad Deal Is Under Scrutiny: Yahoo Agreement Subject of Antitrust Probe, Sources Say*, WASH. POST, July 2, 2008, at D1.

331. *See* Press Release, Fed. Trade Comm'n, Sears Settles FTC Charges Regarding Tracking Software (June 4, 2009), *available at* <http://www.ftc.gov/opa/2009/06/sears.shtml>

also focusing on the development of a new privacy framework to analyze the basis for the harms associated with privacy violations.<sup>332</sup> Furthermore, the FTC has focused on developing rules for online behavioral advertising.<sup>333</sup> In addition, rules governing privacy for online cloud computing services need to be clarified, perhaps by additional legislation.<sup>334</sup>

Consumer protection rules should be updated to apply more effectively to new developments in electronic commerce including the growth of mobile commerce and user-generated content, the greater availability of digital goods online, and increased numbers of consumers acting as online sellers, and new developments in accountability and payment protection. A timely development might be the harmonization of consumer redress and liability rights across various payment mechanisms.<sup>335</sup>

Finally, the discretion given to internet intermediaries over which transactions to allow must be subject to public scrutiny. Today, intermediaries exercise judgment over which transactions are subject to such legal risk that they cannot be allowed. These decisions are made in the context of the business interests and technological capabilities of the intermediaries themselves, but they have important effects on the rights and interests of other parties. Some examples, explained above, include:

- Payment systems effectively decide which internet gambling transactions are illegal. By choosing to block all coded gambling transactions, the system disadvantages horseracing, state lottery, and Indian gaming transactions that are arguably legal.

---

(reporting that in the Sears case the FTC obtained a settlement from Sears after charging that their consent practices in regard to installing an online tracking program on customers' computers constituted an unfair or deceptive practice).

332. See Stephanie Clifford, *Fresh Views at Agency Overseeing Online Ads*, N.Y. TIMES, Aug. 5, 2009, at B1 (stating that David Vladek, the new head of the FTC's consumer protection division, is rethinking privacy). Vladek said that "[t]he frameworks that we've been using historically for privacy are no longer sufficient." *Id.* In his view the FTC will begin to consider not just whether companies caused monetary harm, but whether they violated consumers' dignity because, for example, "[t]here's a huge dignity interest wrapped up in having somebody looking at your financial records when they have no business doing that." *Id.*

333. See Press Release, Fed. Trade Comm'n, FTC Staff Revises Online Behavioral Advertising Principles (Feb. 13, 2009), available at <http://www.ftc.gov/opa/2009/02/behavad.shtm>.

334. See generally ROBERT GELLMAN, WORLD PRIVACY FORUM, PRIVACY IN THE CLOUDS: RISKS TO PRIVACY AND CONFIDENTIALITY FROM CLOUD COMPUTING (2009) (discussing these cloud computing issues).

335. Legal payment protections now differ depending on the type of payment product used (debit or credit) and the nature of the payment provider—traditional payment providers like Visa face legal requirements while new payment providers such as cell phone companies do not.

- Payment systems take complaints from third parties, make an independent legal assessment of the merits of the case, and withdraw service based on these assessments. Effectively, they adjudicate these copyright cases.
- Payment systems looked at the legal arguments from state AGs and BATF and determined that they were stronger than the arguments of the online tobacco merchants.
- Payment systems choose to adopt lists of dangerous substances from the FDA and DEA lists and determine that they cannot be sold by internet pharmacies.

These decisions are sound and sensible ways to balance complex and competing interests. However, they are private sector judgments, inevitably subjective and influenced by the particular interests of the parties involved.

Other intermediaries also have enforcement abilities that they can use at their own discretion. For instance, in January 2009, it was reported that an Irish ISP had agreed to disconnect subscribers who were accused of three instances of infringement by a copyright owner.<sup>336</sup> Allegations of violations would be made by a contractor working for the content owner and transmitted to the ISP.<sup>337</sup> At this point, these decisions are largely up to the payment intermediaries and the ISPs themselves, although in some jurisdictions they are dictated by government requirements,<sup>338</sup> yet their

---

336. Posting of Danny O'Brien to Deeplinks Blog, Irish ISP Agrees to Three Strikes Against Its Customers, <http://www.eff.org/deeplinks/2009/01/irish-isp-agrees-three-strikes-against-its-users> (Jan. 28, 2009).

337. *Id.* Under the agreement the music labels, instead of going to court to get an order to have the ISP shut off a subscriber's connection, provide evidence of infringement to the ISP directly. *Id.* As O'Brien noted,

The difference is that an ISP is not a court; and its customers will never have a chance to defend themselves against the recording industry's accusations and "proof." To whom, without judicial oversight, has the ISP obligated itself to provide meaningful due process and to ensure that the standard of proof has been met?

*Id.*

338. The movement toward graduated response would replace this discretion with government processes. Under the recently passed HADOPI law, French ISPs would be required to suspend internet access for subscribers who have been subject to three allegations of copyright violations. Catherine Saez, *French HADOPI Law, Now Complete, Can Brandish Its Weapons*, INTELL. PROP. WATCH, Oct. 23, 2009, <http://www.ip-watch.org/weblog/2009/10/23/french-hadopi-law-now-complete-can-brandish-its-weapons/>. A court review would be required before suspension. *Id.* A similar graduated response program was adopted in Britain in April 2010. Eric Pfanner, *U.K. Approves Crackdown on Internet Pirates*, N.Y. TIMES, Apr. 8, 2010, <http://www.nytimes.com/2010/04/09/technology/>

decisions will have profound effects on the shape and direction of electronic commerce. Deferring to the norms of the internet community in this context means deferring to these private judgments of intermediaries.

There is a role for internet community decision-making. The best circumstances for deference to law constructed for and by particular internet communities is when an internet community's norms do not "fundamentally impinge upon the vital interests of others who never visit this new space."<sup>339</sup> To the extent that an internet community is self-contained or its activities affect others only on a voluntary basis, then there is a case for deferring.<sup>340</sup>

#### B. PAYMENT SYSTEMS AND THE BORDERED INTERNET

Goldsmith and Wu attack internet exceptionalism, but they also construct a positive vision of a "bordered Internet."<sup>341</sup> This world would work pretty much as the world worked before the Internet. New regulations would be crafted to deal with the new dangers specifically created by the Internet, but there would be no fundamental needed to adjust the basic domestic or international framework.<sup>342</sup>

Resolving jurisdictional disputes would be one significant problem with the bordered Internet. The initial internet exceptionalist argument was that internet activity is simultaneously present in multiple overlapping and inconsistent jurisdictions, and that no one jurisdiction has a better claim to regulate the activity than any other jurisdiction. It would be better to think of the activity as taking place in a separate jurisdiction altogether and have the territorial governments of the world defer to the community norms created there. Goldsmith and Wu's response was that internet activity was real world activity, taking place in particular jurisdictions, and that local governments

---

09piracy.html. Whether these graduated response programs are needed is a point of controversy, but they replace ISP discretion with a system of public accountability.

339. Johnson & Post, *supra* note 1, at 1389.

340. See POST, IN SEARCH OF JEFFERSON'S MOOSE, *supra* note 15, at 178–86 (describing "massively multi-player online games" or MMOGS as good candidates for this effort at online rule creation). This might be. However, Linden Labs, the creator of one of the famous MMOGS, found it necessary to rely on external banking regulators when it decided to ban the offering of interest or any return on investment in-world without proof of an applicable government registration statement or financial institution charter. Posting of Kend Linden to Second Life Blogs, New Policy Regarding In-World "Banks," <https://blogs.secondlife.com/community/features/blog/2008/01/08/new-policy-regarding-in-world-banks> (Jan. 8, 2008 06:43:56 PM). Linden Labs properly concluded that it "isn't, and can't start acting as, a banking regulator." *Id.* New rule-making institutions will emerge only if people think that they are real. For this reason, a policy to defer in certain cases should be public and stable in order to provide the opportunity for the development of alternative rules.

341. GOLDSMITH & WU, *supra* note 7, at viii.

342. *Id.* at 149.

could exert control over this activity by attaching obligations to the local operations of global internet intermediaries.<sup>343</sup> This indirect liability for intermediaries would make it easier to extend local law to the bad actor.<sup>344</sup> Conflict of laws would be handled by the normal mechanisms for resolving these disputes, and ultimately enforced by actions taken against local operations of global intermediaries.<sup>345</sup>

Jurisdiction in cyberspace is a complex topic with many different approaches to assigning both the applicable law and the court of jurisdiction.<sup>346</sup> Questions include determining the location of the transaction, the jurisdiction, and the interests of the parties.<sup>347</sup> An early attempt to deal with these issues in the internet context was the Federal Trade Commission's (FTC) approach to consumer protection in the global marketplace.<sup>348</sup> The simplest cross-border electronic transaction implicates transnational concerns. Choice of law debates inevitably follow. The FTC considered arguments for the "country of origin" approach and the "country of destination" approach.<sup>349</sup> Under the country of origin approach, the law of the merchant would apply and the courts of the merchant's country would adjudicate any disputes.<sup>350</sup> Under the country of destination approach, the law of the consumer would apply and the courts of the consumer's country would adjudicate disputes.<sup>351</sup>

343. *Id.* at 68–72.

344. Mann & Belzley, *supra* note 8, at 259 (“[On the Internet it is] easier for even solvent malfeasors engaged in high-volume conduct to avoid responsibility either through anonymity or through relocation to a jurisdiction outside the influence of concerned policymakers.”). Mann and Belzley also argue that indirect liability makes sense in “cases in which the retailer is located in a jurisdiction outside the United States that will not cooperate with the relevant state regulators.” *Id.* at 277.

345. GOLDSMITH & WU, *supra* note 7, at 158–61.

346. See, e.g., Paul S. Berman, *Towards a Cosmopolitan Vision of Conflict of Laws: Redefining Governmental Interests in a Global Era*, 153 U. PA. L. REV. 1819, 1822 (2005) (arguing that judges should adopt a cosmopolitan approach in internet cases involving choice of law and foreign judgment issues, grounded in the “idea that governments have an interest not only in helping in-state litigants win the particular litigation at issue, but a more important long-term interest in being cooperative members of an international system and sharing in its reciprocal benefits and burdens”).

347. See generally Goldsmith, *supra* note 2 (discussing many of these theories); see also Berman, *supra* note 346, at 1839–40 (discussing various choice-of-law theories that address these questions).

348. FED. TRADE COMM’N, CONSUMER PROTECTION IN THE GLOBAL ELECTRONIC MARKETPLACE: LOOKING AHEAD (2000). The FTC’s discussion of applicable law and jurisdiction is especially relevant. *Id.* at 4–11.

349. *Id.*

350. *Id.* at 2.

351. *Id.* The European Union appeared to take the side of the country of origin in its E-Commerce Directive. European Commission, E-Commerce Directive, <http://ec.europa>.

The defense of the country of origin approach relied on the difficulty of applying any other legal framework to the electronic marketplace.<sup>352</sup> Only this country of origin framework seems to allow for the growth of global e-commerce. The framework considers problems encountered by small businesses selling in many countries of creating and applying a standard for some variety of “purposeful” targeting. Creating a default rule of the country of origin was deemed to better provide needed uniformity and predictability for online businesses.

This approach has defects. First, it forces consumers to rely on unfamiliar consumer protections. If merchants cannot be expected to know the laws of 180 countries, neither can consumers. Second, it creates a race to the bottom, whereby unscrupulous merchants can simply locate in a country with weak consumer protections. Third, consumers cannot reasonably be expected to travel to the country of origin to obtain redress. Fourth, consumers could not rely on their own consumer protection agencies for redress either, since these agencies would also be unable to enforce the consumer’s home jurisdiction protections.

So neither default rule seemed to suffice. As a practical matter, consumer education, self-regulatory efforts, and the development of codes of conduct by multinational organizations were the means chosen to address the cross-border consumer protection issue.<sup>353</sup> For other issues that could not be addressed through these means, the traditional tools of international conflict of law resolution would have to suffice.<sup>354</sup>

---

eu/internal\_market/e-commerce/directive\_en.htm (last visited Feb. 15, 2010). The Directive contains an Internal Market clause “which means that information society services are, in principle, subject to the law of the Member State in which the service provider is established.” *Id.*

352. FED. TRADE COMM’N, *supra* note 348, at 4 (discussing the “two fundamental challenges” to a country-of-destination framework, including “the use of physical borders to determine rights in a borderless medium” and compliance costs).

353. In 1999, the OECD issued its Guidelines for Consumer Protection in the Context of Electronic Commerce, which address principles that could be used by electronic commerce merchants in the absence of global consumer protection rules. ORG. FOR ECON. CO-OPERATION & DEV., GUIDELINES FOR CONSUMER PROTECTION IN THE CONTEXT OF ELECTRONIC COMMERCE (1999) [hereinafter OECD GUIDELINES], available at [http://www.oecd.org/document/51/0,3343,en\\_2649\\_34267\\_1824435\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/51/0,3343,en_2649_34267_1824435_1_1_1_1,00.html).

354. In an interesting twist, some commentators used the presence of these dispute resolution mechanisms to argue against indirect liability for intermediaries. Why deputize intermediaries to stop illegal activities on the Internet when governments can reach the bad actors and resolve any disputes in the normal way? Responding to the argument that indirect liability is needed because the bad actor is unreachable by law enforcement or aggrieved parties, Holland says,

As an initial matter, it is not clear that a significant number of bad actors are beyond the reach of the law. Advances in technology are making it

Some commentators, such as Paul Berman, attempted to reach beyond the traditional dispute resolution mechanisms for resolving conflict of law cases with principles that take into account the realities of multiple community affiliations.<sup>355</sup> His “cosmopolitan pluralism” was “cosmopolitan” because it went beyond the laws of any one particular jurisdiction and recognized the legitimacy of norms created by private parties and communities.<sup>356</sup> It was plural because it did not dissolve the multiplicity of community affiliations and their associated norms into a single world-wide standard. Diversity and conflict would endure and would need to be resolved according to a series of principles that recognized the need to balance competing national norms.<sup>357</sup>

These approaches to resolving jurisdictional disputes in cyberspace have various advantages and disadvantages. However, payment system intermediaries needed a mechanism to address the jurisdictional question that was easy to apply, effective in resolving the dispute, and minimized legal risk to the system or its members. It could not wait for unpredictable, after the fact judgments by courts. The idea they developed, discussed in Section III.A, was that a transaction is unacceptable in the payment system if it is illegal in the jurisdiction of either the buyer or the seller.<sup>358</sup>

The payment card approach provides a simple default rule for intermediaries to apply when determining whether to allow transactions in their systems. It eliminates the heavily fact-based balancing assessments needed to determine, on a case by case basis, whose law applies. The default rule also does not simply adopt a country of origin or country of destination

---

increasingly possible to locate and identify bad actors online, such that online anonymity is difficult to maintain. Likewise, where the bad actor is identified but is found outside the jurisdiction, sovereign governments have developed methods for resolving disputes to permit the direct extraterritorial application of domestic law, such as rules of jurisdiction, conflicts of laws, and recognition of judgments.

Holland, *supra* note 16, at 393.

355. Berman, *supra* note 346, at 1862.

356. *Id.*

357. *Id.* Berman’s work has affinities with that of political philosophers working in the area of national sovereignty in a global world. *See, e.g.,* POGGE, *supra* note 82, at 168–95.

358. Visa’s policy is stated in *International Piracy Hearing*, *supra* note 246, at 71 (statement of Mark MacCarthy, Senior Vice President for Global Public Policy, Visa Inc.). Other payment intermediaries have similar procedures, such as eBay’s restriction about selling and shipping illegal goods to the country where they are illegal. eBay, *Offensive Material Policy*, <http://pages.ebay.com/help/policies/offensive.html> (last visited Feb. 4, 2010) (“[B]ecause eBay is a worldwide community, many of our users live in countries where the possession or sale of items associated with hate organizations is a criminal offense. We can’t allow the sale or shipping of these items there.”).

perspective, each of which is limited. Nor does it leave the transaction in a legal limbo where no law applies.<sup>359</sup>

The payment system experience leads to several observations. First, direct conflicts of law are not as frequent as some anticipated. Technology and payment system practices effectively reduce these conflicts to the rare instance where the law of one country demands what the law of another country forbids. Directly contradicting laws are more common in “political” areas, where governments are seeking information from intermediaries to enforce local laws against its own citizens.<sup>360</sup>

Second, regulating the Internet by focusing on the local affiliates of global payment operations does not require the use of either the traditional or the new “cosmopolitan” conflict resolution methods. By relying on global

---

359. The internal application of this rule involves system efficiency and the balance of interests among the stakeholders in the system. If the merchant is in violation of its own country’s law, then enforcement is conceptually easy. Merchants discovered in violation of local law either have to stop the transactions or be removed from the system. If the merchant is in violation of the law in a different jurisdiction, things are more complicated. Should the bank of the merchant or the bank of the customer be burdened with the enforcement responsibility? If the merchant has this responsibility, then he must not introduce the illegal transaction into the system and the merchant’s bank must not try to process it, then steps must be taken at the merchant’s end to stop the transaction. These steps could include: a system decision requiring the merchant to stop these transaction entirely or leave the system, or coding and programming modifications by the merchant, the merchant’s processor, or the system operator, that would block transactions at the merchant end from entering the system if the customer was from a jurisdiction where the transaction would be illegal or which would restrict the transaction to the merchant’s own jurisdiction. Alternatively, the enforcement measures could be put on the cardholder side. Merchants could introduce properly coded transactions into the system and rely on action on the cardholder’s side to stop the transaction. This seems to fit the case of internet gambling, where U.S. law makes Internet gambling illegal for U.S. citizens, and the payment networks responded to UIGEA with a coding and blocking system that allowed merchants to continue their services in countries where internet gambling was illegal, as discussed earlier in this Article. For instance, should merchants be responsible for knowing the laws of all the countries of all the customers they deal with? Perhaps not, but if 90% of their sales are from an offshore jurisdiction, they should be responsible for knowing that sales of their product are legal in that jurisdiction. Violations of the policy would largely be dealt with on a complaint basis.

360. *See, e.g.*, Press Release, Privacy Int’l, Europe’s Privacy Commissioners Rule Against SWIFT (Nov. 23, 2006), *available at* [http://www.privacyinternational.org/article.shtml?cmd\[347\]=x-347-546365](http://www.privacyinternational.org/article.shtml?cmd[347]=x-347-546365) (describing the SWIFT case, where SWIFT was required to comply with U.S. demands for access to financial information about European customers in virtue of its operations on U.S. soil, while such compliance put them in violation of the European data protection directive). In addition, passage of the Global Online Freedom Act (GOFA) could put internet intermediaries in a conflict of law situation with China and other countries. *See* Global Online Freedom Act of 2007, H.R. 275, 110th Cong. (2007). H.R. 275 was introduced by Representative Chris Smith on January 5, 2007 and would require U.S. intermediaries to resist certain orders from countries in which they are doing business. *Id.*

payment intermediaries, local jurisdictions reach out to the local affiliates that are totally within their jurisdiction. They do not put burdens on entities in foreign jurisdictions at all. There is literally no conflict and thus nothing to which normal mechanisms of conflict resolution may attach.<sup>361</sup>

Some commentators have correctly pointed out that when the laws of different jurisdictions apply to a single transaction, the ability of any particular jurisdiction to unilaterally regulate the Internet is limited.<sup>362</sup> But intermediaries can reduce these conflicts. Global payment systems can simplify transactions to events in which only a buyer in one jurisdiction and a seller in another are implicated. By concentrating enforcement in intermediaries instead of individuals or merchants, local jurisdictions can take advantage of the economies that these institutions make possible.

The experience of payment intermediaries reveals that, within limits, the differences among conflicting jurisdictions can be managed. The bordered Internet works on a small scale. The scale is currently small for two reasons: the number of cases of governments reaching across borders to inflict their laws on internet merchants in other jurisdictions is still relatively small. Moreover, in contrast to the rhetoric about the Internet creating a global marketplace, the scope of cross-border commerce itself is still limited. The reality is that the volume of cross-border transactions is not large enough to create a truly substantial cross-border jurisdictional crisis. Currently, only four percent of the sales for electronic commerce merchants in the U.S. come from abroad.<sup>363</sup> And data from Europe show that cross border online transactions are not increasing as fast as overall e-commerce transactions, staying relatively stable from 2006 to 2008 at six to seven percent.<sup>364</sup>

361. Antigua brought a complaint against the U.S. for the enforcement of its gambling laws, but its success was based only on (1) the U.S.'s failure to exclude internet gambling from the list of services that required open treatment and (2) the idiosyncrasies of U.S. gambling law which appear to allow domestic horse racing to engage in internet gambling while denying similar opportunities to offshore internet gambling merchants. But these are technical obstacles created by the interaction of complex U.S. law and international WTO law and are not real conflict of law problems. See Appellate Body Report, *supra* note 130, 358–64.

362. See, e.g., H. Brian Holland, *The Failure of the Rule of Law in Cyberspace?: Reorienting the Normative Debate on Borders and Territorial Sovereignty*, 24 J. MARSHALL J. COMPUTER & INFO. L. 1, 26 (2005).

363. This is based on transaction data from the Visa system. See *International Piracy Hearing*, *supra* note 246, at 75 (statement of Mark MacCarthy, Senior Vice President for Global Public Policy, Visa Inc.).

364. Comm'n of the European Cmty's., *Commission Staff Working Document: Report on Cross-Border E-commerce in the EU 3*, SEC (2009) 283 final (Mar. 5, 2009), available at [http://ec.europa.eu/consumers/strategy/docs/com\\_staff\\_wp2009\\_en.pdf](http://ec.europa.eu/consumers/strategy/docs/com_staff_wp2009_en.pdf) ("From 2006 to 2008, the share of all EU consumers that have bought at least one item over the Internet increased from 27% to 33% while cross-border e-commerce remained stable (6% to 7%).").

As David Post has warned, the problem the Internet creates for local jurisdictions is one of scale.<sup>365</sup> The bordered Internet simply does not scale up. Global payment systems cannot accommodate an enforcement burden in which each jurisdiction uses payment system mechanisms to enforce each of its local laws on the Internet.

It is not hard to see how a “tragedy of the commons” could arise in this area. Each individual extension of local jurisdiction into cyber space seems small and costless, but collectively the burden becomes unbearable. Governments might feel free to exploit this enforcement mechanism, in the same way that grazers use the commons—under the impression that it is an unlimited resource. However, one of two outcomes will occur as the cross-border rules pile up: either cross-border transactions will remain small and the potential for the Internet to be a global channel of commerce will not be realized, or the political costs of each government attempting to regulate the e-commerce activities of other countries will mount. Either development reveals the limitations of the bordered Internet as a long-term framework for internet governance.

Goldsmith and Wu suggest that enforcement of internet regulations through intermediaries is necessarily limited in size.<sup>366</sup> They suggest that maybe the system will not be able to scale up, but it won’t have to.<sup>367</sup> Small countries such as Antigua cannot enforce internet rules because global intermediaries can simply pull up stakes and leave if the rules are too strict.<sup>368</sup> However, there are a sufficiently large number of countries that global intermediaries will not feel capable of abandoning. If all of them use the intermediary enforcement mechanism, the system will be overwhelmed.

---

365. See Post, *Against “Against Cyberanarchy,”* *supra* note 15, at 1377 (stating that “scale matters”); see also Holland, *supra* note 362, at 29. Holland states,

The online actor cannot know, as a practical matter, the many laws applicable to a particular act, nor when one or more sovereign may decide to attempt regulatory action. This is particularly true in those areas of regulation in which morality, religion and culture are at their most influential, such as speech, race, sex, and even intellectual property. Moreover, it is not simply one actor or a few legal systems. It is an exponential multitude.

*Id.*

366. GOLDSMITH & WU, *supra* note 7, at 81–82.

367. *Id.* at 81.

368. See *id.* at 160 (suggesting that acting as the internet police is just a normal cost of doing business for global companies, which they can avoid in a particular case by leaving a country that tried to impose costs that exceeded the benefits of continued presence in the country and thus creating another objection to the bordered Internet to effectively give larger countries a greater role in internet governance than smaller ones).

## C. INTERNATIONALISM

The fundamentally correct insight of the internet exceptionalists is that the unilateral imposition of one nation's law onto all internet activities that cross borders won't scale.<sup>369</sup>

Internationalism might be the way out. It is the idea that the Internet will eventually be governed, at least for some services, by global institutions and arrangements, and that this is the right public policy for local governments to follow in their dealings with illegal cross border internet transactions.<sup>370</sup> This policy could be implemented through a uniform global standard, or any of a variety of techniques such as WTO rules that bring local laws into harmony. The basic justification for this policy is similar to the justification for establishing a single uniform national policy that prevents the clash of inconsistent rules at the state level: when activities have widespread and significant effects on those outside the local jurisdiction, then uniform principles or some other coordinating mechanism should be adopted at the higher level.<sup>371</sup> This universalism could promise better laws, whereby the “[i]nternational standards could reflect a kind of collection of best practices from around the world—the opposite of the tyranny of the unreasonable.”<sup>372</sup>

Goldsmith and Wu make several criticisms of internationalism. First, a system of universal laws would be unattractive; it would leave the world divided and discontent because the universal law would be unpopular in large segments of the world population. Second, the system of local national laws would better reflect differences among people. Diversity is a good thing and cannot be taken into account by a universal code that overrides local differences. Third, it is not needed. The conflicts of laws, extraterritoriality,

369. See Johnson & Post, *supra* note 1, at 1390 (“One nation’s legal institutions should not monopolize rule-making for the entire Net.”).

370. GOLDSMITH & WU, *supra* note 7, at 26.

371. *Id.* (“If the nations of the world agree to a single global law for questions like libel, pornography, copyright, consumer protection, and the like, the lives of Internet users become much simpler: no conflicting laws, no worries about complying with 175 different legal systems, no race to the bottom.”).

372. *Id.* at 27. Reidenberg also argues that as jurisdictions increasingly conflict there will need to be an overarching harmonization of international rules:

[O]nline enforcement with electronic blockades and electronic sanctions will cause serious international political conflicts. These conflicts arise because of the impact on territorial integrity. Such conflicts are likely to force negotiations toward international agreements that establish the legal criteria for a state to use technological enforcement mechanisms. This progression leads appropriately to political decisions that will define international legal rules.

Joel R. Reidenberg, *States and Internet Enforcement*, 1 U. OTTAWA L. & TECH. J. 213, 230 (2003–2004).

and other considerations are perfectly manageable within the current international framework. For example, since most internet users do not have assets in other countries, they are effectively subject only to the laws of the country where they live. Only large multinational companies with assets all over the world face the multijurisdictional problem, and they already have to live with that because they are already global. Compliance with a plurality of international laws is simply a cost of doing business for global companies. There's nothing new here that would justify a move to a more harmonized global order. There are extra costs to be sure, but nothing so onerous or burdensome that it would require a move to global law.<sup>373</sup>

The responses to these criticisms are straightforward. An unpopular global law is not the goal. Neither is suppression of diversity the goal. The idea is to integrate local laws in some fashion when the regular conflicts among them prove to be intolerable. When diversity does not create this difficulty, there is no need for integration. If, for example, local governments value diversity enough to refrain from using intermediaries to enforce local laws against actors in other jurisdictions, then there is no need for harmonization of these enforcement efforts. But to the extent that governments want to take global enforcement steps, they also need to take steps to integrate the laws they want to enforce. The reason for this is that global intermediaries' costs to mediate the conflicts associated with unilateral attempts at local regulation of the Internet will be so onerous and burdensome that they will cause an unwarranted and unnecessary decline in global interaction.<sup>374</sup>

Berman also describes how the internationalist hope for global standards avoids the conflict of law problem: "if we constructed one universal 'world community' with one set of governing rules, there would never need to be a 'choice of law' in the sense that conflict-of-laws scholars use the term."<sup>375</sup> However, he is critical of this universal world community, for two reasons. First, because of its potential to dissolve community affiliations that provide important emotional connections and opportunities for normative discussion of those connections. Second, he views this universal community as fundamentally unrealistic given the dominance of current notions of nation-state sovereignty.<sup>376</sup>

---

373. See GOLDSMITH & WU, *supra* note 7, at 152–60.

374. Interestingly, the earlier Jack Goldsmith seemed more inclined to accept these practical considerations as a rationale for international harmonization: "When in particular contexts the arbitrariness and spillovers become too severe, a uniform international solution remains possible." Goldsmith, *supra* note 2, at 1235.

375. Berman, *supra* note 346, at 1860.

376. *Id.* at 1860–61.

These objections can be met at the level of generality at which they are cast. We do not need to think of ourselves as primarily world citizens in order to endorse specific global approaches. We can still have deep attachments to local communities and can still debate the relative importance of the overlapping communities we participate in. The global approach endorses the view that self-government “requires a politics that plays itself out in a multiplicity of settings, from neighborhoods to nations to the world as a whole . . .” and “citizens who can abide the ambiguity associated with divided sovereignty, who can think and act as multiply situated selves.”<sup>377</sup> But participation in global community and the wisdom to know when the global perspective should take precedence over more local concerns is essential to this vision of self-government in a global world.

The internationalist proposal is to provide global coordination only when necessary. It is to move to global standards when, as a practical matter, the burdens of allowing diverse local rules are too high. The model of national uniform standards is appropriate: not everything has to be done at the national level, but some things should be done there in order to have an efficient and fair national system. Similarly, there is no need to move from the current system to a world government. But if there are practical ways to improve internet governance through global harmonization, they should be taken.

If governments are going to use payment intermediaries as enforcers of local law, there are a number of steps that could be taken to coordinate their efforts, including:

- In the internet gambling context, a move to an internationally interoperable licensing system that would require each jurisdiction that allows internet gambling to defer to the licensing decisions of other jurisdictions;
- In the controlled substance and child pornography context, a globally coordinated web searching mechanism that would replace the individual monitoring efforts of the intermediaries;
- In the controlled substance context, an internationally accepted list of controlled substances and entities licensed to provide the substances for sale online; and
- In the copyright context, the continued evolution of uniform copyright rules.

---

377. MICHAEL J. SANDEL, PUBLIC PHILOSOPHY: ESSAYS ON MORALITY IN POLITICS 34 (2005).

International agreements are one mechanism to create coordinated action. Although controversial because of the secrecy involved in its development, and the sense that affected parties were excluded from participation, the Anti-Counterfeiting Trade Agreement (ACTA) is a reasonable, though flawed, model for action in this area.<sup>378</sup> There are many mechanisms for international coordination. Decisions regarding which mechanisms to use depend on the issue and the fora available for resolution.

Internationalism has its dangers. Why should each jurisdiction have the same regulations on hate speech and the same regulations on alcohol consumption? The answer is that there will be no harmonization where there are such fundamental differences. Intermediaries will be called upon to resolve the issue themselves or they will be caught between warring governments and forced to choose sides. But efforts should be made to minimize such differences when these differences have global consequences, especially when they are superficial differences that reflect no fundamental divisions. For the same reason that we want uniform global technical standards for information and communications technologies, if possible, we want similar legal frameworks if governments are going to enforce laws on the Internet.

These efforts to ease the friction involved in extending government authority to the Internet through a global framework are in line with other efforts to create global frameworks that promote the growth of the Internet. For example, the thirty-first International Conference of Data Protection and Privacy Commissioners, held in Madrid in November 2009, adopted a set of global privacy standards.<sup>379</sup> There is also likely to be a renewed push for global consumer protection on the occasion of the tenth anniversary of the Organisation for Economic Co-operation and Development's (OECD)

---

378. See Media Statement, Participants in ACTA Negotiations, Anti-Counterfeiting Trade Agreement (ACTA) (June 12, 2009), available at [http://www.med.govt.nz/templates/Page\\_\\_\\_\\_40974.aspx](http://www.med.govt.nz/templates/Page____40974.aspx). For a summary of the ACTA process and the content of the agreement, see the Summary of Key Issues. THE ANTI-COUNTERFEITING TRADE AGREEMENT—SUMMARY OF KEY ELEMENTS UNDER DISCUSSION (2009), available at [http://www.med.govt.nz/templates/MultipageDocumentTOC\\_\\_\\_\\_40563.aspx](http://www.med.govt.nz/templates/MultipageDocumentTOC____40563.aspx).

379. Artemi R. Lombarte, Dir., Agencia Española de Protección de Datos, Slide Presentation: International Standards on Data Protection & Privacy (2009), available at [https://www.agpd.es/portalweb/canaldocumentacion/comparencias/common/IAPP\\_Privacy\\_Summit\\_09.pdf](https://www.agpd.es/portalweb/canaldocumentacion/comparencias/common/IAPP_Privacy_Summit_09.pdf). He describes one of the main criteria of the global privacy standards project as “[t]o elaborate a set of principles and rights aimed to achieve the *maximum degree of international acceptance*, ensuring at once a high level of protection.” *Id.* (emphasis in original). For the standards adopted, see The Madrid Privacy Declaration (Nov. 3, 2009), <http://thepublicvoice.org/TheMadridPrivacyDeclaration.pdf>.

Guidelines for Consumer Protection in the Context of Electronic Commerce.<sup>380</sup>

Both these efforts relate to the growth of the Internet as a vibrant international marketplace. They do this by building online trust. Global information security standards reassure people that their information is safe regardless of the physical location of the websites they visit. Establishing global privacy standards means that the collection and use of online information will be governed by common principles regardless of a website's jurisdiction and will make it easier for global business to transfer information from one jurisdiction to another in a seamless manner. Finally, effective global consumer protection rules will mean that people will have the information and redress rights they need to shop confidently online no matter where the website is located.

## V. CONCLUSION

The initial demand from internet exceptionalists that the online world be left alone by governments has morphed, as this Article explains, into the demand that governments create a global framework to protect and spur the growth of the Internet. The intervening steps in this development are not hard to trace: internet exceptionalists confused their ideal of self-governing internet communities with the idea that the Internet was ungovernable because it was a global communications network that crossed borders. This idea was undermined by the recognition that the coding that underlies internet applications and services is a matter of choice, not unchangeable nature. If something about this system created difficulties for government control, this could be changed. Further, the idea that governments cannot control the Internet was undermined by the ability of global intermediaries' local operations to provide essential services and the practical ability of governments to control these intermediaries. Examples from the payment card world demonstrate how this was done in internet gambling, child

---

380. OECD GUIDELINES, *supra* note 353; *see also* ORG. FOR ECON. CO-OPERATION & DEV., CONFERENCE ON EMPOWERING E-CONSUMERS: STRENGTHENING CONSUMER PROTECTION IN THE INTERNET ECONOMY, PROGRAMME (2009), *available at* <http://www.oecd.org/dataoecd/33/22/44045376.pdf> (describing the conference). The OECD endorsed steps toward global enforcement of some consumer protection rules in a 2003 report on cross-border fraud and a 2007 report on consumer dispute resolution and redress. *See* COMM. ON CONSUMER POLICY, ORG. FOR ECON. CO-OPERATION & DEV., OECD GUIDELINES FOR PROTECTING CONSUMERS FROM FRAUDULENT AND DECEPTIVE COMMERCIAL PRACTICES ACROSS BORDERS (2003), *available at* <http://www.oecd.org/dataoecd/24/33/2956464.pdf>; COMM. ON CONSUMER POLICY, ORG. FOR ECON. CO-OPERATION & DEV., OECD RECOMMENDATION ON CONSUMER DISPUTE RESOLUTION AND REDRESS (2007), *available at* <http://www.oecd.org/dataoecd/43/50/38960101.pdf>.

pornography, controlled substances, online tobacco, and copyright infringement.

These examples prove that intermediaries can control the content of the activities on their online communities, and that government can compel or pressure intermediaries to take these steps. Intermediaries have a general obligation to follow the law, and except in extreme cases, they have no right to resist these lawfully established burdens.

Yet, the question remains: should the government place this enforcement burden on intermediaries? The advantages of government intervention sometimes appear to be substantial, but nothing in the least cost arguments suggests that internet intermediaries are always the best vehicle for government control. The costs, benefits, and equities involved in specific cases have not been adequately assessed. Intermediaries are often in a position to voluntarily police their own communities and have taken steps to do this without explicit government requirements. The equities set out in current law establish a regime that works tolerably well. Even when government requirements are explicit, as in the internet gambling case, they are often crafted to fit the architecture and structure of the intermediaries themselves. While some adjustments would improve these legal regimes, nothing suggests that more liability imposed unilaterally by local governments would be an improvement. A return to internet exceptionalism would not help matters either.

Greater government coordination on the rules that intermediaries must follow on the Internet would be an improvement. To avoid legal liability and to comply with local laws, payment intermediaries are moving toward accepting the laws of all jurisdictions. They also have wide discretion on what activities to allow on their systems. But this situation is problematic. Intermediaries are not the best-situated to decide which rules to follow. Also, no laws are self-interpreting. They often apply to particular situations in obscure and heavily fact-dependent ways. Intermediaries' flexibility in adjudication leaves room for private, strategic, and unaccountable decisions that affect the shape and direction of online activity. Coordinated government rules are best for an additional reason: the intermediary role does not scale well in a world of multiple, overlapping, and conflicting rules. If governments are going to use intermediaries to regulate the Internet, they need to coordinate their own laws to make that role possible.

# LICENSING COMPLEMENTARY PATENTS: “PATENT TROLLS,” MARKET STRUCTURE, AND “EXCESSIVE” ROYALTIES

*Anne Layne-Farrar<sup>†</sup> & Klaus M. Schmidt<sup>‡</sup>*

## TABLE OF CONTENTS

I.	<b>INTRODUCTION</b> .....	1121
II.	<b>COMPLEMENTS, DOUBLE MARK-UPS, AND RAISING ONE’S RIVALS’ COSTS</b> .....	1126
	A. NON-INTEGRATED PATENT HOLDERS.....	1128
	B. VERTICALLY INTEGRATED PATENT HOLDERS .....	1129
III.	<b>POOLING AGREEMENTS AND CROSS-LICENSING</b> .....	1132
IV.	<b>NON-LINEAR AND DISCRIMINATORY ROYALTIES</b> .....	1137
V.	<b>CONCLUSIONS</b> .....	1139
	<b>APPENDIX: BILATERAL CROSS-LICENSING</b> .....	1142

## I. INTRODUCTION

This Article challenges a common definition of “patent troll” as any non-practicing patent holder. The association between non-practicing patent holder and troll was made in the infamous *NTP, Inc. v. Research in Motion, Ltd.* decision (the *BlackBerry* case).<sup>1</sup> In early 2006, the e-mail correspondence of millions of BlackBerry users nearly came to a halt when NTP, Inc. accused Research in Motion, Ltd. (“RIM”), the maker of the popular communication device, of infringing several of its patents.<sup>2</sup> The Federal Circuit found that BlackBerry’s e-mail retrieval system was indeed infringing on some of NTP’s

---

© 2010 Anne Layne-Farrar and Klaus M. Schmidt.

<sup>†</sup> Economist at LECC Consulting.

<sup>‡</sup> Professor of economics at the University of Munich. The authors would like to thank Matthew Bennett, Gerard Llobet, Jorge Padilla, Richard Schmalensee, and Monika Schnitzer for their comments and suggestions and Alina Marinova and Sokol Vako for research assistance. Financial support from Qualcomm is gratefully acknowledged. The ideas and opinions in this paper are exclusively our own. Corresponding author: Klaus Schmidt (klaus.schmidt@LMU.de).

1. 397 F. Supp. 2d 785 (E.D. Va. 2006).

2. *Id.*

patents and awarded damages and a permanent injunction against RIM.<sup>3</sup> Armed with the injunction, NTP threatened to shut down BlackBerry's e-mail services if RIM did not pay royalties for the future use of NTP's patents.<sup>4</sup> In March 2006, in order to avert the injunction, RIM agreed to pay 612.5 million dollars in a last minute settlement, an amount significantly greater than the past damages of 33.5 million dollars awarded by the trial court.<sup>5</sup>

The plaintiff in this highly visible case is often quoted as an example of a so called "patent troll."<sup>6</sup> This term is used to describe a company that uses a patent to "hold-up" manufacturing companies and to extort "excessive" royalties that are higher than the "fair share" dictated by the contribution of its patent. Unfortunately, it is notoriously difficult to determine whether royalties are "excessive," and to distinguish between a hold-up and aggressive, but legitimate, bargaining. Because a workable definition is lacking, patent trolls are sometimes associated with entities that do not develop innovations of their own and even more frequently with "non-practicing" or "non-manufacturing" patent holders (often referred to as "NPEs," for non-practicing entities).<sup>7</sup> In the Blackberry example, NTP is an NPE. It does not produce cell phones or any other goods but owns a portfolio of patents that it licenses to manufacturing companies. In contrast to a vertically integrated firm (which both holds a patent and uses it to produce a good in a downstream market), an NPE does not require cross-licenses from competitors in the downstream market. Therefore, it is claimed

---

3. NTP, Inc. v. Research in Motion, Ltd., 418 F.3d 1282 (Fed. Cir. 2005).

4. Stephen Lawson, *BlackBerry Users Face Shutdown*, PC WORLD, Dec. 1, 2005, available at [http://www.pcworld.com/article/123761/blackberry\\_users\\_face\\_shutdown.html](http://www.pcworld.com/article/123761/blackberry_users_face_shutdown.html).

5. For a more detailed discussion of this case, see Mark A. Lemley & Carl Shapiro, *Patent Holdup and Royalty Stacking*, 85 TEX. L. REV. 1991, 2009 n.36 (2007); James F. McDonough III, Comment, *The Myth of the Patent Troll: An Alternative View of the Function of Patent Dealers in an Idea Economy*, 56 EMORY L.J. 189, 194-95 (2006).

6. Peter Detkin, former assistant general counsel for Intel, coined the term "patent troll" after Intel was sued for libel for its use of the term "patent extortionist." Detkin explains that "[a] patent troll is somebody who tries to make a lot of money off a patent that they are not practicing and have no intention of practicing and in most cases never practiced." McDonough III, *supra* note 5, at 192; see also Ian Austen & Lisa Guernsey, *A Payday for Patents 'R' Us; Huge Blackberry Settlement Is Grist for Holding Company*, N.Y. TIMES, May 2, 2005, at C1.

7. See, e.g., John M. Golden, "Patent Trolls" and Patent Remedies, 85 TEX. L. REV. 2111, 2112 (2007); McDonough III, *supra* note 5, at 189; Carl Shapiro, Injunctions, Hold-Up, and Patent Royalties 3 (Aug. 2006) (unpublished manuscript, on file with the Haas School of Business, University of California). As Lemley & Shapiro put it: "Defining a patent troll has proven a tricky business, but that does not mean the problem does not exist. Nonpracticing entities file 30-40% of all patent suits in the computing and electronics industries, for example." Lemley & Shapiro, *supra* note 5, at 2009.

that an NPE is not constrained in its behavior and may choose unjustifiably high royalty rates that are not in proper relation to the contribution of its patents.

The identification of NPEs and patent trolls has important legal consequences. In *eBay Inc. v. MercExchange, L.L.C.*,<sup>8</sup> Justice Kennedy stated in his concurring opinion that there are firms that “use patents not as a basis for producing and selling goods but, instead, primarily for obtaining licensing fees . . . . For these firms, an injunction . . . can be employed as a bargaining tool to charge exorbitant fees . . . .”<sup>9</sup> Kennedy’s statement implied that the lower courts should be careful in granting injunctive relief to non-practicing patent holders. In fact, since *eBay*, many district courts have denied injunctive relief to non-manufacturing and non-competing firms.<sup>10</sup>

This Article analyzes how patent holders choose their royalties depending on their business model (vertically integrated or not), the structure of the upstream and downstream markets, and the type of licensing agreements feasible: linear and non-linear royalties, cross-licensing, and patent pools. This Article shows that an NPE has different incentives regarding royalty rates than a vertically integrated company. However, there is no reason to presume that a non-integrated patent holder will charge *higher* rates than vertically integrated companies. To the contrary, the vertically integrated firm has an incentive to raise its royalties in order to raise its rivals’ costs and to restrict entry in the downstream market, which does not hold true for non-integrated patent holders. Thus, an integrated firm may charge higher royalties than an NPE.

8. *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388 (2006).

9. *Id.* at 396 (Kennedy, J., concurring) (joining in concurrence Stevens, Souter, and Breyer, JJ.) (citation omitted).

10. See John L. Dauer Jr. & Sarah Elizabeth Cleffi, *Trends in Injunctive Relief in Patent Cases Post-eBay*, THE METRO. CORP. COUNSEL, Feb. 2007, at 16. Drauer and Cleffi state:

[A] survey of the cases decided since *eBay* proves useful in identifying one trend in the decisions. What has become apparent thus far is the district courts’ attention to the considerations expressed in Justice Kennedy’s concurrence and whether the parties are in direct competition. . . .

. . . .  
 . . . When the parties are not in direct competition, the courts . . . would likely find monetary damages are adequate . . . . To date, district courts appear to have thus far heeded Justice Kennedy’s warnings in his *eBay* concurrence and not issued injunctions to such parties.

*Id.*; see also Bernard H. Chao, *After eBay, Inc. v. MercExchange: The Changing Landscape for Patent Remedies*, 9 MINN. J. L. SCI. & TECH. 543, 553 (2008) (“[T]he existence of direct competition generally results in a permanent injunction. The converse is also true. Lack of direct competition generally results in the denial of a permanent injunction.”).

There are a few recent papers in the law and economics literature dealing with the problem of patent trolls and NPEs.<sup>11</sup> Several authors have pointed out that NPEs, such as universities, government-sponsored research labs, and some high technology companies, can and do perform important and valuable functions in a market economy.<sup>12</sup> For example, many NPEs concentrate on their comparative advantage of conducting research and development while leaving the manufacturing of final products to other companies.<sup>13</sup> NPEs can also foster the dissemination of new technologies and encourage entry in the downstream market because their primary, if not only, source of revenue relies on making their innovations tradable by patenting and licensing.<sup>14</sup> Other NPEs can act as intermediaries, buying and selling patents to provide liquidity and increased efficiency to technology markets.<sup>15</sup> This literature establishes that a general condemnation of non-practicing patent holders as “trolls” is clearly not warranted. However, none of these papers addresses the question of whether NPEs may have an incentive to charge higher royalties than vertically integrated firms.

In a seminal paper, Carl Shapiro identifies two sets of conditions under which a hold-up problem may give rise to patent troll behavior.<sup>16</sup> First, the manufacturing firm is unaware of the patent when it invests in the production of its product.<sup>17</sup> In this case, after the investment is sunk, the patent holder can extort supra-normal royalties by threatening to obtain an

---

11. See, e.g., Mark Lemley, *Are Universities Patent Trolls?*, 18 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 611 (2008).

12. For instance, universities are arguably NPEs, but are productive elements of the economy. See *id.* at 629–30. Likewise, research firms with no manufacturing arms, such as Qualcomm and Palo Alto Research Center (PARC) specialize in R&D and can contribute valuable technologies to society, albeit for others to make and commercialize. Finally, patent aggregators such as RPX Corporation, Intellectual Ventures, and OPTI INC can play an important market-making function. See McDonough III, *supra* note 5, at 211.

13. Motivated by the example of universities, Lemley refines the definition of a patent troll. He argues that universities are not patent trolls because they are actively engaged in technology transfer, while patent trolls are non-manufacturing entities that do not engage in technology transfer but instead license only the right not to be sued. However, Lemley concedes that this definition is too abstract and could easily be gamed if applied by the courts. See Lemley, *supra* note 11, at 629–30.

14. See Damien Gerardin, Anne Layne-Farrar, & Jorge Padilla, *Elves or Trolls? The Role of Non-Practicing Patent Owners in the Innovation Economy* (TILEC Discussion Paper No. 2008-018, 2008), available at <http://ssrn.com/abstract=1136086>; Sannu K. Shrestha, Note, *Trolls or Market-Makers? An Empirical Analysis of Nonpracticing Entities*, 110 COLUM. L. REV. 114, 126–31 (2010) (reviewing the various arguments put forth in support of NPEs).

15. See McDonough III, *supra* note 5, at 213–15.

16. Shapiro, *supra* note 7, at 10–11. See generally Lemley & Shapiro, *supra* note 5, for a more detailed discussion of the implications of this model.

17. Shapiro, *supra* note 7, at 11.

injunction and shut down production of the entire product. Second, the manufacturing firm is aware of the patent before it invests in the production of its product, but the patent is “weak,” i.e., the probability that it will be declared valid by a court is considerably smaller than one.<sup>18</sup>

In the weak patent scenario, Shapiro’s model assumes two possible outcomes. First, the manufacturing firm can start production without a license and wait for a decision of the court, facing the same sunk cost hold-up problem as described above. Alternatively, the manufacturer can threaten to invent around the patent, such that negotiations with the patent holder are based on the assumption that the patent is valid with certainty. In both cases the patent holder will get supra-normal royalties.

The scenarios, *supra*, considered by Shapiro are illustrations of the hold-up problem and fit the stylized facts of the *BlackBerry* case, but they have nothing to do with the distinction between non-practicing and vertically integrated patent holders. If one of Shapiro’s model hold-up problems arises, both an NPE and a vertically integrated patent holder would have the exact same incentive to exploit it.

This Article considers a set-up that allows for different market structures and different types of licensing contracts. In the upstream market, there are one or more patent holders, each of whom has at least one patent that is essential for the production of the downstream good.<sup>19</sup> All parties are aware of all patents and their validity is not in dispute. Thus there is no hold-up problem stemming from weak patents,<sup>20</sup> but each patent holder has considerable market power because he can threaten to interrupt downstream production if the downstream firms are unwilling to accept his royalties.

In Part II this Article analyzes how patent holders choose their royalties depending on their business model (vertically integrated or not) and the structure of the upstream and downstream markets (in situations where all firms act non-cooperatively and set individual linear and non-discriminatory royalties). It is well known in economics that vertical integration eliminates

---

18. *Id.* at 21.

19. This Article restricts its attention to the case in which the patents at issue are essential for downstream production, so there are no feasible substitutes. This is the most suitable case for its purposes since the presence of viable substitutes prevents hold-up. If there are imperfect but viable substitutes for the patent upstream, the analysis is more complicated. See Daniel Quint, *Economics of Patent Pools When Some (But Not All) Patents Are Essential* (SIEPR Discussion Paper 06-028, 2009), available at <http://www.ssc.wisc.edu/~dqunt/papers/patent-pools-quint.pdf>.

20. This Article finds this form of hold-up far less plausible since licensees can and do challenge the validity of a patent and need not negotiate as if the patent were valid and infringed.

the vertical double mark-up problem within the vertically integrated firm (but not across firm boundaries), which leads to lower royalties.<sup>21</sup>

Part III evaluates the situation of coordinated royalty setting in the upstream market, i.e., cross-licensing agreements or patent pools. Part IV extends the analysis to the case where patent holders can charge non-linear royalties, such as those that combine an upfront fee with a running royalty rate.

These results will show that non-integrated companies are constrained by the impact their behavior has on the downstream market and, in some cases, may charge *lower* royalties than their vertically integrated counterparts. Ultimately, whether a company charges excessive royalties depends on whether there is scope for hold-up, because of either sunk investments or weak patents (coupled with high litigation costs). These factors are orthogonal to whether patent holders are vertically integrated or not. The results suggest that two commonly applied definitions of patent troll—either all non-practicing patent holders or all non-innovating patent collectors—are misleading. The business model of the patent holder is not the key factor for identifying a patent troll. Instead, the crucial factor is the weakness of the patents at issue or the presence of factors that facilitate hold-up in the circumstances at hand.

## II. COMPLEMENTS, DOUBLE MARK-UPS, AND RAISING ONE'S RIVALS' COSTS

To clarify the assumptions and the logic behind this Article's conclusions, this Part develops a simple economic model. Consider a high technology good, such as a cell phone or a DVD player, that is based on a technological standard requiring the use of a number of different patented technologies to be operational. Each of the patents is essential in the sense that no firm can legally produce the good without access to each patent. The essential patents are often owned by different companies and each firm that produces the good requires a license from each of the patent holders. The market for licenses will be referred to as the "technology" or "upstream market" and the market for the good (for example, the cell phone or the DVD player) will be the "product" or "downstream market." Some firms may be vertically integrated in that they hold an essential patent and produce the final good. Other companies may be non-vertically integrated: either they hold a patent but do not use it to produce the final good themselves, or they produce the

---

21. *See, e.g.*, DENNIS CARLTON & JEFFREY PERLOFF, MODERN INDUSTRIAL ORGANIZATION 417 (4th ed. 2005).

final good but do not hold any essential patents. These will be referred to as non-integrated companies—upstream and downstream firms, respectively. To simplify the model, this Article assumes that each company in the upstream market owns exactly one essential patent and that all firms in the downstream market have the same cost functions and produce end goods that are substitutes.

The focus of this analysis is on the royalties that will be charged by vertically integrated and non-integrated patent holders. In this Part, it will be assumed that all patent holders are restricted to using linear, non-discriminatory royalties.<sup>22</sup> The prevalence of linear royalty rates in real world licensing contracts can be explained by asymmetries of information between the licensee and the licensor and their risk-sharing properties.<sup>23</sup> Non-discriminatory royalties are often explicitly noted in contracts and are a common commitment made during standard setting (known as Reasonable and Non-Discriminatory licensing (RAND)). Part IV extends this analysis to the case where firms can use two-part tariffs and can discriminate between different downstream producers.

What royalties will be charged by a vertically integrated as opposed to a non-integrated patent holder? The answer depends on the market structure. There are several effects that have opposing impacts on vertically integrated patent holders when compared with non-integrated patent holders.<sup>24</sup> In general, it is ambiguous whether a non-integrated patent holder or a vertically integrated firm charges higher or lower royalties. Nevertheless, it is highly instructive to understand the different effects in order to evaluate their relative importance.

---

22. In other words, for now it can be assumed that the straightforward royalty rates comprise the license payment; non-linear royalty schedules and lump sum fees are not used.

23. If there are asymmetries of information, linear royalties can be used as optimal screening or signaling devices. See, e.g., A.W. Beggs, *The Licensing of Patents Under Asymmetric Information*, 10 INT'L J. INDUS. ORG. 171 (1992) (discussing optimal screening); Alain Bousquet et al., *Risk Sharing in Licensing*, 16 INT'L J. INDUS. ORG. 535 (1998) (characterizing licensing agreements with optimal risk sharing properties); Jay P. Choi, *Technology Transfer with Moral Hazard*, 19 INT'L J. INDUS. ORG. 249 (2001) (showing that linear royalties are optimal when there is moral hazard); Nancy T. Gallini & Brian D. Wright, *Technology Transfer Under Asymmetric Information*, 21 RAND J. ECON. 147 (1990) (discussing signaling devices).

24. As explained in Part II.B, vertical integration reduces double marginalization by combining the entities seeking a profit margin, which tends to lower the royalty rate. But vertical integration also creates incentives to raise downstream rivals' costs, which tends to increase the royalty rate.

## A. NON-INTEGRATED PATENT HOLDERS

A non-integrated upstream firm owning a single patent required for downstream production will exercise its monopoly over the patent and charge a royalty that maximizes profits. If the patent holder increases its royalty rate, it will increase the marginal cost that each downstream firm incurs. Downstream firms will, at least partially,<sup>25</sup> pass this cost increase through to their customers. Thus, downstream prices will increase and the quantities of the final good sold in the downstream market will decrease. When the monopolistic patent holder chooses its royalty rate, it considers that a higher royalty rate raises its profit on each unit but lowers the number of units sold. Because marginal costs are essentially zero for patent licensing, the patent holder will raise its royalties to the point that a one percent increase in the royalty rate yields a one percent decrease in the quantity sold. This is just the standard monopoly profit maximization problem as applied to patent licensing.

However, in contrast to the standard monopoly problem, there are several additional externalities that the monopolistic patent holder imposes on other active producers in these markets. First, when upstream patent holders choose their royalties they do consider that higher royalties might reduce the profits of the downstream firms. Specifically, if downstream firms cannot pass the increase in the royalty rate through to end prices, then higher royalties reduce the margins of downstream firms. Furthermore, any increase in end prices reduces the quantities that they can sell downstream. Since both the upstream and downstream producers require a profit margin—whereas for an integrated firm a single profit margin suffices—the result is known as a “double mark-up problem.” This problem typically arises when two vertically related firms both have market power.

Second, an upstream patent holder does not take into account that if he raises the royalty rate for his patent, he reduces the profits of the other patent holders, because the patents are perfect complements. This is called the “complements effect.”<sup>26</sup> Within a technical standard, patents are perfect complements: each is essential for producing the final good, so each unit sold in the downstream market requires exactly one license of each of the patents. If one patent holder raises its royalty and thereby reduces the total quantity

---

25. The pass-through rate may be larger than one, for example, in the case of a constant elasticity demand function.

26. *See generally* AUGUSTIN A. COURNOT, RESEARCHES INTO THE MATHEMATICAL PRINCIPLES OF THE THEORY OF WEALTH (A. M. Kelley 1971) (N.T. Bacon trans., 1838) (describing the effects for the first time).

of the final good sold downstream, this increase reduces the revenues of all the other patent holders.

In recent IP literature, the complements problem has acquired the moniker of “royalty stacking” because many firms’ royalty rates stack up to form a large cumulative burden for manufacturers.<sup>27</sup> Both the complements and the double mark-up effects may result in an aggregate royalty rate that is higher than the royalty rate a fully integrated monopolist (i.e., a monopolist who owns all essential patents and all downstream firms) would choose.

Both the complements effect and the double mark-up effect tend to raise royalties above their optimal level.<sup>28</sup> In fact, if all patent holders set their royalty rates independently and non-cooperatively, then the sum of all royalties would be higher than the total royalties a fully integrated monopolist charges. In this case, upstream and downstream firms, as well as consumers, would benefit if total royalties were lower. The effect occurs because as royalties fall, quantities sold rise, meaning higher royalty revenues for the patent holder.<sup>29</sup>

#### B. VERTICALLY INTEGRATED PATENT HOLDERS

A well-known remedy to mitigate the double mark-up problem is vertical integration. In a simple chain of monopolies, vertical integration eliminates the double mark-up problem. However, in a more complicated world with several upstream and downstream firms, vertical integration does not necessarily improve social welfare. For example, suppose one upstream and one downstream firm vertically integrate. When the upstream division of this integrated firm increases its royalty rate, it fully internalizes the effect on the profits of its downstream division. Note that the total firm profits remain unaffected because the royalty rate of its own downstream division is just an internal transfer payment that shifts profits from the downstream to the upstream division. The internalization mitigates the double mark-up problem within the integrated firm. However, with vertical integration there is a new strategic effect pointing in the opposite direction. By raising its royalty rate, the vertically integrated firm raises the costs of its downstream competitors without raising its own cost. Thus, by raising the marginal costs of its rivals, the integrated firm gains a competitive advantage, pushes its rivals’ prices

---

27. See Lemley & Shapiro, *supra* note 5, at 1993.

28. See *id.* at 2014.

29. In the region where prices exceed the monopoly level, the effect of increasing quantities outweighs the effect of falling royalty rates. See CARLTON & PERLOFF, *supra* note 21, at 91.

higher, and as a result, gains a higher market share downstream for the sale of its own good.<sup>30</sup>

The source of this “raising one’s rivals’ costs effect” is vertical integration.<sup>31</sup> A non-integrated patent holder does not participate in downstream profits. Thus an upstream patent holder benefits if the downstream market becomes more competitive, if there is market entry, and if more units of the final good are sold as a result. In contrast, a vertically integrated firm makes part, or most, of its profits in the downstream market. Therefore, since the vertically integrated firm wants this market to be less competitive, it will oppose market entry, and it will use the royalty rate for the patent of its upstream division to achieve an advantage over rival downstream firms that do not hold patents for cross-licensing.

To illustrate this point, suppose that there is just one upstream patent holder and a large number  $N$  of potential downstream firms with identical cost functions that compete in quantities in the downstream market. Suppose that a Cournot equilibrium<sup>32</sup> exists in the downstream market for any royalty rate  $r > 0$  charged by the upstream monopolist.<sup>33</sup> No other assumptions need to be imposed on the cost and demand functions.

---

30. The “raising one’s rivals’ costs” effect was first described by Salop and Scheffman. See Steven C. Salop & David T. Scheffman, *Raising Rivals’ Costs*, 73 AM. ECON. REV. 267 (1983); see also Steven C. Salop & David T. Scheffman, *Cost-Raising Strategies*, 36 J. INDUS. ECON. 19 (1987) (generalizing and extending the results from *Raising Rivals’ Costs*, as well as developing a theory of raising rivals’ costs through vertical integration showing that vertical integration can be anticompetitive). However, they restrict attention to a dominant firm that can affect marginal and average costs of a competitive fringe. They show that the dominant firm will raise its rivals’ cost in order to either foreclose the market or to induce competitors to raise their prices and to relax competition. The situation this Article is interested in is closer to the models examined by Ordober et al. and Kim, and considers a two-stage duopoly model with price competition and differentiated products. Their model looks at the more conventional case where the goods produced upstream are perfect substitutes while this Article looks at the opposite case of perfect complements. See Janusz A. Ordober, Garth Saloner & Steven C. Salop, *Equilibrium Vertical Foreclosure*, 80 AM. ECON. REV. 127, 138–40 (1990) (finding that a vertically integrated firm must be able to commit ex ante to a price for the input good, even though there is an incentive to reduce this price ex post). No such commitment is necessary in this Article’s analysis. Kim analyzes a similar model, but he restricts attention to the case of a linear demand curve. See Sung H. Kim, *Vertical Structure and Patent Pools*, 25 REV. INDUS. ORG. 231, 236–38 (2004).

31. See Klaus M. Schmidt, *Licensing Complementary Patents and Vertical Integration* (CEPR Discussion Paper No. 5987, 2006), available at <http://www.cepr.org/pubs/new-dps/dplist.asp?dpno=7005>.

32. See COURNOT, *supra* note 26, and accompanying text.

33. That is, an equilibrium where firms compete on quantities supplied and the equilibrium price is determined by the sum of these quantities. Novshek offers a set of fairly weak sufficient conditions that guarantee existence and uniqueness of an equilibrium in the Cournot model. He requires that there exists a  $\bar{q}$  (where the upper bar indicates an upper

Suppose that the patent holder is non-integrated and not active in the downstream market. In this case it will charge the monopolistic royalty rate  $r^M > 0$  that maximizes profits given the Cournot equilibrium in the downstream market. The larger the  $N$ , the more competitive the downstream market, the smaller the mark-up charged by the downstream firms, the greater the volume of goods sold, and the higher the patent holder's profits.

If the upstream patent holder vertically integrates with one of the downstream firms, one possible strategy of the integrated firm is to charge a royalty rate that is so high that no other downstream firm can make a profit which forces them to exit the market. The rate increase does not affect the downstream division of the integrated firm because within the integrated firm the royalty rate is a mere transfer price that has no effect on overall firm profits: each dollar spent on higher royalties by the downstream division is a dollar of revenue for the upstream division. Thus, by raising its royalty rate to a prohibitively high level, the integrated firm can foreclose all other downstream firms and monopolize the downstream market. Because a fully integrated monopolist (who controls both the upstream and the downstream market) maximizes profits, this strategy is indeed optimal.<sup>34</sup>

This extreme example shows that a vertically integrated firm has a natural inclination to use its royalty rate to raise the costs of its rivals in order to increase its profits in the downstream market. If the downstream firms are not all identical, but instead sell sufficiently differentiated products or have lower marginal costs than the downstream division of the integrated firm, then the integrated firm will not want to completely shut out all of its downstream rivals from the market.<sup>35</sup> The presence of other differentiated downstream firms extends the market, and increases the royalty income of the upstream division. However, the integrated firm still charges a royalty rate that discriminates against its downstream rivals. Thus the integrated firm's royalty rate may be higher than the royalty rate chosen by a non-integrated upstream firm under identical circumstances and patented technology.

---

bound for  $Q$ ) such that  $P(Q) > 0$  for all  $Q < \bar{Q}$  and that  $P(Q) = 0$  for all  $Q > \bar{Q}$ , and that  $P(Q)$  is twice continuously differentiable with the first derivative  $P'(Q) < 0$  for all  $Q < \bar{Q}$  (note that a single apostrophe is the typical mathematical symbol for a first derivative). Furthermore,  $P'(Q) + q \cdot P''(Q) < 0$  for all  $0 \leq q \leq \bar{Q}$ . William Novshek, *On the Existence of Cournot Equilibrium*, 52 REV. ECON. STUD. 85, 90–94 (1985).

34. And has thus been a real world problem. See, e.g., *United States v. Aluminum Co. of America*, 148 F.2d 416, 427 (2d Cir. 1945); Case T-5/97, *Industrie des Poudres Sphériques SA v. Comm'n*, 2000 E.R.C. II-3755.

35. See also X.H. Wang & B.Z. Yang, *On Licensing Under Bertrand Competition*, AUSTL. ECON. PAPERS, June 1999, at 106, for a similar result with Bertrand competition.

In conclusion, vertical integration has two effects on the market outcome. First, vertically integrated firms internalize the double mark-up effect between upstream and downstream divisions, which tends to improve efficiency. Second, the “raising one’s rivals’ costs effect” tends to raise royalties and to reduce efficiency. The specific structure of the cost and demand functions determines which of the two effects dominates. Klaus Schmidt shows in a much more general model, with arbitrarily many firms in the upstream and downstream markets and a general model of downstream competition, that a market in which all firms are vertically integrated may give rise to higher *or* lower total royalties than a market in which all firms are non-integrated.<sup>36</sup> For the special, but natural example of Cournot competition with linear demand and cost functions and identical firms, Sung H. Kim shows that if the number of vertically integrated firms is not too large, then vertical integration induces an equilibrium price in the downstream market that is strictly higher than the equilibrium price obtained under non-integration.<sup>37</sup> Thus, vertical integration may well reduce total output, total industry profit, and social welfare.

If there are several vertically integrated upstream firms, each of them has an incentive to raise the costs of the other vertically integrated firms by raising its royalties. This gives rise to the prisoners’ dilemma. All vertically integrated upstream firms would be better off if they would all charge lower royalties, but for each firm, it is optimal to charge high royalties. Part III shows that cross-licensing agreements or patent pools can solve this prisoners’ dilemma, although they do not necessarily solve the problem of raising rivals’ costs.

### III. POOLING AGREEMENTS AND CROSS-LICENSING

One assumes thus far that all firms active in the upstream market choose their royalty rates independently and non-cooperatively. However, it is quite common that firms coordinate their behavior through bilateral cross-licensing agreements.<sup>38</sup> Moreover, within standard setting contexts, it is becoming more common for at least some patent holders among a group of firms with complementary patents to coordinate their licensing through patent pools.<sup>39</sup> Such cross-licensing and pooling agreements are often

---

36. See Schmidt, *supra* note 31, at 13–14.

37. See Kim, *supra* note 30, at 245 (Theorem 3).

38. See Carl Shapiro, *Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting*, in 1 INNOVATION POLICY AND THE ECONOMY 119, 127 (Adam B. Jaffe, Josh Lerner & Scott Stern eds., 2000).

39. See, e.g., Sabra Chartrand, *The Federal Government Will Allow a Group of Companies to*

reciprocal.<sup>40</sup> Each firm agrees to a low royalty rate for its own patents as long as the other firms involved in the agreement also charge a low royalty rate.

This Part shows that if all firms are vertically integrated—as was largely the case within standards setting in past decades—it is possible to solve both the complements and the double mark-up problem as well as sustain the fully integrated monopoly outcome with a set of cross-licensing agreements. If the firms are symmetric, they will charge symmetric royalty rates.<sup>41</sup> Because everything is symmetric, the royalty payments cancel out and there are no net payments in equilibrium for other vertically integrated firms. However, if non-integrated downstream firms are present in the market (e.g., semiconductor chip fabricators that are not active in chip design), the problem of raising rivals' costs returns.

Some have argued that because the need to get cross-licenses from fellow patent holders does not restrict a non-integrated patent holder's behavior, it will turn such a patent holder into a patent troll charging unjustifiably high royalties.<sup>42</sup> Without the constraints that reciprocal cross-licensing imposes, it is claimed that non-integrated patent holders will hold-up vertically integrated firms and charge excessively high royalties.<sup>43</sup> However, this argument misses the mark. Pooling agreements may solve the complements and the double

---

*Unify Administration of 27 Patents*, N.Y. TIMES, June 30, 1997, at D8 (discussing the pool that is under development for radio frequency identification (RFID) standards); Philip Goldstein, *SISVEL Launches CDMA2000 Patent Pool*, FIERCEWIRELESS, June 11, 2009, <http://www.fiercewireless.com/press-releases/sisvel-launches-cdma2000-patent-pool>; Olga Kharif, *The RFID Patent Pool: Will It Make Waves?*, BUSINESSWEEK, Sept. 11, 2006, (The Tech Beat Blog), [http://www.businessweek.com/the\\_thread/techbeat/archives/2006/09/the\\_rfid\\_patent.html](http://www.businessweek.com/the_thread/techbeat/archives/2006/09/the_rfid_patent.html); Lynette Luna, *Patent Pool Can Work in the 4G World*, FIERCEWIRELESS, June 22, 2009, <http://www.fiercewireless.com/story/patent-pool-can-work-4g-world/2009-06-22>; Press Release, Via Licensing Corp., *Via Licensing Announces Progress Toward Establishing AVC License Terms* (Oct. 15, 2003), available at [http://www.vialicensing.com/news/details.cfm?VIANEWS\\_ID=309](http://www.vialicensing.com/news/details.cfm?VIANEWS_ID=309) (discussing many of the video and audio MPEG standards already involving patent pools).

40. See Shapiro, *supra* note 38.

41. In other words, if the firms hold similar patent portfolios (in terms of estimated value and composition), serve the same markets and similar customers, and generate roughly the same size in terms of revenues, then they will charge each other roughly similar licensing fees, so that a royalty free swap is likely.

42. See, e.g., Carl Shapiro, *Technology Cross-Licensing Practices: FTC v. Intel (1999)*, in THE ANTI-TRUST REVOLUTION: ECONOMICS, COMPETITION, AND POLICY 350, 357 (4th ed. 2004). On the other hand, some authors recognize that cross-licensing represents just one possible constraint, with other constraints remaining in place. See, e.g., Anne Layne-Farrar, Gerard Llobet & A. Jorge Padilla, *Preventing Patent Hold Up: An Economic Assessment of Ex Ante Licensing Negotiations in Standard Setting*, 37 AIPLA Q.J. 445 (2009).

43. See Shapiro, *supra* note 42.

mark-up problem even if some upstream firms are non-integrated.<sup>44</sup> In this case, firms are no longer symmetric. The non-integrated patent holders make their profits in the upstream market only and must insist on charging relatively higher royalties, while the integrated firms make their profits both upstream and downstream and tend to prefer lower royalties that shift profits downstream. This follows because with imperfect competition in the downstream market, vertically integrated firms earn additional margins downstream. For equally allocated profits amongst all integrated and non-integrated patent holders, the non-integrated patent holders must charge higher royalties than the integrated firms.

To make these arguments more precise, consider the case where there are  $N \geq 2$  symmetric and vertically integrated companies. Suppose that these firms negotiate a set of cross-licensing agreements according to which each firm charges each other firm the royalty  $r \geq 0$  for using its patent. Thus, the total royalty that each firm has to pay for each unit of the final good it sells in the downstream market is equal to the sum of rival upstream firm's royalty rates, which, given the assumed symmetry, amounts to multiplying the royalty rate,  $r$ , by the number of rivals:  $R = (N-1)r$ . Let  $R^*$  denote the total royalty payment that induces each firm to produce  $1/N$  of the output in the downstream market. If the firms choose  $r = r^* = R^* / (N-1)$ , then, given the set of cross-licensing agreements, each firm will produce  $1/N$  of the quantity downstream and each firm will earn  $1/N$  of the fully integrated profit. Thus, if all vertically integrated firms agree to charge the same royalty,  $r^*$ , the complements and the double mark-up problems disappear, just as if the firms had literally merged into a single monopoly firm. Furthermore, the cross-licensing agreements prevent the firms from raising their vertically integrated rivals' costs.

The above example mimics the outcome of a patent pool where a single licensing authority handles licensing of the full bundle of patents. Note that  $N$  vertically integrated firms forming such a pool is an agreement that fixes input prices. Antitrust authorities could regard this as an illegal cartel. In fact, the vertically integrated firms charge each other royalties that pass through to consumers and induce downstream divisions to charge the monopoly price. But, as previously discussed, this price is lower than if firms did not coordinate their behavior. The reason is the complements problem. If the  $N$  upstream monopolists do not coordinate their behavior, each will try to

---

44. Keeping with the chip example, these firms would be the design-only shops, having no manufacturing plants.

exploit their monopoly power. Total royalties will be even higher than the royalty a fully integrated monopolist charges.

To be sure, the outcome of a patent pool is a monopoly outcome. But, the monopoly is based on the IP rights of the patent holders, which are the rewards for their innovations. The patent pool ensures that none of the patent holders monopolizes their invention, which would impose negative externalities on the other monopolists; instead, patent holders coordinate their behavior to make benefit everyone.

Nevertheless, antitrust authorities tend to view patent pools that fix input prices with great suspicion, while they are generally much less skeptical about bilateral, reciprocal cross-licensing agreements.<sup>45</sup> However, in terms of economic effects, there is little difference between a patent pool and a bilateral cross-licensing agreement. For example, with a linear Cournot game, the same outcome can be sustained in equilibrium if only bilateral cross-licensing agreements are feasible, rather than a full-fledged pool.<sup>46</sup> To see this, consider two firms,  $i$  and  $j$ , who agree to a reciprocal royalty rate amongst themselves. This royalty rate has to be optimal given the royalty rates that all other pairs of firms have already agreed to. If all firms bilaterally agree to charge each other  $r^*$ , then firms jointly produce the monopoly quantity and earn the monopoly profit. Suppose now that firms  $i$  and  $j$  reduce the royalty rate that they charge each other. This reduces their marginal costs, so they gain an advantage in the downstream market. However, all the other firms now produce less downstream which implies less licensing income for firms  $i$  and  $j$  upstream. It turns out that these two effects just cancel out.<sup>47</sup> Thus, the patent pool outcome is also an equilibrium outcome if bilateral cross-licensing agreements are the only contract choice for firms.

If, in addition to the  $N$  vertically integrated firms, there are some non-integrated downstream firms, the double mark-up and the raising one's rivals' costs effects reappear. This scenario, where the more traditional large

---

45. U.S. DEP'T JUSTICE & FED. TRADE COMM'N, ANTITRUST ENFORCEMENT AND INTELLECTUAL PROPERTY RIGHTS: PROMOTING INNOVATION AND COMPETITION 8, 66 (2007), available at <http://www.ftc.gov/reports/innovation/P040101PromotingInnovationandCompetitionrpt0704.pdf>; Commission Notice, Guidelines on the Application of Article 81 of the EC Treaty to Technology Transfer Agreements, 2004 O.J. (C 101/2); see also Richard Schmalensee, *Standard-Setting, Innovation Specialists and Competition Policy*, 57 J. INDUS. ECON. 526, 542–47 (2009) (arguing that antitrust policy should not allow or encourage collective negotiation of patent royalty rates).

46. See *infra* Appendix.

47. This equilibrium need not be unique. However, even if there are multiple equilibria, the equilibrium that implements the monopoly outcome maximizes total surplus of all firms and is a natural focal point.

vertically integrated firms compete with both upstream and downstream specialists, is the best description of the composition of firms participating in most cooperative standard-setting bodies today. In this common case, under this Article's model, the vertically integrated firms will charge the low reciprocal royalty rate,  $r^*$ , only among themselves but a higher royalty rate,  $r > r^*$ , to their non-integrated downstream competitors. Part of the increased royalty rate is due to the downstream specialists' lack of patents to cross-license,<sup>48</sup> which eliminates a payment in kind. Some portion of the increase, however, is set in order to discriminate against the non-integrated downstream rivals.

What happens if, in addition to the  $N$  vertically integrated firms, there is also some number  $M > 0$  of non-integrated patent holders? Is it still possible to solve the complements and the double mark-up problem with a multi-lateral cross-licensing agreement? If the total royalty rate that each downstream division has to pay equals the sum set before,  $R^*$ , then each firm will again produce  $1/N$  of the monopoly output, just as before. However, the non-integrated patent holders have no interest in cross-licensing agreements because they do not require a license for the other patents. Furthermore, note that the integrated firms make part of their profits upstream and part of them downstream, while the non-integrated patent holders rely entirely on their royalty income. Finally, vertically integrated firms can also accept a portion of their upstream payment in the form of a cross-license, which tends to reduce the explicit royalty rate sought. For these reasons, non-integrated patent holders tend to ask for higher royalties than vertically integrated firms. In particular, if all firms have the same bargaining power and agree to share profits equally, the non-integrated patent holders must get a higher royalty rate than the vertically integrated firms.

If constraints forced firms to charge equal royalties independent of whether they are active in the downstream market or not, a conflict of interest would arise. Vertically integrated firms prefer a royalty rate,  $r$ , that is somewhat smaller than the monopoly royalty rate,  $r^* = R^*/(M+N-1)$ . This rate corresponds to the monopoly outcome and lowers total industry profits, but it does so at the expense of the non-integrated upstream firms, while the vertically integrated firms benefit from the lower downstream costs. In comparison, the non-integrated upstream firms prefer a royalty rate that is somewhat higher than the monopoly rate,  $r^* = R^*/(M+N-1)$ , because they want to shift profits from the downstream to the upstream market.

---

48. For an introduction to the interaction between contract terms, see generally THE ECONOMICS OF CONTRACTS: THEORIES AND APPLICATIONS (Eric Brousseau & Jean-Michel Glachant eds., 2002).

#### IV. NON-LINEAR AND DISCRIMINATORY ROYALTIES

Firms may also use non-linear royalty schemes, such as two-part tariffs, which are licenses with an upfront lump sum fee paid in conjunction with a linear running royalty rate based on sales spread over a specified time period. Such licenses are common in practice. The upfront fee reflects the agreed upon value of the patented technology, while the running royalty shares the risk over any remaining uncertainty regarding the commercial success of the product employing the patented technology. If written cross-licensing agreements are an option, the ability to employ similar non-linear payment schedules does not change the analysis.<sup>49</sup> The parties can set the linear portion of the royalties to implement the monopoly outcome and then use the fixed fees to redistribute profits.

Non-linear royalties are more interesting if firms have to set their rates independently and non-cooperatively, as is often the case in standard-setting bodies. The following analysis shows that the monopoly outcome can still be achieved if all firms are non-integrated. Suppose that each non-integrated upstream patent holder  $u$ ,  $u=1, \dots, N$ , makes a take-it-or-leave-it offer of a two-part tariff consisting of a fixed fee,  $R_u$ , and a linear royalty rate,  $r_u$ .<sup>50</sup> The first step of the argument is to show that all of the non-integrated downstream firms must make zero economic profits in equilibrium.<sup>51</sup> Suppose, to the contrary, that downstream firms make an economic profit that is strictly positive. Then an upstream patent holder could raise its fixed fee,  $R_u$ , in order to capture this profit for itself, without affecting the downstream firm's decision to continue participating in the market. As long as economic profits are non-negative, the downstream firm will stay in the market. In fact, if the sum of the fixed fees of the other upstream firms leaves any profit for the downstream firms, then each upstream firm has an incentive to further raise its fixed fee until all downstream firms make zero

---

49. Note that cross-licensing agreements are not always possible. For instance, if patented technology has many applications across multiple industries, it is possible that a particular licensee has no patents of interest to offer the licensor. Alternatively, if the licensee is a patent aggregator or an upstream R&D specialist, it may have no interest in any cross-licenses.

50. The combination of take-it-or-leave-it offers and two-part tariffs implies that upstream patent holders have all the bargaining power and that downstream firms do not get any rents. In a more complex bargaining procedure that gives some bargaining power to downstream firms, this extreme result disappears. However, all the qualitative results, described *infra*, still hold true.

51. Recall that accounting profits are very different from economic profits, which consider opportunity costs in addition to costs of production. Zero economic profits simply indicate a competitive market, where downstream firms do not earn anything above the competitive return for their production investments.

economic profits. Thus, in equilibrium, upstream firms will extract the entire surplus from downstream firms.<sup>52</sup> If a downstream firm controls an input factor that increases the value of the downstream good, such as a brand name, its bargaining power would be stronger and the allocation of profits would surely differ.

The second step of the argument is to show that linear royalties will be set efficiently so as to maximize the downstream profits that can then be captured by the upstream firms. To illustrate, suppose that the sum of the linear royalties is higher than the royalty rate that implements the monopoly price. This circumstance has been described as “royalty stacking” in the literature.<sup>53</sup> If rates were to stack to a level higher than the monopolistic rate, the upstream patent holder,  $u$ , could reduce its royalty rate,  $r_u$ , which would increase total profits because rates higher than a monopolist’s yield strictly smaller profits for the patent holder. The patent holder can simultaneously increase its fixed fee,  $R_u$ , as a means of capturing the entire increase in total profits for itself, leaving the downstream firms with zero economic profits, described *supra*. Thus, if the sum of the linear royalties does not lead to the monopoly outcome, each upstream firm has an incentive to change its royalty structure in order to obtain the resulting increase in profits through an increase of the fixed fee. There is a symmetric pure strategy equilibrium in which all patent holders charge the same linear royalties and fixed fees such that the sum of the linear royalties implements the monopoly outcome and the sum of the fixed fees captures all the downstream profits. Thus, with the flexibility of two-part tariffs, both the double mark-up and the complements problem, including royalty stacking, disappear. To summarize, if all firms are non-integrated and two-part tariffs can be used, the same outcome will manifest with a patent pool even if no written cross-licenses are utilized.

However, this result holds only if all patent holders are non-integrated—an unlikely scenario. Today, most complex industries, where the complements and double mark-up problems are likely to emerge, are characterized by a mixture of vertically integrated and non-integrated firms. In fact, under this Article’s model, if there are at least two vertically integrated firms along with the other non-integrated firms, then there is no symmetric, pure strategy equilibrium where all patent holder royalty rates are the same—some asymmetry must persist. For example, suppose that there is an equilibrium in which all firms charge symmetric fees,  $r$ , and royalties,  $R$ .

---

52. This extreme outcome is due to the assumption that the upstream firms have all the bargaining power and can make take-it-or-leave-it offers. See Schmidt, *supra* note 31, at 10.

53. See Lemley & Shapiro, *supra* note 5, at 1993.

First of all, it must be the case that the fixed fees extract all the profits from the downstream market, otherwise each firm would have an incentive to further raise its fixed fee. But if fees and royalties extract all of the downstream economic profits, then all vertically integrated firms are indifferent as to whether or not to produce downstream. Suppose that vertically integrated firm,  $i$ , further increases its linear royalty rate or its fixed fee. The increase does not affect  $i$ 's own costs, but it does raise the costs of its downstream rivals. As a result, some other firms will now decide not to buy the licenses at all and to produce a quantity of zero—all to the benefit of firm  $i$ . Therefore, firm  $i$  has an incentive to deviate from the proposed equilibrium candidate. Thus, the “raising rivals’ costs” effect implies that a symmetric, pure strategy equilibrium does not exist when both integrated and non-integrated firms are present.

As seen *supra*, when all downstream firms are vertically integrated, firms can deal with the “raising rivals’ costs” effect by writing a cross-licensing agreement. However, they nonetheless have an incentive to discriminate downstream against non-integrated firms in order to jointly monopolize the market. Non-integrated upstream firms do not have this incentive. To the contrary, they benefit from more downstream competition because it increases downstream quantities sold and therefore increases their royalty income. This result leads us to reject the prevalent definition of a patent troll as any non-practicing or non-innovating entity. Indeed, NPEs are the least likely to exhibit troll behaviors. Instead, a better gauge is the presence of special conditions for a patent hold-up and the exploitation of irreversible investments, regardless of the business model of the patent holder.

## V. CONCLUSIONS

This analysis shows that the presumption that NPEs always charge higher royalties than vertically integrated companies is not warranted. It is true that if firms set linear royalties non-cooperatively, a vertically integrated firm will internalize the vertical double mark-up problem which tends to reduce royalties. However, a vertically integrated firm also has an incentive to raise its royalties in order to raise its rivals’ costs and to restrict entry in the downstream market, especially among non-integrated downstream firms. The overall effect on rates is ambiguous. In fact, under some circumstances, a vertically integrated firm may charge *higher* royalties than its non-integrated counterpart.<sup>54</sup>

---

54. In particular, when the downstream market is comprised of both vertically integrated firms and downstream manufacturing specialists, the integrated firms have strong

If firms can coordinate their royalties through patent pools or cross-licensing agreements, they can solve the complements and the double-mark-up problem independently of whether or not they are vertically integrated. However, if there are non-integrated downstream firms, vertically integrated firms will nonetheless want to increase royalties in order to discriminate against them, while non-integrated patent holders will want to decrease royalties to benefit from increased competition and entry downstream.

When both non-integrated and integrated firms hold patents and are able to coordinate their royalties, a conflict of interests arises. Non-integrated firms have to make their profits upstream, while the vertically integrated firms make some part, or even the bulk of their profits in the downstream market. Furthermore, vertically integrated firms may reduce their royalty rates in exchange for a cross-license payment in kind—an option not available to upstream firms. Therefore, if firms want to split *profits* equally, or if they want to achieve the same aggregate licensing payments, then non-integrated firms must receive higher explicit *royalties* than the integrated firms.

This Article's analysis suggests that there is no justification for the presumption that non-integrated patent holders always charge higher royalties than vertically integrated companies. Moreover, even when non-integrated patent holders charge "higher" royalties than their vertically integrated counterparts, it does not imply that the rates are "excessive" or that the firm is exhibiting troll-like behavior. Rather, non-integrated patent holders naturally require higher royalty earnings because they earn no profits downstream and receive no payments in kind in the form of cross-licenses.

The findings presented in this Article are consistent with the larger patent literature. In a recent empirical paper, Allison, Lemley, and Walker analyzed a data set of litigated patents.<sup>55</sup> Almost by definition, litigated patents are patents that are highly valuable. A large fraction of these patents are owned by NPEs, held mainly by invention specialists and by patent holding companies. This shows that NPEs play a major role in the modern patent system. Allison et al. compare the characteristics of the most litigated patents, specifically patents that have been litigated at least eight times, to patents that have been litigated only once. They find that the most litigated patents are much more likely to be owned by NPEs.<sup>56</sup> This is consistent with this

---

incentives to raise their non-integrated rivals' costs through high royalty rates. This strategy softens downstream competition and enables the integrated firms to earn higher profits.

55. John R. Allison, Mark A. Lemley & Joshua Walker, *Extreme Value or Trolls on Top? The Characteristics of the Most-Litigated Patents*, 158 U. PA. L. REV. 1 (2009); see also Shrestha, *supra* note 14 (similar findings).

56. *Id.* at 23–26.

Article's analysis that suggests that there is more potential for conflict if NPEs are involved.<sup>57</sup> Furthermore, in the overwhelming majority of cases studied by Allison et al., the patent holders were invention specialists.<sup>58</sup> Only seven percent of all lawsuits involved patent holders that did not invent the patent themselves, but rather, acquired it.<sup>59</sup> This suggests that court-based discrimination against all NPEs could have severe adverse effects on the investment incentives of innovation specialists who play an important role in advanced market economies.

In conclusion, remedy rules, such as entitlement to injunctive relief, should not depend on the plaintiff's business model status. Whether a company charges "excessive" royalties depends on the potential scope for a hold-up, either because of sunk investments or weak patents. These factors are orthogonal to whether patent holders are vertically integrated or not.

---

57. *See supra* Part II.

58. Allison et al., *supra* note 55, at 23–26.

59. *Id.* at 32.

### APPENDIX: BILATERAL CROSS-LICENSING

Consider  $N$  vertically integrated firms producing with identical constant marginal production cost  $k$ . Each firm owns one essential patent. Firms compete in quantities in the downstream market and face a linear inverse demand function,  $P(Q)=a-bQ$ . Each pair  $(i,j)$ ,  $i, j \in \{1, \dots, N\}$ , of firms agrees to symmetric cross-licensing at rate  $r_{ij}=r_{ji}$ , where  $r_{ij}$  is the linear royalty charged by firm  $i$  to firm  $j$ . Thus, the total licensing costs of firm  $i$  are given by  $R_i \sum_{j \neq i} r_{ji}$ .

We look for a subgame perfect equilibrium with the two following properties:

- Given the licensing cost  $R_i$  the firms play a Cournot-Nash equilibrium at the second stage of the linear Cournot game.
- There does not exist a pair  $(i, j)$  of firms that could increase its joint profits by agreeing to a different cross-licensing rate.

#### CLAIM

There is a symmetric subgame perfect equilibrium in the cross-licensing game in which all firms agree to charge the same cross-licensing royalty rate  $r_{ij} = r = \frac{a-k}{2N}$ . This licensing rate implements the monopoly outcome.

#### PROOF:

Suppose that each pair of firms agreed to the cross-licensing rate  $r = \frac{a-k}{2N}$ . Then each firm has marginal cost  $c = k + \frac{(N-1)(a-k)}{2N}$ . It is straightforward to compute the symmetric Cournot-Nash equilibrium at the second stage of the game. In equilibrium, each firm produces  $q = \frac{a-k}{2bN}$ , so the total quantity supplied is  $Q = \frac{a-k}{2b}$  which is equal to the monopoly quantity. The resulting price is the monopoly price  $p = \frac{a+k}{2}$ , and each firm makes  $\Pi_i = \frac{(a-k)^2}{4bN}$  which is  $1/N$  of the monopoly profit.

Suppose now that two firms 1 and 2 consider a deviation and charge each other  $r_{12} = r_{21} = r - d$ , (where  $d$  may be positive or negative). This deviation changes the marginal costs of these two firms to  $\underline{c} = k + (N-2)r + r - d = (N-1)r - d$ , while the marginal costs of all other firms remain unchanged. We now have to solve for the Cournot-Nash

equilibrium in the downstream market with asymmetric cost functions. Note that downstream profit functions are given by

$$\Pi_i = \left( a - b \sum_{j=1}^N q_j - k - (N-1)r + d \right) \cdot q_i \text{ if } i \in \{1,2\}$$

$$\Pi_i = \left( a - b \sum_{j=1}^N q_j - k - (N-1)r \right) \cdot q_i \text{ if } i \in \{3, \dots, N\}$$

The first order conditions for profit maximization are given by

$$\frac{\partial \Pi_i}{\partial q_i} = a - 2bq_i - b \sum_{j \neq i} q_j - k - (N-1)r + d = 0 \text{ if } i \in \{1,2\}$$

$$\frac{\partial \Pi_i}{\partial q_i} = a - 2bq_i - b \sum_{j \neq i} q_j - k - (N-1)r = 0 \text{ if } i \in \{3, \dots, N\}$$

Symmetry of firms 1 and 2 and firms 3, ..., N requires  $q_i = q_2 = \bar{q}$  and  $q_3 = \dots = q_N = \underline{q}$ . Solving for  $\bar{q}$ ,  $\underline{q}$ , and  $P$  and plugging in  $r = \frac{a-k}{2N}$  yields:

$$\bar{q} = \frac{a - k - (N-1)(r - d)}{b(N+1)} = \frac{a - k}{2Nb} + \frac{N-1}{N+1} d$$

$$\underline{q} = \frac{a - k - (N-1)r - 2d}{b(N+1)} = \frac{a - k}{2Nb} - \frac{2d}{N+1}$$

$$P = \frac{a + Nk + rN(N-1)}{N+1} = \frac{a + k}{2} - \frac{2d}{N+1}$$

Thus, the profit of firm  $i \in \{1,2\}$  is given by

$$\Pi_i = \underbrace{(P - k - (N-1)r + d) \cdot \bar{q}}_{\text{downstream profit}} + \underbrace{(r - d)\bar{q} + (N-2)r\underline{q}}_{\text{upstream profit}}$$

Substituting  $P$ ,  $r$ ,  $\bar{q}$ , and  $\underline{q}$  yields

$$\Pi_i = \frac{-4akN + a^2N^2 + k^2N^2 + 2a^2N + k^2 - 2ak - 2aN^2k - 8d^2N(N-1)}{4bN(N+1)^2}$$

Differentiating with respect to  $d$  we get:

$$\frac{\partial \Pi_i}{\partial d} = \frac{-16dN(N-1)}{4bN(N-1)} = -\frac{4d}{b}$$

Note that this profit function is globally concave and maximized at  $d = 0$ . Thus, no pair of firms has an incentive to deviate and to change its bilateral royalty rate. *Q.E.D.*

**1144**

**BERKELEY TECHNOLOGY LAW JOURNAL [Vol. 25:1121**