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

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

Berkeley Technology Law Journal
University of California
School of Law
3 Boalt Hall
Berkeley, California 94720-7200
btlj@law.berkeley.edu
http://www.btlj.org

BERKELEY TECHNOLOGY LAW JOURNAL

VOLUME 28 NUMBER 2 FALL 2013

TABLE OF CONTENTS

ARTICLES

SUBSCRIBER INFORMATION

The *Berkeley Technology Law Journal* (ISSN1086-3818), a continuation of the *High Technology Law Journal* effective Volume 11, is edited by the students of the University of California, Berkeley, School of Law (Boalt Hall) and is published in print three times each year (March, September, December), with a fourth issue published online only (July), by the Regents of the University of California, Berkeley. Periodicals Postage Rate Paid at Berkeley, CA 94704-9998, and at additional mailing offices. POSTMASTER: Send address changes to Journal Publications, University of California, Berkeley Law—Library, LL123 Boalt Hall—South Addition, Berkeley, CA 94720-7210.

Correspondence. Address all correspondence regarding subscriptions, address changes, claims for non-receipt, single copies, advertising, and permission to reprint to Journal Publications, University of California, Berkeley Law—Library, LL123 Boalt Hall—South Addition, Berkeley, CA 94705-7210; (510) 643-6600; JournalPublications@law.berkeley.edu. *Authors*: see section titled Information for Authors.

Subscriptions. Annual subscriptions are $65.00 for individuals and $85.00 for organizations. Single issues are $30.00. 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 ($30.00) 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 (16th ed. 2010) and to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (Columbia Law Review Ass'n et al. eds., 19th ed. 2010). Please cite this issue of the *Berkeley Technology Law Journal* as 28 BERKELEY TECH. L.J. ___ (2013).

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*, and the Bolt, a collection of short comments and updates about new developments in law and technology written by members of BTLJ.

INFORMATION FOR AUTHORS

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

Format. Submissions are accepted in electronic format through the Express*O* online submission system. Authors should include a curriculum vitae and resume when submitting articles, including his or her full name, credentials, degrees earned, academic or professional affiliations, and citations to all previously published legal articles. The Express*O* submission website can be found at http://law.bepress.com/expresso.

Citations. All citations should conform to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (Columbia Law Review Ass'n et al. eds., 19th ed. 2010).

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

DONORS

The *Berkeley Technology Law Journal* and the Berkeley Center for Law & Technology acknowledge the following generous donors to Berkeley Law's Law and Technology Program:

Partners

COOLEY LLP

COVINGTON & BURLING LLP

FENWICK & WEST LLP

ORRICK, HERRINGTON & SUTCLIFFE LLP

Benefactors

FISH & RICHARDSON P.C.

KASOWITZ BENSON TORRES & FRIEDMAN LLP

KIRKLAND & ELLIS LLP

LATHAM & WATKINS LLP

MCDERMOTT WILL & EMERY

MORRISON & FOERSTER LLP

SKADDEN, ARPS, SLATE, MEAGHER & FLOM LLP & AFFILIATES

WEIL, GOTSHAL & MANGES LLP

WHITE & CASE LLP

WILMER CUTLER PICKERING HALE AND DORR LLP

WILSON SONSINI GOODRICH & ROSATI

WINSTON & STRAWN LLP

Members

ALSTON & BIRD LLP

BAKER BOTTS LLP

BAKER & MCKENZIE LLP

BINGHAM MCCUTCHEN LLP

DURIE TANGRI LLP

FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, LLP

GTC LAW GROUP & AFFILIATES

GUNDERSON DETTMER STOUGH VILLENEUVE FRANKLIN & HACHIGIAN, LLP

HAYNES AND BOONE, LLP

HICKMAN PALERMO TRUONG BECKER BINGHAM WONG, LLP

KEKER & VAN NEST LLP

HOGAN LOVELLS LLP

IRELL & MANELLA LLP

KILPATRICK TOWNSEND & STOCKTON LLP

KNOBBE MARTENS OLSON & BEAR LLP

MORGAN, LEWIS & BOCKIUS LLP

MUNGER, TOLLES & OLSON LLP

ROPES & GRAY LLP

SIDLEY AUSTIN LLP

SIMPSON THACHER & BARTLETT LLP

TURNER BOYD LLP

VAN PELT, YI & JAMES LLP

WEAVER AUSTIN VILLENEUVE & SAMPSON, LLP

BOARD OF EDITORS 2012–2013

Executive Committee

Editor-in-Chief
DAVID ROSEN

Managing Editor
LAUREN ESCHER

Senior Articles Editors
YVONNE LEE
HANNAH MINKEVITCH

Senior Executive Editor
JOE SEXTON

Senior Annual Review Editors
WINNIE HUNG
JANE LEVICH

Senior Scholarship Editor
COURTNEY BOWMAN

Editorial Board

Submissions Editors
JULIA KOLIBACHUK
PENNI TAKADE

Production Editor
JARAD BROWN

Bluebook Editors
GABRIEL MILITELLO
YUE WANG
ZACH WOOD

Annual Review Editors
ANGEL DIAZ
LALI MADDURI

Notes & Comments Editors
ELISE EDLIN
BRITT HARWOOD

Symposium Editors
KAREN KOPEL
KILEY WONG

Publishing Editor
DINA ROUMIANTSEVA

Web Content Editor
CHRIS CIVIL

Web & Technology Editors
EDUARD MELESHINSKY
ANDREA YANKOVSKY

External Relations Editor
GAVIN LIU

Member Relations Editor
CASEY HULTIN

Articles Editors

BRADY BLASCO
JULIE BYREN
EMILY CHEN
ROSS COHEN
CHRISTINA FARMER
GRANT GARBER
ROBIN KUNTZ
NICHOLAS LAMPROS
MARK LANGER
MARIENNA MURCH
JUSTIN ORR
SARAH ORRICK
LAUREN SMITH
ETHAN WEINER

MEMBERSHIP

Vol. 28 No. 2

Associate Editors

LAUREN BLAKELY
DALLAS BULLARD
MICHAEL BURSHTEYN
DOMINIQUE CAAMANO
LAURE CARDINET
SIENA CARUSO
ANTHONY CHANRASMI
RAFAEL CRUZ
LINDSAY DOCTO
MATTHEW DONOHUE
WILSON DUNLAVEY
JONATHAN FRANCIS
NYDIA GALARZA
S. ZUBIN GAUTAM
MARK GRAY
JEFFREY HIRSCHEY
SAMANTHA HUANG
LEON KOTLYAR
BAILEY LANGNER
ALEX S. LI
DEBBIE LI
GI-KUEN LI
EMILY LY
DARIUS MASSOUDI
STEVEN S. MCCARTY-SNEAD
JUSTIN MCGUIRK
JOSEPH MORNIN
PURBA MUKERJEE
ARAM ROSTOMYAN
LEXI RUBOW
ALLISON SCHMITT
ROSALIND SCHONWALD
DONGBIAO SHEN
DANIELA SPENCER
JENNA STOKES
CHRISTOPHER SUNG
STEVEN SWANSON
AMY TU
VAIBHAV VENKATESH
BONNIE WATSON
ANDREW YOCOPIS
MARK ZAMBARDA
LUIS ZAMBRANO
OGNJEN ZIVOJNOVIC

Members

ASHWIN ANAND
CRAIG ARMSTRONG
NATASHA ARORA
IRENA CHEN
VICTOR CHIU
ALEX COLEY
NATHALIE DAVID
BEVAN DOWD
MISA EIRTIZ
MICHAEL GAFFNEY
DANNY GOLDSTEIN
LIBBY HADZIMA
JULIA HOFFMAN
DAMION JURRENS
ROHAN KADAM
DROFN KAERNESTED
KAMOLA KOBILDJANOVA
ANAGHA KRISHNAN
SAHAR MAALI
KASSIE MALDONADO
ANGELIKA MAREK
NATALIA MIKOLAJCZYK
RUSS NELDAM
WADE ORBELIAN
STUYVIE PYNE
MEREDITH REED
NAZLI SAKA
ANDREW SIMON
MARTIN STIERLE
MICHAEL LIU SU
BENJAMIN TAKEMOTO
JONATHAN TAMIMI
EMMANUELLE TANG
MATTHEW WASSERMAN
KEVIN XU
ZELIN YANG
MARGARET YI

BTLJ ADVISORY BOARD

BERKELEY CENTER FOR LAW & TECHNOLOGY 2012–2013

Executive Director

ROBERT BARR

Faculty Directors

KENNETH BAMBERGER
PETER MENELL
ROBERT MERGES
DEIRDRE MULLIGAN
PAMELA SAMUELSON
PAUL SCHWARTZ
SUZANNE SCOTCHMER
MOLLY VAN HOUWELING

Associate Director

LOUISE LEE

Affiliated Faculty and Scholars

AARON EDLIN
JOSEPH FARRELL
RICHARD GILBERT
BRONWYN HALL
THOMAS JORDE
MICHAEL KATZ
DAVID MOWERY
DAVID NIMMER
DANIEL RUBINFELD
ANNALEE SAXENIAN
JASON SCHULTZ
HOWARD SHELANSKI
CARL SHAPIRO
MARJORIE SHULTZ
LON SOBEL
TALHA SYED
DAVID TEECE
JENNIFER M. URBAN
HAL R. VARIAN
DAVID WINICKOFF

A SIMPLE APPROACH TO SETTING REASONABLE ROYALTIES FOR STANDARD-ESSENTIAL PATENTS

Mark A. Lemley[†] *& Carl Shapiro*[††]

ABSTRACT

A Standard Setting Organization ("SSO") typically requires its members to license any standard-essential patents on fair, reasonable, and non-discriminatory ("FRAND") terms. Unfortunately, numerous high-stakes disputes have recently broken out over just what these "FRAND commitments" mean and how and where to enforce them. We propose a simple, practical set of rules regarding patents that SSOs can adopt to achieve the goals of FRAND commitments far more efficiently with far less litigation. Under our proposed approach, if a standard-essential patent owner and an implementer of the standard cannot agree on licensing terms, the standard-essential patent owner is obligated to enter into binding baseball-style (or "final offer") arbitration with any willing licensee to determine the royalty rate. This obligation may be conditioned on the implementer making a reciprocal FRAND commitment for any standard-essential patents it owns that read on the same standard. If the implementer is unwilling to enter into binding arbitration, the standard-essential patent owner's FRAND commitment not to go to court to enforce its standard-essential patents against that party is discharged. We explain how our proposed FRAND regime would work in practice. Many of the disputes currently arising around FRAND commitments become moot under our approach.

---

† Lemley is the William H. Neukom Professor at Stanford Law School and a partner at Durie Tangri LLP. Lemley has represented Google in matters related to the subject of this Article, but Google has provided no financial support for this project and the views offered here have neither been reviewed nor approved by Google.

†† Shapiro is the Transamerica Professor of Business Strategy at the Haas School of Business, University of California at Berkeley and a Senior Consultant at Charles River Associates.

We speak only for ourselves. We thank Robert Barr, Jorge Contreras, Thomas Cotter, Joseph Farrell, Richard Gilbert, Rose Hagan, Robert Harris, Brian Love, Gil Ohana, Fiona Scott-Morton, Greg Sidak, Jeffrey Wilder, and participants at a workshop at the Federal Trade Commission for helpful comments on an earlier draft.

TABLE OF CONTENTS

Voluntary standard-setting organizations enable industry participants to meet and establish technical standards. These standards can greatly facilitate competition and innovation.[1] For example, the Institute of Electrical and Electronics Engineers ("IEEE") has established Ethernet and Wi-Fi standards.[2] However, complications arise when implementing a standard requires practicing certain patents. Patents covering technology necessary to comply with a standard are "standard-essential patents."

The vast majority of standard-setting organizations ("SSOs") require their members to commit to license any standard-essential patents on fair,

1. *See* U.S. DEP'T OF JUSTICE & FED. TRADE COMM'N, ANTITRUST ENFORCEMENT AND INTELLECTUAL PROPERTY RIGHTS: PROMOTING INNOVATION AND COMPETITION 33–56 (2007), *available at* http://www.justice.gov/atr/public/hearings/ip/222655.pdf.

2. For more information about the 802.3 series of Ethernet standards, see *IEEE 802.3: Ethernet*, IEEE STANDARDS ASS'N, http://standards.ieee.org/about/get/802/802.3.html (last visited Mar. 30, 2013). For information about the 802.11 series of WiFi standards, see *IEEE 802.11: Wireless LANs*, IEEE STANDARDS ASS'N, http://standards.ieee.org/about/get/802/802.11.html (last visited Mar. 30, 2013).

reasonable, and non-discriminatory ("FRAND") terms.[3] These FRAND commitments serve two primary goals: (1) to promote the standard by assuring companies implementing the standard that they will not be blocked from bringing their products to market or held up so long as they are willing to pay reasonable royalties for any standard-essential patents, and (2) to provide reasonable rewards to those who have invested in research and development to develop the technology used by the standard.[4]

FRAND commitments have taken on increasing importance in recent years as courts have been called upon to decide what they mean,[5] and as the Federal Trade Commission has brought antitrust actions to enforce those commitments.[6] This litigation is largely a function of ambiguities and

3. A recent survey of SSO patent policies can be found in RUDI BEKKERS & ANDREW UPDEGROVE, A STUDY OF IPR POLICIES AND PRACTICES OF A REPRESENTATIVE GROUP OF STANDARDS SETTING ORGANIZATIONS WORLDWIDE (2012), http://sites.nationalacademies.org/PGA/step/IPManagement/PGA_072197, which is part of a project by the National Academies of Science. *See Patent Challenges for Standard-Setting in the Global Economy: Lessons from Information and Communications Technology*, NATIONAL ACADEMIES, http://nap.edu/catalog.php?record_id=18510 (last visited Nov. 3, 2013). Earlier surveys of SSO patent policies can be found in Mark A. Lemley, *Intellectual Property Rights and Standard-Setting Organizations*, 90 CALIF. L. REV. 1889 (2002) [hereinafter Lemley, *SSOs*] and Benjamin Chiao, Josh Lerner & Jean Tirole, *The Rules of Standard-Setting Organizations: An Empirical Analysis*, 38 RAND J. ECON. 905 (2007). For a recent survey of licensing *disclosure* policies, see Jorge Contreras, *Technical Standards and Ex Ante Disclosure: Results and Analysis of an Empirical Study*, 53 JURIMETRICS 163 (2013). We do not discuss in this paper SSOs that merely require disclosure of patents but do not require licensing of those patents. That was more common at the time of Lemley's 2002 paper, but disclosure-only policies have fallen into disfavor, in part because of abuse of those policies by companies like Rambus. *See id.* at 166 n.9, 180; Lemley, *SSOs*, *supra*.

4. *See, e.g.*, U.S. DEP'T OF JUSTICE & U.S. PATENT & TRADEMARK OFFICE, POLICY STATEMENT ON REMEDIES FOR STANDARDS-ESSENTIAL PATENTS SUBJECT TO VOLUNTARY F/RAND COMMITMENTS 5 (2013), *available at* http://www.justice.gov/atr/public/guidelines/290994.pdf.

5. *See, e.g.*, Microsoft Corp. v. Motorola, Inc., 696 F.3d 872 (9th Cir. 2012); Broadcom Corp. v. Qualcomm, Inc., 501 F.3d 297 (3d Cir. 2007); Apple, Inc. v. Motorola Mobility, No. 11-CV-178-BBC, 2012 WL 5416941 (W.D. Wisc. Oct. 29, 2012); Apple, Inc. v. Samsung Elecs. Co., No. 11-CV-01846-LHK, 2012 WL 2571719 (N.D. Cal. June 30, 2012); Apple, Inc. v. Motorola Inc., No. 11-CV-08540, 2012 WL 1959560 (N.D. Ill. May 22, 2012); Microsoft Corp. v. Motorola, Inc., 864 F. Supp. 2d 1023 (W.D. Wash. 2012). For a table listing all FRAND litigation, see Jorge L. Contreras, Fixing FRAND: A Pseudo-Pool Approach to Standards-Based Patent Licensing, app. 1 (Mar. 13, 2013) (unpublished manuscript), http://works.bepress.com/jorge_contreras/6/.

6. *See, e.g.*, Complaint, Motorola Mobility LLC, No. 121-0120 (F.T.C. Jan. 3, 2013), *available at* http://www.ftc.gov/os/caselist/1210120/130103googlemotorolacmpt.pdf; Statement of the Federal Trade Commission, Robert Bosch GmbH, No. 121-0081 (F.T.C. 2013), *available at* http://www.ftc.gov/os/caselist/1210081/121126boschcommissionstatement.pdf; Complaint, Negotiated Data Solutions LLC (*N-Data*), No. C-4234, 2008 WL 4407246 (F.T.C. Sept. 23, 2008); Dell Computers, Inc., 121 F.T.C. 616 (1996).

omissions in the FRAND system used by most SSOs. The effectiveness of the FRAND commitment has been undermined by these ambiguities and omissions, especially for standards in the information technology sector.

In this Article, we propose best practices for SSOs in implementing the FRAND commitment. Under our proposal, owners of standard-essential patents agree to license their portfolio of standard-essential patents on FRAND terms, with the portfolio royalty rate determined through binding arbitration if necessary. SSOs adopting our proposal will more effectively achieve the twin goals of FRAND commitments noted, *supra*: freedom to implement the standard along with reasonable returns to inventors who contribute patented technology to the standard. Our approach is flexible: SSOs can adopt our basic structure with variations designed to fit their individual needs and circumstances. Indeed, patentees can choose to abide by the arbitration commitment even if the SSO does not compel it.

Substantively, our proposal is designed to steer bilateral, ex post negotiations towards royalty rates that reflect the outcome of ex ante technology competition.[7] Our proposal achieves this (a) by protecting implementers from the threat that they will face an injunction or exclusion order, but (b) only for implementers who agree to pay a reasonable royalty rate, as determined through binding arbitration if necessary. Procedurally, our proposals are designed to be as simple, unambiguous, and transparent as possible.

The key to our approach is to bind patentees to engage in arbitration over the reasonable royalty with any willing licensee, rather than litigating the patents in court. We favor baseball-style arbitration, under which each party submits its final offer to the arbitrator, who then must pick one of those two offers. Under our proposal, a patentee who makes a FRAND commitment promises to forego court enforcement of its standard-essential patents in favor of arbitration over the royalty rate with any implementer of the standard willing to engage in such arbitration.[8]

If SSOs follow our proposal for resolving FRAND disputes, many of the issues that have occupied the courts will become moot, at least for future

7. We refer to negotiations that take place before the standard is adopted as *ex ante negotiations*. Negotiations that take place after the standard has been implemented are referred to as *ex post negotiations*. Actual ex ante negotiations are often difficult or infeasible, in part because not all of the parties with an interest in deploying a standard belong to the SSO. Nothing in our proposal discourages or prevents negotiations at an early stage.

8. For other proposals to have a neutral fact-finder determine a FRAND royalty rate, see Kai-Uwe Kuhn, Fiona Scott Morton & Howard Shelanski, *Standard Setting Organizations Can Help Solve the Standard Essential Patents Licensing Problem*, 3 COMPETITION POL'Y INT'L ANTITRUST CHRON. (SPECIAL ISSUE) 1, 4–5 (Mar. 2013).

standards. Under a FRAND regime of the sort we propose, there is no need for the SSO to be substantively involved in deciding what is reasonable, no need to decide whether one party or another breached a contract so long as they participated in the arbitration, no need to decide whether a patent holder's offer was actually a FRAND offer, no need to worry about which jurisdiction is litigating the issue, and no need for antitrust law to intervene so long as the parties are abiding by their FRAND commitments.

We recognize that the vast majority of patent litigation does not involve standard-essential patents.[9] SSO FRAND policies must be understood in the context of the broader operation of the patent system. Making FRAND commitments clear and enforceable may reduce the leverage of standard-essential patent owners in negotiations with implementers who own patents that are not standard-essential. Our hope is that the U.S. Patent and Trademark Office will continue to improve patent quality and reduce patent pendency, and the courts will continue to make improvements in the general treatment of patent remedies, so that owners of FRAND-encumbered patents are not unfairly disadvantaged in comparison with owners of patents that are not standard-essential. We do not regard broader problems with the patent system as a reason to preserve a poorly functioning set of SSO patent rules.

In Part I, we address the issues that must be resolved to give effect to a FRAND commitment, and we propose a set of best practices for resolving those issues.[10] In Part II, we consider a number of questions that frequently arise in standard-essential patent FRAND litigation that by and large become moot under our proposed FRAND regime.

I. PROPOSED APPROACH TO FRAND COMMITMENTS

We describe and explain here our proposed approach to FRAND commitments. SSO best practice begins with an explicit articulation in the SSO's intellectual property ("IP") rules of the twin goals of the FRAND regime: freedom to implement the standard along with reasonable returns to inventors who contribute patented technology to the standard.[11] Those

9. Timothy Simcoe et al., *Competing on Standards? Entrepreneurship, Intellectual Property, and Platform Technologies*, 18 J. ECON. & MANAGEMENT STRATEGY 775, 787 (identifying a total of 949 standard-essential patents litigated over a period of many years; for comparison, there were over 5000 patent suits filed in 2012 alone).

10. For an argument encouraging the adoption of best practices by SSOs, see Michael A. Lindsay, *Safeguarding the Standard: Standards Organizations, Patent Hold-up, and Other Forms of Capture*, 57 ANTITRUST BULL. 17, 30–31 (2012).

11. Looking at a large number of SSOs, Bekkers and Updegrove noted:

reasonable returns reflect the ex ante value of the patented technology, not the additional ex post value resulting from the standardization itself. As one of us wrote fifteen years ago, "Reasonable *should* mean the royalties that the patent holder could obtain in open, up-front competition with other technologies, not the royalties that the patent holder can extract once other participants are effectively locked in to use technology covered by the patent."[12] As we discuss below, this interpretation of "reasonable royalties" comports with patent law and is now widely accepted by economists and policy makers.

A. BASIC STRUCTURE OF THE FRAND REGIME

Patent owners have the right to exclude others from practicing their technology (subject to the limits of equity, which will not always grant injunctions),[13] or to trade that right for valuable consideration. SSO FRAND rules are designed to get parties to pre-commit to license their essential patents on reasonable and non-discriminatory terms, preventing later efforts to disrupt the technology or hold up users of the standard for supra-competitive royalties.[14] Put another way, the FRAND commitment is at its base an agreement not to exercise the full scope of the patentee's rights in exchange for having its technology adopted as an industry standard, likely resulting in increased licensing opportunities.

For that commitment to be effective, it must be a legally binding obligation. A "FRAND commitment" that is nothing more than a promise to

> [N]one of the policies attempts to even define what "fair" or "reasonable" fees are intended to mean in context. Nor do they state that at minimum, such fees must bear a reasonable relationship to the economic value of the IPR, despite the fact that this benchmark is stated explicitly by the FTC in its report on evolving [*sic*] IP marketplace, as well as in the European Commission's relevant Guidelines on horizontal cooperation agreements.

Bekkers & Updegrove, *supra* note 3, at 102–03 (citations omitted).

12. CARL SHAPIRO & HAL R. VARIAN, INFORMATION RULES: A STRATEGIC GUIDE TO THE NETWORK ECONOMY 241 (1999).

13. *See* eBay Inc. v. MercExchange LLC, 547 U.S. 388, 391–97 (2006).

14. On the holdup problem, see generally Mark A. Lemley & Carl Shapiro, *Patent Holdup and Royalty Stacking*, 85 TEX. L. REV. 1991 (2007); Mark R. Patterson, *Inventions, Industry Standards, and Intellectual Property*, 17 BERKELEY TECH. L.J. 1043 (2002) [hereinafter Patterson, *Inventions*] (discussing the holdup inherent in ex post valuations of standard-essential patents); Mark R. Patterson, *Leveraging Information About Patents: Settlements, Portfolios, and Holdups*, 50 HOUS. L. REV. 483 (2012) [hereinafter Patterson, *Leveraging Information About Patents*]; Mark R. Patterson, *Antitrust and the Costs of Standard-Setting: A Commentary on Teece & Sherry*, 87 MINN. L. REV. 1995, 2001 n.33 (2003) [hereinafter Patterson, *Antitrust*]; Joseph Scott Miller, *Standard Setting, Patents, and Access Lock-in: RAND Licensing and the Theory of the Firm*, 40 IND. L. REV. 351 (2007).

later license to a party only if the patentee feels like it—the position some patentees have taken[15]—is not a commitment at all. Rather, to work properly, the FRAND commitment must itself be an undertaking by the patentee to limit its rights, and SSOs should make it clear that they regard it as one. Our preferred approach is that the FRAND commitment be treated as an enforceable license agreement with reasonable terms to be determined in the future, though if SSOs impose the arbitration system we propose, that approach is not strictly necessary.

We propose that the FRAND commitment be defined as follows. An SSO participant who makes a FRAND commitment is obliged to make a "FRAND offer" to any interested party who agrees to reciprocate. A "FRAND offer" means a purely monetary offer to license the SSO participant's entire portfolio of standard-essential patents on reasonable and non-discriminatory terms for the purpose of making, using, or selling products that comply with the standard. Crucially, the SSO participant promises that, if it cannot come to terms with another party implementing the standard, the question of the proper FRAND royalty rate will be subject to binding, baseball-style (or "final offer") arbitration.[16] SSO best practices include specifying a reputable arbitration association with established, unbiased rules for the conduct of the proceeding. The patentee's agreement to binding arbitration precludes it from going to court to enforce its standard-essential patents against implementers of the standard, except in very limited circumstances that we detail *infra*.

The obligation to make a FRAND offer does not prevent the standard-essential patent owner from entering into an alternative licensing arrangement, such as a portfolio cross license, with an implementer of the standard. It will often make sense for private parties to enter into a deal that reflects their specific circumstances. To help facilitate these deals while giving effect to the non-discrimination prong of FRAND, SSO best practices should include a mechanism by which the owner of standard-essential patents is obligated to disclose to any willing licensee the terms on which it has already licensed its standard-essential patents to other parties, subject to a suitable mechanism to protect the owner's confidential non-price business information.

The FRAND commitment in no way prevents or discourages private licensing agreements, and indeed we think they will be the norm.[17] Rather,

15. *See, e.g.*, Broadcom Corp. v. Qualcomm, Inc., 501 F.3d 297, 313–14 (3d Cir. 2007).

16. *See infra* Section I.B.2.

17. One complication comes from the "non-discriminatory" prong of the FRAND commitment. A patent owner making a FRAND commitment is obliged to offer similar

the FRAND commitment provides a fallback position (what economists call a threat point) should those negotiations fail. All of the nuances discussed *infra* come into play only if the owner of a patent subject to a FRAND commitment and an implementer are unable to reach an agreement giving the implementer the patent licensing rights necessary to produce products that comply with the standard. And no government actor is compelling this fallback position; it is a voluntary commitment between patent owners and the SSO, for the benefit of future implementers of the standard.[18]

B. WHAT HAPPENS WHEN NEGOTIATIONS FAIL?

Under our proposal, if a standard-essential patent owner and an implementer of the standard cannot agree on license terms, the patent owner is obligated to enter into binding arbitration to determine the FRAND royalty rate for its entire portfolio of standard-essential patents, so long as the implementer makes a reciprocal FRAND commitment for patents reading on the standard in question.[19] We denote an implementer who agrees to reciprocity and binding arbitration as a "willing licensee." If a standard-essential patent owner offers to enter into binding arbitration and the implementer refuses to make a reciprocal commitment or to submit to arbitration, the patent owner's FRAND commitment not to go to court to enforce its standard-essential patents against that party has been discharged.

1. Injunctions

As a best practice, SSOs should explicitly state in their IP policies that a patent holder making a FRAND commitment has given up its right to seek

terms to similarly situated parties. However, that does not mean that everyone must always pay the same price. Different types of buyers may reasonably be treated differently, and buyers who bring their own value to the table in the form of other patents to trade should generally expect to pay less than those who do not. For discussion of cross licenses and their implicit valuation of patents, see Mark A. Lemley & A. Douglas Melamed, *Missing the Forest for the Trolls*, 113 COLUM. L. REV. (forthcoming 2013), *available at* http://ssrn.com/abstract=2269087.

18. For this reason, whatever concerns have been expressed over Federal Trade Commission ("FTC") enforcement actions about the First Amendment right to petition the courts do not apply here. *See, e.g.*, Statement of Commissioner Maureen K. Ohlhausen, Statement of the Federal Trade Commission, Robert Bosch GmbH, No. 121-0081 (F.T.C. 2013), *available at* http://www.ftc.gov/os/caselist/1210081/121126boschohlhausen statement.pdf (articulating the First Amendment argument). An arbitration agreement is not state action limiting a patentee's right to sue, only a private agreement that does not raise constitutional issues.

19. While we want to encourage negotiation, patents have a limited life, and "willing" licensees should not be able to drag unproductive negotiations out forever. We suggest that either party can give the other a sixty-day notice of intent to seek binding arbitration; failing to participate in the arbitration triggers the remedies we discuss, *infra*.

an injunction against any willing licensee for infringement of any of its standard-essential patents. The commitment to binding arbitration aligns well with that approach. The matter of the FRAND royalty rate will be litigated in front of a private decision maker who does not have the power to issue an injunction. The FRAND commitment is also not consistent with seeking an exclusion order at the International Trade Commission ("ITC"), which is an injunction by another name. Making binding arbitration the exclusive remedy will preclude resort to the ITC in an effort to end-run the FRAND commitment, something scholars, the PTO, and antitrust agencies have urged in recent years.[20]

Explicitly ruling out injunctions will tend to steer bilateral negotiations towards a reasonable royalty rate. A key principle of bargaining theory is that the threat points of the two parties, along with their bargaining skills (which determine how their combined gains from reaching a deal are split), govern the outcome of bilateral negotiations.[21] So long as the arbitration procedure itself is unbiased, bargaining in the shadow of binding arbitration will tend to lead to reasonable rates. Introducing injunctions would drive negotiated royalty rates away from reasonable rates to artificially high ones reflecting the threat of holdup.[22]

---

20. U.S. DEP'T OF JUSTICE & U.S. PATENT & TRADEMARK OFFICE, POLICY STATEMENT ON REMEDIES FOR STANDARDS-ESSENTIAL PATENTS SUBJECT TO VOLUNTARY F/RAND COMMITMENTS, *supra* note 4; Federal Trade Commission, Third Party United States Federal Trade Commission's Statement on the Public Interest, Certain Gaming and Entertainment Consoles, Related Software, and Components Thereof, Inv. No. 337-TA-752 (USITC 2013), *available at* http://www.ftc.gov/os/2012/06/1206ftcgamingconsole.pdf; *see also* Colleen V. Chien & Mark A. Lemley, *Patent Holdup, the ITC, and the Public Interest*, 98 CORNELL L. REV. 1, 2–5 (2012). Notably, in a recent ITC investigation instigated by InterDigital Communications, the Commission refused to allow the investigation to go forward where the dispute arose under a license agreement that provided for arbitration. Notice, Certain Wireless Devices with 3G Capabilities & Components Thereof, Inv. No. 337-TA-800, 2013 WL 3361874 (USITC June 28, 2012). The fact that the ITC will enforce an arbitration agreement unless the claim of arbitrability is "wholly groundless," *Qualcomm Inc. v. Nokia Corp.*, 466 F.3d 1366, 1371, 1373 n.5 (Fed. Cir. 2006), while courts must stay resolution of a patent lawsuit pending an ITC proceeding, is an important benefit of the arbitration approach over judicial resolution.

21. *See* Lemley & Shapiro, *supra* note 14, at 1995–98 (providing an example of a bargaining model).

22. Chien & Lemley, *supra* note 20, at 8; Joseph Farrell et al., *Standard Setting, Patents, and Hold-Up*, 74 ANTITRUST L.J. 603, 616 (2007); Lemley & Shapiro, *supra* note 14; Doug Lichtman, *Understanding the RAND Commitment*, 47 HOUS. L. REV. 1023, 1033 (2010); Carl Shapiro, *Injunctions, Hold-Up, and Patent Royalties*, 12 AM. LAW & ECON. REV. 280 (2010). Ratliff and Rubinfeld find otherwise, but that's because they assume in their model that an injunction will only be granted after a FRAND rate has been finally set by a court and an implementer refuses to pay it. James Ratliff & Daniel L. Rubinfeld, *The Use and Threat of Injunctions in the RAND Context*, 9 J. COMPETITION L. & ECON. 1, 13, 21–22. That's not the

The standard-essential patent owner may seek an injunction against an unwilling licensee, as we discuss, *infra*. However, the court may well not grant an injunction. The court may well conclude that an SSO participant who has made a FRAND commitment has already declared that royalties are sufficient to compensate it for infringement by compliant products, so that the SSO participant will suffer no irreparable harm from infringement of its standard-essential patents.[23]

2. Procedure

We suggest that SSOs specify that disputes over what is FRAND be resolved through binding arbitration, "baseball-style."[24] In baseball-style arbitration, the parties produce evidence and argument before the arbitrator, and then they each propose a royalty number. The arbitrator must pick one of the two numbers offered and cannot come up with her own number. Using baseball-style arbitration logically drives the parties towards making reasonable proposals, because the party that asks for too much (or offers too little) risks losing the case altogether.[25] FRAND disputes are well suited to baseball-style arbitration, because the only thing at issue is which of two numbers in fact represents the more reasonable royalty.[26] We provide more details on how to determine that royalty in Section I.B.3, *infra*.

way things have happened in the real world. Patentees have sought injunctions in German courts and in the ITC without any resolution of FRAND issues. *See* Order, Interdigital Commc'ns Inc. v. Huawei Techs. Co., No. 1:13-cv-00008-RGA (D. Del. Mar. 14, 2013) (refusing to stay ITC proceeding that might lead to an injunction pending resolution of a FRAND royalty rate to which all defendants had agreed to be bound).

23. *See, e.g.*, Apple, Inc. v. Motorola, Inc., 869 F. Supp. 2d 901 (N.D. Ill. 2012) (Posner, J., sitting by designation). There may be policy reasons to prefer granting injunctions against unwilling licensees to encourage implementers to use the arbitration system we propose. In prior work, we have argued that injunctions should be available against defendants who copy a technology and seek to game the patent system by refusing to pay a reasonable royalty. Lemley & Shapiro, *supra* note 14, at 2036. Some, though not all, unwilling licensees will fit into this category.

24. Some SSOs already require arbitration of disputes. *See, e.g.*, VMEBUS INT'L TRADE ASS'N, VSO POLICIES AND PROCEDURES § 10.5 (2009), http://www.vita.com/home/VSO/vso-pp-r2d6.pdf; DIGITAL VIDEO BROADCASTING, MEMORANDUM OF UNDERSTANDING § 14.7 (2011), http://www.dvb.org/documents/DVB-MoU-2011.pdf.

25. J. Gregory Sidak, *Court-Appointed Neutral Economic Experts* 31 (J. Competition L & Econ., Working Paper, Apr. 21, 2013) ("[Baseball-style] arbitration has the effect of generating more credible estimates by altering the incentives of experts for either side to generate extreme values for their clients.").

26. One complication is that royalties in the real world are sometimes structured as lump-sum payments and sometimes as a percentage of ongoing sales. One of us has argued elsewhere that running royalties are better when calculating ongoing rather than past commitments. Mark A. Lemley, *The Ongoing Confusion over Ongoing Royalties*, 76 MO. L. REV. 695 (2011). But if both parties phrase their proposals in the same units, it doesn't matter

Baseball-style arbitration has a number of other advantages. The arbitrator does not need to decide whether any given patent is valid and infringed. Nor does she need to decide whether a particular patent is essential except in unusual circumstances.[27] Both of those things may be contested, and the evidence on each question will likely influence the reasonableness of the competing royalty proposals. But unlike a court that might have to rule on any number of subsidiary factual issues, the only thing the arbitrator needs to do is pick the better of two proposed royalty rates.

Under SSO best practices, any arbitration decision will be disclosed to willing licensees.[28] This disclosure is justified by the non-discrimination provision; it is hard to know whether a royalty unfairly favors one party unless we also know what other parties had to pay. Disclosure to willing licensees has other advantages: it will encourage implementers to submit to arbitration, reduce the need for duplicative arbitrations, avoid giving one party an informational advantage if they have already been involved in an arbitration, and help build a record of what constitutes FRAND royalty rates that may encourage subsequent parties to resolve their disputes themselves. In any given arbitration, the standard-essential patent owner and the licensee may well prefer to keep the arbitration outcome secret. For the reasons just given, such secrecy would undermine the effectiveness of the FRAND regime. And in any event, courts are not likely to permit it, at least when a

which they choose. If one party argues for a lump sum and the other for a running royalty, the arbitrator is choosing between apples and oranges. That makes the arbitrator's job harder, but by no means impossible; she simply must decide which approach better reflects what hypothetical negotiators would have chosen in that particular instance. If necessary, she can specify the royalty structure (e.g., lump sum vs. running royalties, or the royalty base to be used) to facilitate an apples-to-apples comparison of the rates proposed by the two parties. Beyond this, we do not think the arbitrator needs to or should resolve disputes over other non-price license terms. A FRAND license is by definition neither temporally limited nor limited to producing a set number of products; it provides terms that apply whenever the licensee makes products implementing the standard during the term of the patents. So there is no need for an arbitrator to decide on non-price terms like duration or output limits; there are no such terms. We can imagine parties wanting to include other limits in a license, and of course they are free to do so if they agree. But the FRAND obligation doesn't compel any such terms, so the arbitrator should not have to resolve them.

27. We can imagine a situation in which the parties to an arbitration dispute whether a particular patent is within the definition of an "essential" patent, and therefore whether the award will include a license to that patent. In that circumstance, it would be best if the arbitrator specified whether the patent in question is "essential" to minimize future litigation over whether or not certain patents have been licensed under the arbitration award.

28. Willing licensees also should have access to the terms on which these same standard-essential patents have been licensed to others, subject to suitable protections of confidential business information.

party to a subsequent dispute can show that the information is potentially relevant.[29]

Finally, like any arbitration, opportunities for appeal will be limited under this approach. Generally, parties to an arbitration can appeal only in cases of legal error or some procedural deficiency.[30]

3. Reasonable Royalties

The concept of a "reasonable royalty" is the heart of the patentee's right under the FRAND commitment, and it is the one thing the arbitrator will be called upon to decide. SSOs may differ in how they prefer to implement the FRAND concept. In an ideal world, SSOs would offer detailed guidance on what constitutes a reasonable royalty for a portfolio of standard-essential patents, whether or not they adopt all of the principles we favor. But they rarely do so.[31] In the absence of particularized guidance from an SSO, we offer a set of principles regarding the "reasonable royalty" concept that we believe should have widespread support.

Our starting point is the concept of reasonable royalties from U.S. patent law. The courts are very familiar with this concept, which they calculate in most patent damages cases.[32] That is not to say they always do it perfectly; far

29. *See, e.g.*, *In re* MSTG, Inc., 675 F.3d 1337, 1348 (Fed. Cir. 2012).

30. Hall St. Assocs. LLC v. Mattel, Inc., 552 U.S. 576 (2008) (enumerating limited grounds for appeal of arbitration award). There is some risk that if the patentee gets to choose the arbitration service, it will choose one known to be biased in its favor. But while courts are generally deferential to arbitral decisions, they have proven willing to intervene to reject arbitration agreements that are procedurally unfair. *See, e.g.*, Wheeler v. Noteworld LLC, No. 11-35984, 506 F. App'x. 543 (9th Cir. 2013) (affirming conclusion that abusive provisions in arbitration agreement were unconscionable).

31. *See* Lemley, *SSOs*, *supra* note 3 at 1913–14.

32. For an extended discussion of the relationship between damages under patent law and the FRAND concept in an SSO context, see Suzanne Michel, *Bargaining for RAND Royalties in the Shadow of Patent Remedies Law*, 77 ANTITRUST L.J. 889 (2011).

from it.[33] But we focus here on issues specific to the FRAND regime, not the more general challenge of determining reasonable royalty rates.[34]

Under patent law, a reasonable royalty normally is based on a hypothetical, arms-length negotiation between a willing buyer and a willing seller that takes place at the time the infringement begins.[35] For standard-essential patents, a reasonable royalty should be based on a hypothetical, arms-length negotiation that takes place at the time the SSO is setting the standard.[36] For parties making a FRAND commitment during the standard-setting process, the reasonable price is the price they would negotiate at that point, not a price that differs for each implementer depending on the happenstance of when that party begins implementing the standard. SSO

33. For criticism of particular reasonable royalty doctrines and suggestions for improvement, see, for example, Daralyn J. Durie & Mark A. Lemley, *A Structured Approach to Calculating Reasonable Royalties*, 14 LEWIS & CLARK L. REV. 627 (2010); Amy L. Landers, *Patent Claim Apportionment, Patentee Injury, and Sequential Innovation*, 19 GEO. MASON L. REV. 471 (2012); Amy L. Landers, *Let the Games Begin: Incentives to Innovation in the New Economy of Intellectual Property Law*, 46 SANTA CLARA L. REV. 307 (2006) [hereinafter Landers, *Let the Games Begin*]; Mark A. Lemley, *Distinguishing Lost Profits from Reasonable Royalties*, 51 WM. & MARY L. REV. 655 (2009); Brian J. Love, *The Misuse of Reasonable Royalty Damages as a Patent Infringement Deterrent*, 74 MO. L. REV. 909 (2009); Brian J. Love, *Patentee Overcompensation and the Entire Market Value Rule*, 60 STAN. L. REV. 263, 278 (2007) [hereinafter Love, *Patentee Overcompensation*].

34. For example, issues of royalty base, double payment, and upstream/downstream rights have come up in ordinary patent cases as patentees have moved from targeting makers of components to going after downstream customers and even individual end users. *See, e.g.*, Tim Steller, *Local Firm Faces Heat of Patent Enforcer*, ARIZ. DAILY STAR, Feb. 17, 2013, http://azstarnet.com/news/local/tim-steller-local-firm-faces-heat-of-patent-enforcer/article_456aaa3f-893e-5465-8e93-73c849415fad.html (describing a growing trend of large companies sending demand letters to small businesses stating that they own patents to "common technologies or processes"). Those issues, thorny as they are, are not unique to the FRAND context. *See, e.g.*, Quanta Comp. v. LG Elecs., 553 U.S. 617 (2008) (discussing the rules of patent exhaustion when patentees seek to recover from both upstream and downstream companies).

35. *See* Lemley, *supra* note 33, at 666–68.

36. As noted, *supra*, the idea that a reasonable royalty should reflect the ex ante value of the patented technology, over and above the best alternative, is far from new. This is the approach recommended by the Federal Trade Commission. *See* FED. TRADE COMM'N, THE EVOLVING IP MARKETPLACE: ALIGNING PATENT NOTICE AND REMEDIES WITH COMPETITION 22–23 (2011), *available at* http://www.ftc.gov/os/2011/03/110307patentreport.pdf. The European Commission also takes this approach. European Comm'n, Guidelines on the Applicability of Article 101 of the Treaty on the Functioning of the European Union to Horizontal Co-operation Agreements, 2011 O.J. (C 11) ¶ 289, *available at* http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2011:011:0001:0072:EN:PDF. This basic idea is discussed in Farrell, et al., *supra* note 22, and is reflected in the ex ante auction model advocated by Dan Swanson and William Baumol. Daniel G. Swanson & William J. Baumol, *Reasonable and Nondiscriminatory (RAND) Royalties, Standard Selection, and the Control of Market Power*, 73 ANTITRUST L.J. 15 (2005).

best practice includes explicitly noting the timing and context for the hypothetical ex ante negotiation underlying the reasonable royalties concept.

The hypothetical negotiation needs to take place under conditions where the alternative specifications have been identified, so that the parties are well informed about the best potential non-infringing alternatives to the proposed standard.[37] In some cases, the best ex ante technological alternatives would have required some development effort by SSO participants, and could not simply have been taken off the shelf. The key idea here is that a reasonable royalty should reflect what would happen as a result of well-informed ex ante technology competition. The incremental value of the patented technology over and above the next-best alternative serves as an upper bound to the reasonable royalties. To this end, SSO best practice includes maintaining records, such as minutes from SSO meetings, that will inform subsequent negotiators and arbitrators of the ex ante technical alternatives that were feasible or considered, along with their pros and cons.

By construction, the reasonable royalty rate does *not* include the value attaching to the creation and adoption of the standard itself.[38] To allow patentees to capture that value, which flows from the collective adoption decisions of the group rather than from the underlying value of the technology chosen, would undermine the goals of the FRAND commitment.[39]

---

37. The hypothetical ex ante negotiation is not intended to reflect what an actual ex ante negotiation would have looked like. For starters, SSO participants were unlikely to have known what patents covered what specifications. But reasonable royalty calculations have always assumed a counterfactual world. The point is not to reconstruct what the parties actually did; obviously they did not agree on a price ex ante, or there would be no dispute. Rather, the point of the hypothetical negotiation rule in patent damages is to determine what hypothetical reasonable parties might have done, had they had all the facts, including knowledge of non-infringing alternatives. *See, e.g.*, Grain Processing Corp. v. Am. Maize-Prods. Co., 185 F.3d 1341, 1350–53 (Fed. Cir. 1999). The FRAND royalty concept is no different in this respect.

38. Apple, Inc. v. Motorola, Inc., 869 F. Supp. 2d 901, 913 (N.D. Ill. 2012); Patterson, *Inventions*, *supra* note 14, at 1048; Mark A. Lemley, *Ten Things to Do About Patent Holdup of Standards (and One Not to)*, 48 B.C. L. REV. 149, at 158–59 (2007).

39. For this reason, the hypothetical bilateral negotiation can, if necessary, involve some communication, but not explicit coordination, among SSO members who are implementers to avoid an outcome in which the patent holder uses a "divide and conquer" strategy to support royalty terms and an associated equilibrium technology adoption outcome that is inferior to another equilibrium for a blocking coalition of implementers. *See generally* Ilya Segal & Michael D. Whinston, *Robust Predictions for Bilateral Contracting with Externalities*, 71 ECONOMETRICA 757 (2003) (studying bilateral contracting between one principal and some number of agents where each agent's utility depends on the principal's unobservable trades with other agents). For an analysis of how independent, bilateral negotiations, along with the non-discrimination prong of FRAND, can deter patent holdup,

The hypothetical negotiation over the FRAND commitment is a bilateral negotiation between the patent holder and one implementer. That doesn't mean other implementers are irrelevant. Deals with those parties may be evidence of a reasonable price, and certainly constitute relevant evidence in the arbitration. Plus, deals with other parties may be binding due to the non-discrimination commitment.[40] But the hypothetical ex ante negotiation is not one in which all the buyers act collectively to reduce prices.[41] It is one in which patentees negotiate with individual licensees just as they would in any other circumstance, subject to the constraint that they have pre-committed not to discriminate.[42] The commitment to license to all comers makes the auction approach proposed by some[43] inappropriate.

Royalty stacking arises when implementers must pay royalties to multiple patent owners, so those royalties cumulate or "stack" on top of each other from the perspective of the implementer.[44] To address the problem of royalty stacking, the hypothetical negotiation needs to reflect the presence of patents held by others that read on the same product. A real-world negotiation would not consider in a vacuum one party's standard-essential patent portfolio, or even the standard-essential patents associated with one of many standards being implemented in a given product. The price any implementer would be willing to pay for a given standard-essential patent portfolio depends on the other royalty payments they will be asked to make to bring their product to market. For that reason, the hypothetical negotiation needs to reflect and account for reasonable royalties for standard-essential patents held by others,

see Richard Gilbert, *Deal or No Deal? Licensing Negotiations in Standard-Setting Organizations*, 77 ANTITRUST L.J. 855 (2011).

40. We do not discuss the non-discrimination commitment in detail here. For other work considering it, see, for example, Lemley, *supra* note 3, at 1913–14; *Intellectual Property and Standard-Setting*, *in* ABA HANDBOOK ON THE ANTITRUST ASPECTS OF STANDARDS SETTING 95 (2d ed. 2011); Gilbert, *supra* note 39.

41. On the risk of buyers' cartel inherent in SSO IP policies, see 2 HERBERT HOVENKAMP ET AL., IP AND ANTITRUST § 36.6b (2010); see also Carl Shapiro, *Setting Compatibility Standards: Cooperation or Collusion?*, *in* EXPANDING THE BOUNDS OF INTELLECTUAL PROPERTY 81 (2001).

42. For example, a patentee who makes or sells compliant products cannot charge a higher royalty to an implementer against whom the patentee competes directly than to another party who sells very distinct compliant products.

43. *See, e.g.*, Damien Geradin, Anne Layne-Farrar & A. Jorge Padilla, The Ex Ante Auction Model for the Control of Market Power in Standard Setting Organizations (Apr. 9, 2007) (unpublished manuscript), http://ssrn.com/abstract=979393; David L. Newman, *Going Once . . . Going Twice . . . Licensed Under the Most Reasonable and Non-Discriminatory Bidding Terms!*, 11 NW. J. TECH. & INTELL. PROP. 139 (2013). An auction presupposes licenses to a subset of bidders, instead of all of them, as FRAND requires.

44. *See*, *e.g.*, Lemley & Shapiro, *supra* note 14, at 1993.

just as a reasonable royalty in patent damages, properly understood, must include the concept of apportionment of the value of a product among multiple contributors to that value.[45] This is part of what is intended by the "well-informed" aspect of the hypothetical ex ante negotiations.

SSO best practices should include an instruction to the arbitrator to consider all patents declared essential to the standard in question, not just the portfolio of standard-essential patents submitted to arbitration. The economics literature demonstrates that when multiple essential inputs are priced independently, collective overpricing tends to result, due to the "Cournot complements" problem.[46] This overpricing, which appears in practice in the form of royalty stacking, reduces the collective returns to standard-essential patent owners *and* to implementers.[47] To avoid this mutually undesirable outcome, SSO best practices should acknowledge the problem of royalty stacking, empower the arbitrator to account for royalty stacking, and provide the arbitrator with the best possible information to do so.[48] Arbitrators (and courts) can gain insight from the commercial arrangements companies have employed to deal with royalty stacking, notably in the context of patent pools or other mechanisms involving the aggregation of essential patents.[49] An arbitrator does not have the luxury of resolving all standard-essential patent disputes together, at least absent some

45. *See, e.g.*, Landers, *Let the Games Begin*, *supra* note 33, at 354–62; Love, *Patentee Overcompensation*, *supra* note 33, at 268–69; Patterson, *Leveraging Information About Patents*, *supra* note 14, at 503–13.

46. The "Cournot complements" problem arises when multiple necessary inputs are supplied by separate firms, each with market power. In Cournot's original example, one firm had a monopoly over copper and another had a monopoly over zinc, two inputs essential for making brass. ANTOINE AUGUSTIN COURNOT, RESEARCHES INTO THE MATHEMATICAL PRINCIPLES OF THE THEORY OF WEALTH 99–116 (Nathaniel T. Bacon trans., Augustus M. Kelley Publishers 1971) (1838). For discussion in the context of patents, see Lemley & Shapiro, *supra* note 14, at 2013–17; Carl Shapiro, *Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting*, 1 INNOVATION POLICY & ECON. 119, 122–24 (2000).

47. A lower aggregate royalty rate would lead to more sales of the final product, raising the profits of the patent holders and the implementers. *See* Shapiro, *supra* note 14, at 2013–17.

48. Royalty stacking applies to *all* of the patents reading on a given product, not just the standard-essential patents associated with the standard at issue. Therefore, in principle, the reasonable aggregate royalty rate for all standard-essential patents reading on a given standard should depend on the entire set of patents reading on the product. SSOs may want to point this out, as part of providing guidance to the arbitrator. However, it may be difficult in practice for the arbitrator to learn about and account for patents other than standard-essential patents associated with the standard at issue. The arbitrator may find it more workable to maintain a focus on the ex ante incremental value of the standard-essential patents at issue.

49. *See* Shapiro, *supra* note 46, at 134–36; Richard J. Gilbert, *Antitrust for Patent Pools: A Century of Policy Evolution*, 2004 STAN. TECH. L. REV. 3, ¶ 2.

mechanism in the SSO rules to handle all standard-essential patent disputes on the same patent in the same arbitration proceeding.[50] But the arbitrator can and should take evidence on the existence of other standard-essential patent portfolios the licensee would have to pay; that evidence bears on the royalty rate.[51]

We close this Section by noting two significant respects in which a FRAND royalty arbitration is different than a patent damages case. First, in a FRAND royalty arbitration, there is no need to determine the reasonable royalty on a patent-by-patent basis. Indeed, doing so would be exceedingly difficult and costly for large patent portfolios. The FRAND concept involves a reasonable rate for a party's entire portfolio of standard-essential patents.[52] Establishing a FRAND rate for an entire standard-essential patent portfolio is simpler than—and matches more closely to—real-world licensing practices in the information technology sector, where implementers commonly negotiate portfolio licenses that give them freedom to operate.[53]

Second, unlike the "reasonable royalty" concept used to calculate damages in patent infringement cases, the hypothetical negotiation for FRAND purposes does *not* assume that any particular patent, much less the entire standard-essential patent portfolio, is valid and infringed.[54] Unlike a patent infringement case, the reasonable royalty for a portfolio is a function of the probability that the patents in that portfolio are actually valid and infringed.[55] Reasonable royalties will logically be lower for patents that are more likely to be found invalid or not infringed. That certainly does not

50. Some have suggested such a "pseudo-pool." *See* Contreras, *supra* note 5, at 26–33.

51. Evidence on pricing from patent pools may be relevant to pricing decisions in any event. *See* Jorge L. Contreras, Rethinking RAND: SDO-Based Approaches to Patent Licensing Commitments 14–15 (Oct. 10, 2012), http://ssrn.com/abstract=2159749.

52. There are some advantages to establishing a FRAND rate for a portfolio of standard-essential patents equal to the sum of the reasonable royalty rates for all of the individual standard-essential patents in that portfolio. Under this "neutrality principle," patent holders have no incentive to split up their portfolios, or combine them, to increase the overall FRAND royalty rate. However, the neutrality principle does not generally follow from the hypothetical, ex ante negotiation construct. Nor does it mean in practice that one needs to compute the FRAND rate for each patent.

53. *See, e.g.*, Gideon Parchomovsky & R. Polk Wagner, *Patent Portfolios*, 154 U. PA. L. REV. 1 (2005); Lemley & Melamed, *supra* note 17.

54. For this reason, we cannot simply apply the reasonable royalty standards used in patent law, as some have suggested. *See* Anne Layne-Farrar, A. Jorge Padilla & Richard Schmalensee, *Pricing Patents for Licensing in Standard-Setting Organizations: Making Sense of FRAND Commitments*, 74 ANTITRUST L.J. 671, 679–82 (2007).

55. A patent is inherently a probabilistic right. *See* Mark A. Lemley & Carl Shapiro, *Probabilistic Patents*, 19 J. ECON. PERSP. 75 (2005) (discussing the economics of probabilistic patents).

mean that the parties are required to litigate validity and infringement on a patent-by-patent basis, and we very much doubt they will want to do that for the entire portfolio. But the significance and strength of the portfolio is key to determining a reasonable price for that portfolio.

When arbitration is invoked, we expect that both parties will often present evidence on validity because it may affect the royalty awarded.[56] This is entirely appropriate and desirable, just as parties negotiating a patent license typically spar over validity and infringement. The arbitrator should account for the likelihood of validity and infringement, along with the significance, of the patents at issue in deciding which of the two royalty rates proposed by the parties for a portfolio of standard-essential patents is more reasonable.

We don't mean to suggest that determining the FRAND royalty for a portfolio of standard-essential patents is an easy matter. It isn't. Indeed, we expect FRAND arbitration will often involve extensive discovery, given the range of information relevant to determining which of the two offers made is more reasonable. Nonetheless, the principles articulated here will make arbitration far more predictable than litigation, greatly increasing the efficiency and accuracy with which FRAND disputes are resolved.

C. THE BOUNDARIES OF THE FRAND ARBITRATION

There are some complications to the basic framework, though fewer than one might expect. Most center on determining the circumstances under which patent owners are able to enforce their standard-essential patents by means other than submitting them to binding arbitration for the determination of a FRAND royalty rate.

1. Unwilling Patentees and Unwilling Licensees

What happens if a standard-essential patent owner who has made a FRAND commitment with binding arbitration goes to court to enforce its standard-essential patents? The defendant in that proceeding would have the opportunity to argue that the patents asserted are essential to the standard and that the court should therefore compel arbitration, per the patentee's commitment. Motions to compel arbitration are routinely granted in the United States.[57] And they are considered at the outset of a case.[58] So a patentee cannot simply opt out of a binding arbitral commitment.

---

56. If both parties stipulate that a given patent is essential, infringement of that patent may not be disputed.

57. The Federal Arbitration Act, 9 U.S.C. §§ 1–16 (2012), declares a federal policy in favor of arbitration, Nitro-Lift Techs. v. Howard, 133 S. Ct. 500 (2012), and courts have

Implementers, by contrast, have made no such commitment. When a patentee makes an arbitration demand to a potential licensee, the licensee can either choose to participate in the process or refuse to do so. If the implementer participates in the process, it will be bound by the result. It can't later decide it dislikes the result and refuse to pay. An arbitral award can be enforced in court. In addition, under our approach, if the patent holder can convince an authorized fact finder that the implementer is likely to evade paying the royalties or be unable to do so, the fact finder may require the implementer to make payments into escrow or post a suitable bond. If the implementer then fails to comply with this requirement, they are effectively refusing to participate in the process.

If the potential licensee refuses to participate in a royalty-setting arbitration as to a standard-essential patent, the patentee can sue that party in court for infringing that patent, seeking damages and injunctive relief. A commitment to license on reasonable terms is not a commitment to be whipsawed by a potential licensee. An implementer who agrees to participate only if it gets a result it likes[59] is no different than a patentee who agrees to license on reasonable terms only if it gets to decide what is reasonable. Neither party is acting in good faith. A patentee who makes a FRAND commitment is obligated to agree to reasonable licensing terms, but does not have to license to someone who will not make a similar commitment to accept reasonable terms set by the arbitrator.

2. Essential Versus Non-Essential Patents

The FRAND commitment applies to standard-essential patents. Under our proposal, a patentee who makes a FRAND commitment is free to litigate normally over other, non-essential patents in its portfolio. But what happens when the parties cannot agree whether a patent is essential?

For starters, SSOs can and should limit disputes over what is an "essential" patent by clearly defining that term. Based on our experience, we suggest that the SSOs define a patent to be essential if any product complying with the standard will infringe on that patent. Any narrower

proven willing to enforce that policy even in controversial circumstances. *See, e.g.*, AT&T Mobility LLC v. Concepcion, 131 S. Ct. 1740 (2011) (enforcing clause banning class actions in favor of arbitration).

58. Indeed, normally any court proceedings must be stayed in favor of arbitration. *See, e.g.*, *In re* Pharmacy Benefit Managers Antitrust Litig., 700 F.3d 109, 116 (2d Cir. 2012) ("a party to a valid and enforceable arbitration agreement is entitled to a stay of federal court proceedings pending arbitration as well as an order compelling such arbitration.").

59. *See, e.g.*, Apple, Inc. v. Motorola Mobility, No. 11-CV-178-BBC, 2012 WL 5416941 (W.D. Wisc. Oct. 29, 2012).

definition would not prevent patentees from holding up implementers. However, some SSOs may conclude that a broader definition, capturing some notion of "commercially essential" rather than just "technically essential," is superior for them.

One possible issue is over-declaring: patentees claiming patents are essential when they aren't. As a general matter, we don't see much risk in over-declaring given our approach. If a patentee declares patents essential, it commits itself to the arbitration procedure we have outlined. If the patent is not in fact truly essential, the patentee could have sued to enforce the patent in court and sought an injunction. The fact that the patentee voluntarily gives up those rights when it didn't need to doesn't create a problem for the rest of the world. We can imagine a patentee wanting to do this to pad its portfolio if the default or starting FRAND rate on a portfolio is proportional to the number of declared essential patents, but that can easily be solved by giving the other side the ability to argue that a particular patent isn't worth much because it is easy to design around (so it isn't essential). The arbitrator could determine whether a patent is truly essential, but we doubt she would have to in very many cases.

Another issue is under-declaring: trying to avoid a FRAND commitment on patents that truly are essential. This presents a more serious risk; a patent holder may seek to narrow, avoid, or evade its FRAND commitment by claiming that a given patent is not essential.[60] A patentee that claims its patent is not essential presumably will not invoke the arbitration proceeding, but will instead sue in court or threaten litigation. If the SSO has made arbitration mandatory for essential patents, as we propose, a defendant in that lawsuit would have the opportunity to argue that the patent was in fact essential to the standard and therefore to compel arbitration. The court would have to determine whether the patent was in fact essential because it was technically necessary to comply with the standard. Similarly, if a patentee asserts a patent in court after binding arbitration is over and claims that the patent was not essential and therefore not part of the arbitration, the defendant could argue that the patent was essential and thus covered by the license.

We don't see a good alternative to courts making these determinations. We will note, however, that the behavior of the parties themselves will sometimes offer evidence on this score. A patentee asserting that

60. Some SSOs require a listing of essential patents, but most do not, see Lemley, *SSOs*, *supra* note 3, and it is often impractical for a patentee to make that judgment as the standard is being set, since many patents will not yet have issued and their claims may change over time.

implementation of the standard itself infringes is necessarily arguing that the patent is essential to the standard.[61] A patentee who successfully denies its patent is essential to the standard should have a hard time then persuading a court that a product infringes simply because it complies with that standard.[62] Conversely, a defendant that successfully compels arbitration over a patent should not be permitted to deny in that arbitration that its compliant product infringes that patent.

One particular species of under-declaring has to do with patent applications and other yet-to-be-patented inventions. It seems clear that the FRAND obligation should extend not just to issued patents but to pending patents as well,[63] and we urge SSOs to make that explicit.[64] Inventors can keep patent applications pending in the PTO for years or even decades and can even seek additional new patents from old applications,[65] and the PTO takes years to issue a patent.[66] It would make little sense to limit a FRAND commitment only to those patents that happen to have issued by the time the standard is adopted. And while the scope of patent claims can certainly change during prosecution, so that we might not know at the outset whether a particular patent was essential to the standard, as noted, *supra*, we don't need to know exactly which patents are essential to include them in the FRAND commitment.

We would apply the same analysis to an idea developed at the time a standard issues but not yet the subject of a patent application. For an invention to be essential to a technical standard, it presumably must have been made at the time that standard is adopted. With limited exceptions, an inventor with an idea essential to a technical standard must file an application within one year after adoption of the standard or lose rights to the invention under the statutory bars.[67] Accordingly, we suggest that SSOs specify that the

61. Fujitsu Ltd. v. Netgear Inc., 620 F.3d 1321, 1327 (Fed. Cir. 2010) (allowing owner of a standard-essential patent to show infringement by showing the defendant complied with the standard).

62. *See* Apple, Inc. v. Samsung Elecs. Co., 695 F.3d 1370, 1374 (Fed. Cir. 2012) (requiring a causal nexus between infringement and sales for injunctive relief).

63. *See* Lemley, *supra* note 3, at 1958 (making this argument).

64. Some already do so. *See id.* at 1904–11; Bekkers & Updegrove, *supra* note 3, at 44–45 (identifying SSOs that apply FRAND policies to patent applications).

65. *See, e.g.*, Mark A. Lemley & Kimberly A. Moore, *Ending Abuse of Patent Continuations*, 84 B.U. L. Rev. 63, 72–80 (2004).

66. *See* John R. Allison & Mark A. Lemley, *Who's Patenting What? An Empirical Exploration of Patent Prosecution*, 53 VAND. L. REV. 2099, 2118–21 (2000) (finding that the average patent takes 2.77 years to issue).

67. The America Invents Act changed the definition of prior art, but if anything it shortened the grace period an inventor is allowed before filing. *See* 35 U.S.C. § 102 (2012).

FRAND commitment applies not only to existing patents and applications but also—at the very least—to those applications filed within one year after the SSO adopts the standard.[68]

3. Reciprocity

A patentee that makes a FRAND commitment to an SSO covering a particular standard may reasonably expect that others with essential patents covering the same standard will make the same commitment. Under our proposal, therefore, a FRAND offer to a party that owns standard-essential patents can be made conditional on the would-be licensee itself making a reciprocal FRAND offer. By definition, that offer must cover the would-be licensee's portfolio of standard-essential patents reading on the same standard.

When the would-be licensee is a member of the same SSO, as will often be the case since those with patents that read on a standard often participate in the SSO that sets that standard, we don't need any special rule or mechanism to make this happen. Each party will have made the FRAND commitment, and each party is bound by their commitment. If one party tries to go back on that commitment, the other can simply move to compel arbitration, and that should be their remedy. There is no need for defensive suspension of the FRAND commitment in this circumstance.

By contrast, when a potential licensee also owns patents essential to the same standard but has *not* made a FRAND commitment to the SSO, the patentee can find itself in an unfair position: threatened with injunctive relief over the standard but unable to respond in kind. Thus, we think it makes sense to permit the patent owner to require reciprocity by other essential patent owners who did not participate in the SSO and thus did not themselves make a FRAND commitment. In this circumstance, defensive suspension helps protect SSO members from holdup by non-members and encourages participation in the FRAND commitment and perhaps the SSO itself.

Defensive suspension also helps resolve another thorny issue: SSO members who make a commitment to license their patents on FRAND terms and then later acquire different standard-essential patents from a third party who made no similar commitment. We suggest in Section I.C.4 that the

68. We favor applying the FRAND commitment to all future applications filed by the party making that commitment. However, for applications filed much later, the danger of the implementer gaming the system by claiming that a subsequent patent is a standard-essential patent may exceed the danger of the patentee gaming the system by obtaining a truly essential patent not subject to the FRAND commitment.

FRAND commitment should run with the patent; applying that position consistently would lead to the conclusion that the newly-acquired standard-essential patents are not subject to the FRAND commitment. While this might seem troubling, since it allows a patentee to make a FRAND commitment and later acquire standard-essential patents that avoid that commitment, the fact is that those standard-essential patents would not have been subject to that commitment in anyone else's hands. The fact that the acquirer participated in the SSO shouldn't change that result. In any event, if the FRAND commitment is subject to defensive suspension, acquiring patents becomes much less of an issue: a party that acquires and asserts non-FRAND-encumbered standard-essential patents can expect to be sued for injunctive relief and damages by a FRAND patentee invoking its defensive suspension rights.

Whether defensive suspension can extend beyond patents essential to the standard at issue is a harder question. If an implementer has a patent essential to a different standard that ends up in the same product, and that standard is not promulgated by an SSO that requires a FRAND commitment following our best practices, the patentee who has made a commitment will be at a disadvantage.[69] And if the effect of such a provision was to encourage more reciprocal FRAND commitments the world would probably be a better place. On the other hand, extending defensive suspension too far would undo the benefits of the FRAND commitment. A patentee is not entitled to demand that no one assert any patents in any field against it as a condition of following through on its FRAND commitment.[70] And it is at least as plausible that an overbroad defensive suspension clause would end up nullifying FRAND commitments as that it would induce reciprocal ones. While the issue is not free from doubt, we think that an offer made conditional on the would-be licensee licensing any patents *other* than standard-essential patents reading on the standard at issue is *not* a FRAND offer. This approach also has the benefit that it does not require affirmative

69. Thus, we sympathize with the plight of Google-Motorola, which was investigated for antitrust violations for suing to enforce its patents only in response to having its Android operating system challenged in patent suits by Apple and Microsoft. Motorola Mobility LLC, No. 121-0120, 2013 WL 124100 (F.T.C. Jan. 3, 2013). But the fact that Motorola made a FRAND commitment and Apple and Microsoft didn't means the situations are not the same. *See* Thomas H. Chia, Note, *Fighting the Smartphone Patent War with RAND-Encumbered Patents*, 27 BERKELEY TECH. L.J. 209 (2012).

70. *See* News Release, Fed. Trade Comm'n, FTC Accepts Settlement of Charges Against Intel (Mar. 17, 1999), http://www.ftc.gov/opa/1999/03/intelcom.shtm (settling antitrust claim based on refusal to share intellectual property with any company that sued Intel seeking an injunction).

action by each patentee. The SSO can simply provide that a FRAND commitment can be suspended as to any party that holds essential patents covering the same standard who is not willing to make the same commitment.[71]

4. Transfer of Standard-Essential Patents

The FRAND obligation should travel with the patent. The economic case for this is overwhelming. The very point of the commitment is to comfort implementers that they will not be held up by parties refusing to license patents essential to the standard.[72] If a patentee can undo the FRAND commitment merely by selling its patents to someone who has not personally made that commitment, that comfort is illusory.[73] Large technology companies have increasingly turned to "patent privateering" in an effort to raise money or to raise rivals' costs, selling part of their portfolios to patent assertion entities to enforce those patents against others in the industry.[74] Nokia has spun out patents to Mosaid to use against its rivals; Micron has sold thousands of patents to patent assertion entities, and so on.[75]

Fortunately, this is a relatively easy problem for SSOs to solve with suitably crafted patent rules. These rules should prevent a patentee from avoiding or weakening its FRAND commitment merely by selling encumbered patents. Patent law has previously encountered the problem of people who try to sell the same right twice, and the general rule is that if you transfer all or part of a right to one party and record that transfer with the patent office, any subsequent buyer takes subject to that restriction even if

71. One implication of our approach is that most open source licenses will not be FRAND compliant because they condition a license on things beyond reciprocal licensing of standard-essential patents. *Cf.* Greg Vetter, *Patenting Cryptographic Technology*, 84 CHI.-KENT L. REV. 757, 774 (2010) (noting that open-source advocates oppose RAND in favor of royalty-free or, better still, viral patent licensing); Jason Schultz & Jennifer M. Urban, *Protecting Open Innovation: The Defensive Patent License as a New Approach to Patent Threats, Transaction Costs, and Tactical Disarmament*, 26 HARV. J. L. & TECH. 1 (2012) (proposing an open-source viral patent license).

72. *See supra* text accompanying note 11.

73. In fact the comfort can only ever be partial, because there is no way to assure that there are no parties out there who don't belong to the SSO and will pop up claiming to own an essential patent. But if the major industry players participate in the SSO, that risk is at least minimized.

74. For discussions of privateering, see, for example, Tom Ewing, *Indirect Exploitation of Intellectual Property Rights by Corporations and Investors*, 4 HASTINGS SCI. & TECH. L.J. 1 (2012); Lemley & Melamed, *supra* note 17.

75. Lemley & Melamed, *supra* note 17; Andrei Hagiu & David B. Yoffie, *The New Patent Intermediaries: Platforms, Defensive Aggregators, and Super-Aggregators*, 27 J. ECON. PERSP. 45 (2013).

they didn't know about it.[76] Put another way, after recordation there is no bona fide purchaser for value doctrine in patent law.[77] So the solution to the problem is easy if we think of a FRAND commitment as an executory license without a price term, rather than a mere promise to license in the future.[78] But even if we don't, the law can and should treat a binding commitment to license a patent to all comers as something that encumbers the patent itself, so that merely selling the patent cannot release the commitment.[79] Just as a mendacious patentee can't whitewash inequitable conduct by selling the patent to someone who didn't lie to the patent office,[80] a patentee that has promised that a patent will not be enforced by means of an injunction can't wipe away that commitment by finding a buyer who didn't make that promise. While we think patent or antitrust law would find their way to that result if need be,[81] the SSO could help matters along by making clear that a patentee's FRAND commitments bind not only itself but also its successors, and by keeping a record of the FRAND commitments it receives and making them available to willing licensees. Coupled with an accurate and up-to-date PTO database of patent assignments, it should be straightforward for a willing licensee to determine whether a patent is owned by a company that made a FRAND commitment.

Once we understand that a FRAND commitment travels with the patent, many of the complexities around transfer fall away. Patentees will have no incentive to sell off part of a FRAND-encumbered portfolio to privateers to evade their FRAND commitment. Legitimate transactions may leave

76. *See* 35 U.S.C. § 261 (2012). *But see* Board of Trustees of Leland Stanford Jr. Univ. v. Roche Molecular Sys., 131 S. Ct. 2188 (2011) (finding an exception to this rule). *Stanford* dealt with a situation in which the Court concluded there was no transfer of the right to the first acquirer because the contract with Stanford only promised to later assign patents, rather than actually operating to assign them. *Id.* at 2197. Treating a FRAND commitment as an executory license would solve this problem, though if SSOs take our approach it is not necessary.

77. *See* 35 U.S.C. § 261.

78. For an argument that we should do just that, see Lemley, *supra* note 3, at 1914–16.

79. For an argument that this is permissible under the law of servitudes, see Jay Kesan, *FRAND's Forever: Standards, Patent Transfers and Licensing Commitments*, IND. L.J. (forthcoming 2013), *available at* http://ssrn.com/abstract=2226533.

80. *See, e.g.*, 1st Media, LLC v. Elec. Arts, Inc., 694 F.3d 1367, 1369–72 (Fed. Cir. 2012) (considering inequitable conduct claim even though inventor accused of inequitable conduct was no longer the patent owner).

81. Implied license or equitable estoppel doctrines might limit enforcement of such a patent. *See* Mark A. Lemley & David McGowan, *Could Java Change Everything? The Competitive Propriety of a Proprietary Standard*, 43 ANTITRUST BULL. 715 (1998). Alternatively, the Federal Trade Commission has shown a willingness to rely on Section 5 of the FTC Act to prevent fraudulent conveyances like this. *See* Negotiated Data Solutions LLC (*N-Data*), No. C-4234, 2008 WL 4407246 (F.T.C. Sept. 22, 2008).

standard-essential patents in different hands, but there is no reason to think they will change the process we have described, *supra*, or the royalty rate the patents will command in an arbitration proceeding. A buyer may get a higher total royalty if it now owns more standard-essential patents, but the seller should get a correspondingly smaller royalty.[82] And an implementer who has already paid for a license to a patent does not need to pay again for that patent just because it was sold, though of course the implementer may have to pay to license different patents from the buyer.

II. SIMPLIFYING THE FRAND DEBATES

Much of the time and effort that has been spent by courts and commentators in the FRAND debate seems misdirected. Many of the contested issues are sideshows that become moot under our proposals. Here are some important examples.

A. BREACH OF FRAND COMMITMENT

Parties have spent a great deal of time litigating the question of whether one or both sides have breached a FRAND commitment. Implementers argue that patentees' offers to license are not really reasonable, and thus breach the patentees' commitment to license on FRAND terms. Patentees argue that implementers are not negotiating in good faith and therefore are infringers rather than putative licensees working to figure out the right royalty payment.[83] Courts are asked to resolve what constitutes a breach of the FRAND commitment. Doing so necessarily requires the court to make a substantive judgment regarding what is a reasonable royalty, and then a second judgment of whether one or both parties' offers were reasonably close to a reasonable royalty.[84] And it blurs the line between settlement

82. Disaggregating patent rights to try to artificially increase the total royalty is socially costly. *See* Lemley & Melamed, *supra* note 17, at 30 (discussing the ways disaggregation can increase the Cournot Complements problem, reducing the efficiency of patent licensing). Disaggregation to evade a FRAND commitment should be discouraged or prevented by SSO patent rules. This is one reason why it is desirable, in principle, that the reasonable royalty rate for a standard-essential patent portfolio equal the sum of the reasonable royalties for the individual patents in that portfolio, as noted, *supra* text accompanying note 52. If that "neutrality principle" applies, a patent owner has no incentive based on the FRAND regime to sell its portfolio to a third party or split up its portfolio.

83. *See* Microsoft Corp. v. Motorola, Inc., 696 F.3d 872, 878 (9th Cir. 2012); Apple, Inc. v. Motorola Mobility, No. 11-CV-178-BBC, 2012 WL 5416941, at *1–2 (W.D. Wisc. Oct. 29, 2012); Microsoft Corp. v. Motorola, Inc., 864 F. Supp. 2d 1023, 1033–36 (W.D. Wash. 2012).

84. *See Microsoft*, 864 F. Supp. 2d at 1036–39.

negotiations and litigation in ways that may give parties pause before entering into candid pre-litigation license negotiations.

Our response to this debate is simple: who cares? The fight over breach of the FRAND commitment occurs because implementers have claimed that a standard-essential patent holder has failed to make a FRAND offer, and thus is in breach and cannot seek an injunction, perhaps due to patent misuse. Patentees counter by saying the implementer acts in bad faith and so loses the benefit of the license. But if injunctions are generally off the table, this argument becomes moot. Under our proposal, if an implementer thinks an offer is not FRAND, the implementer can just say "no." If the standard-essential patent holder does not want to budge, they go to arbitration to determine what is FRAND. There is no need to have a preliminary assessment of whether the offer was FRAND or close enough to FRAND.

Several other legal concerns also melt away under our proposal. If the parties have a mechanism in place to determine a reasonable royalty and bind the parties to that judgment, we no longer need to litigate whether the patentee's commitment creates an implied license, promissory estoppel, or equitable estoppel, and whether it is intended to benefit third parties. Our approach also sweeps away all the discussion of whether an offer was "so unreasonable" as to constitute a breach of the FRAND commitment.[85] This includes not only seeking "excessive royalties" but also "coercing" a license to non-standard-essential patents in exchange for standard-essential patents, or simply seeking an injunction.[86] All this is moot. The only question courts need to ask is whether each party agreed to participate in the arbitration and be bound by the results. If so, it doesn't matter how reasonable or unreasonable their negotiating position was.

B. DECLARATORY JUDGMENTS

Implementers sometimes seek declaratory judgments regarding what is FRAND or finding that an offer was not FRAND.[87] This becomes unnecessary under our proposed regime. If you want a decision on what a FRAND royalty rate is, go to arbitration.

As noted, *supra*, while a court can and will compel a patentee to arbitrate, there is no legal way to compel an implementer who wants to resolve the question in court to go to arbitration if they refuse.[88] But what SSOs can and

85. *Id.* at 1038.

86. *Id.*

87. *See, e.g.*, Complaint at 21, Microsoft Corp. v. Motorola, Inc., No. C10-1823JLR, (W.D. Wash. 2013), 2010 WL 4466798.

88. *Supra* Section I.C.1.

should do in that situation is release the patentee from the FRAND commitment not to seek an injunction. An implementer who wants to challenge the validity or infringement of the patents is free to do so in court, but if they are not willing to pay a FRAND royalty, they shouldn't benefit from the patentee's forbearance in seeking damages and injunctive relief. That won't make declaratory judgments of invalidity impossible, but should make them rare.

Implementers may try a different strategy: agree to arbitration, pay the royalty set by the arbitrator, and then file suit later to try to invalidate the patent. Under *MedImmune v. Genentech*, licensees have the power to go to court to challenge the validity of a patent even if they promised in the license agreement they wouldn't.[89] In effect, they can take advantage of the patentee's license offer and still challenge the validity of the patent without fear of consequences—at least, consequences for that license deal.[90] The same is true today with standard-essential patents: an implementer that takes a license to a standard-essential patent is free to challenge the validity of that patent by declaratory judgment in court, or to seek reexamination of the patent in the PTO.[91]

Reasonable minds can differ on the wisdom of this rule. It encourages challenges to validity, which are public goods that might otherwise be under-supplied by settlements.[92] But the rule also makes it harder to achieve finality. In any event, two things may make such challenges less important under our approach. First, the FRAND rate negotiated or set by arbitration applies to all of the essential patents owned by an SSO participant. A declaration that a particular patent is invalid may not benefit an implementer much if the rest of the portfolio remains untouched. Second, under our approach, the arbitrator's award is based on an ex ante probabilistic assessment of the

89. *See* MedImmune, Inc. v. Genentech, Inc., 549 U.S. 118, 113, 137 (2007).

90. For discussion of ways patentees may be able to punish efforts by licensees to challenge patents, see Michael Risch, *Patent Challenges and Royalty Inflation*, 85 IND. L.J. 1003 (2010). For limits on no-challenge clauses, see, for example, Rates Technology Inc. v. Speakeasy, Inc., 685 F.3d 163 (2d Cir. 2012).

91. *See MedImmune*, 549 U.S. at 133, 137; *In re* Baxter Int'l, Inc., 678 F.3d 1357, 1360, 1364–65 (Fed. Cir. 2012).

92. *See, e.g.*, Christopher A. Cotropia, *Modernizing Patent Law's Inequitable Conduct Doctrine*, 24 BERKELEY TECH. L.J. 723, 752–53 (2009); Joseph Farrell & Robert P. Merges, *Incentives to Challenge and Defend Patents: Why Litigation Won't Reliably Fix Patent Office Errors and Why Administrative Patent Review Might Help*, 19 BERKELEY TECH. L.J. 943, 952 (2004); Joseph Scott Miller, *Building a Better Bounty: Litigation-Stage Rewards for Defeating Patents*, 19 BERKELEY TECH. L.J. 667, 687–88 (2004); John R. Thomas, *Collusion and Collective Action in the Patent System: A Proposal for Patent Bounties*, 2001 U. ILL. L. REV. 305, 317.

entire portfolio. That ex ante assessment necessarily assumes that some patents in the portfolio may be invalid or not infringed. So there is no reason that a subsequent finding of invalidity should change the reasonable royalty determination. Some SSOs may want to make this point explicitly in their IP rules by stating that FRAND rates established through arbitration remain in force regardless of the outcome (either way) of subsequent validity challenges to a subset of the patents in the portfolio.

Second, while *MedImmune* held that licensees could challenge validity,[93] the same does not hold for defendants in a patent infringement suit who settle the suit by taking a license. Once a case is at issue, a settlement resolves the dispute with prejudice and the parties cannot reopen it.[94] Suing a defendant and then settling, in other words, offers patentees a way around *MedImmune*. The same may be true of disputes resolved by arbitration, though the courts have not resolved this issue. If so, resolution of a legal dispute over the FRAND commitment by arbitration will bar further challenges to the validity of the patent, at least as they relate to that implementer practicing that patent to implement the standard. An implementer will always be able to seek ex parte reexamination of the patent, though not inter partes reexamination or post-grant opposition.[95] But if they participate in the FRAND arbitration process, they are likely to be bound by the results.

C. CONFLICT BETWEEN JURISDICTIONS

Different jurisdictions have very different rules regarding injunctions. Germany and the U.S. International Trade Commission, for instance, grant injunctions on different terms than do the U.S. courts after *eBay*.[96] As a result, some patentees have turned to courts that will automatically grant injunctions, even on FRAND-encumbered standard-essential patents.[97] Implementers in turn have sought to prevent resort to these courts by a variety of means, including antitrust law and anti-suit injunctions.[98] This competition between jurisdictions not only creates comity concerns, but also

93. *MedImmune*, 549 U.S. at 118.

94. *See* Pactiv Corp. v. Dow Chem. Co., 449 F.3d 1227, 1230–32 (Fed. Cir. 2006); *see also* Aspex Eyewear, Inc. v. Marchon Eyewear, Inc., 672 F.3d 1335, 1345–46 (Fed. Cir. 2012).

95. 35 U.S.C. §§ 317, 329 (2012) (establishing that participants in inter partes reexamination and post-grant opposition must not have litigated the validity of the same patent before).

96. eBay Inc. v. MercExchange LLC, 547 U.S. 388 (2006); Chien & Lemley, *supra* note 20.

97. *See, e.g.*, Microsoft Corp. v. Motorola, Inc., 696 F.3d 872, 879–80 (9th Cir. 2012) (describing Motorola's suit against Microsoft in Germany for patent infringement in which Motorola sought—and was granted—an injunction).

98. *See, e.g.*, *id.* at 880–81; Motorola Mobility LLC, No. 121-0120, 2013 WL 124100 (F.T.C. Jan. 3, 2013).

raises questions about whether the patentee has impliedly licensed the implementer's use, and how different jurisdictions treat implied licenses. All this becomes moot if the SSO rules are clarified as we propose. A patentee who has made a FRAND commitment has agreed to resolve all disputes via arbitration. Under U.S. law, that commitment can be enforced by sending the dispute to arbitration, regardless of where it arises.[99]

D. ANTITRUST

Another significant benefit of our approach is that it significantly reduces the need for antitrust litigation to effectuate meaningful FRAND commitments. Courts, commentators, and regulatory agencies have devoted quite a bit of attention to antitrust scrutiny of standard-essential patents over the past decade.[100] Some of these cases have argued that seeking an injunction after making a FRAND commitment or failing to disclose a standard-essential patent constitutes monopolization (or attempted monopolization).[101] Others have argued that SSOs create an unlawful buyer's cartel by conspiring to fix royalty rates on a patent.[102]

Antitrust has an important role to play where SSOs do not set clear rules or set rules that can readily be gamed. Patentees who game those less-desirable rules by hiding information from the SSO,[103] making unreasonable

99. Not all jurisdictions may have the same rules regarding arbitration. We think the U.S. law is particularly suitable to our approach, but we are not experts in foreign law. To ensure uniformity in implementing our approach, SSOs should specify the law that applies to the enforcement of the arbitration agreement. If they don't and if a court in a foreign jurisdiction ignores the FRAND commitment and allows a suit for injunctive relief to proceed, the most the SSO can do is permit the implementer to engage in defensive suspension of the FRAND commitment with respect to its own patents.

100. *See, e.g.*, *Microsoft*, 696 F.3d 872; Rambus Corp. v. FTC, 522 F.3d 456 (D.C. Cir. 2008); Broadcom Corp. v. Qualcomm, Inc., 501 F.3d 297 (3d Cir. 2007); Apple, Inc. v. Motorola Mobility, No. 11-CV-178-BBC, 2012 WL 5416941 (W.D. Wisc. Oct. 29, 2012); Apple, Inc. v. Samsung Elec. Co., No. 11-CV-01846-LHK, 2012 WL 2571719 (N.D. Cal. June 30, 2012); Apple, Inc. v. Motorola Inc., No. 1:11-CV-08540, 2012 WL 1959560 (N.D. Ill. May 22, 2012); Microsoft Corp. v. Motorola, Inc., 864 F. Supp. 2d 1023 (W.D. Wash. 2012); Motorola Mobility LLC, No. 121-0120, 2013 WL 124100 (F.T.C. Jan. 3, 2013); Statement of the Federal Trade Commission, Robert Bosch GmbH, No. 121-0081 (F.T.C. 2013), *available at* http://www.ftc.gov/os/caselist/1210081/121126boschcommissionstatement.pdf; Negotiated Data Solutions LLC (*N-Data*), No. C-4234, 2008 WL 4407246 (F.T.C. Sept. 22, 2008); Dell Computers, Inc., 121 F.T.C. 616 (1996). The leading treatise on IP and antitrust devotes an entire chapter to the issue. HOVENKAMP ET AL., *supra* note 41, ch. 35.

101. *Rambus Corp.*, 522 F.3d 456; *Broadcom Corp.*, 501 F.3d 297; *Apple v. Samsung*, 2012 WL 2571719.

102. *See, e.g.*, HOVENKAMP ET AL., *supra* note 41, § 35.6 (collecting cases).

103. *Rambus Corp.*, 522 F.3d 456.

demands,[104] or trying to avoid a FRAND commitment by selling the patents may face antitrust liability or other forms of government regulatory scrutiny.[105] The SSO itself or its members may even face Sherman Act Section 1 antitrust exposure if the SSO rules are so vague as to facilitate patent holdup or a buyer's cartel.[106] Even if the standard overall is pro-competitive (that is, it generates consumer benefits), under a "least restrictive alternative" approach an SSO and its members may not be immune from antitrust scrutiny if the rules are significantly flawed in a way that creates market power for some of the members.

Under our approach, many of these issues should become moot, since the patentee cannot obtain an injunction (or transfer the patent to someone who can) against a willing licensee, and since competitors are not involved in jointly setting the reasonable royalty rate. If SSOs set clear, reasonable rules following the best practices we recommend, and parties follow those rules, there should be little or no need for antitrust to intervene. Indeed, even the risk of non-disclosure of a patent is lessened, since the patentee has committed to license its essential patents whether or not it discloses them. For the most part, the rules we have described are self-executing, meaning that even if a party tries to break the rules set by the SSO there still may be no need for antitrust to intervene. Thus, we suggest that parties who abide by these procedures—patentees, implementers, and the SSOs themselves—should be immune from antitrust liability for activities that merely follow those rules.[107] They have entered into an arrangement that is on balance good for competition, one that allows patentees to receive reasonable royalties but prevents holdup and reduces the risk of monopolization by trickery.

The fact that antitrust remains a last resort available when SSOs don't follow best practices may have two practical benefits, however. First, under our approach the promise of avoiding the risk of antitrust liability will be a powerful incentive for both SSOs and patent owners to adopt the best practices we propose. Second, the risk of antitrust liability may be relevant

104. *Broadcom Corp.*, 501 F.3d 297.

105. *N-Data*, 2008 WL 4407246. *N-Data* was brought under the FTC's section 5 authority, which applies where an unfair trade practice harms competition without necessarily rising to the level of an antitrust violation. Notably, only the FTC, not private parties, can enforce section 5. 15 U.S.C. § 45 (2012).

106. *See* U.S. DEP'T OF JUSTICE & FED. TRADE COMM'N, ANTITRUST ENFORCEMENT AND INTELLECTUAL PROPERTY RIGHTS: PROMOTING INNOVATION AND COMPETITION, *supra* note 1, at 53–56.

107. *See* Thomas F. Cotter, *Reining in Remedies in Patent Litigation: Three (Increasingly Immodest) Proposals*, 30 SANTA CLARA COMPUTER & HIGH TECH. L.J. (manuscript at 4) (forthcoming 2013) (arguing that FRAND royalties are properly resolved as a matter of patent, not antitrust law).

when an individual patentee wants to adopt best practices but the SSO governing the standard has not yet done so. We propose that a patentee that *unilaterally* commits to the FRAND procedures we describe here should be immune from antitrust liability for following these procedures.[108] A patentee's unilateral binding commitment to arbitration could be enforced whether or not it was elicited by an SSO. Thus, just as the prospect of antitrust immunity might lure SSOs to adopt best practices, it might also lure patentees to implement those practices even if the SSO has not done so. Given the large number of standard-essential patents based on preexisting standards,[109] and given that SSOs tend to update their IP rules rather slowly,[110] this is not a small matter.

III. CONCLUSION

Most of the litigation and debate over the FRAND commitment is unnecessary. SSOs can and should adopt best practices that will prevent patentee holdup while ensuring that the question of the appropriate royalty is resolved in a fair and predictable way. True, there will still be hard questions to face, notably determining the appropriate royalty in the face of the complexities of modern technology. But we think the FRAND commitment should be understood to create a simple, binding commitment that allows the patent holders and willing licensees to resolve those difficult questions through baseball-style binding arbitration if they cannot come to terms on their own. Rarely will they need to go to court.

108. A patentee that acts unilaterally may reasonably worry more about reciprocity than those who act under the aegis of an SSO. Thus, a broader defensive suspension clause might be justifiable in this circumstance.

109. *See, e.g.*, Lemley & Shapiro, *supra* note 14 (documenting thousands of patents declared essential to just two standards).

110. *Compare* Contreras, *supra* note 5, *with* Lemley, *SSOs*, *supra* note 3 (studying SSO policies ten years apart).

"GENTLY DOWN THE STREAM": WHEN IS AN ONLINE PERFORMANCE PUBLIC UNDER COPYRIGHT?

Daniel Brenner[†]

---

† Judge, Superior Court, Los Angeles County; Adjunct Professor, University of Southern California Gould School of Law. Formerly Partner, Hogan Lovells LLP. The Author wishes to thank Professor Paul Goldstein, Stanford Law School, and Rose Perez, Esq. for helpful suggestions to this Article.

I. INTRODUCTION

The "public performance" right in copyright law may lack the eponymous primacy of the right to copy, but it is a close second. Creators of literary works, sound recordings, and audio-visual productions rely on the right to copy as a significant inducement to create. But the latter two groups, along with playwrights and specifically composers, look to public performances of their works as a major source of their economic reward. Online audio-visual works present a challenge for courts tasked with defining the scope of the public performance right. The economic implications are significant for ASCAP, BMI, and SESAC[1]—the nation's three performance rights organizations ("PROs"), which collect royalties on behalf of member composers (and their estates)—as well as owners of audio-visual works, when the distributor is not under license.[2]

This Article concludes that the language of the Copyright Act supports a limited online public performance right. It argues that it is wrong to conclude every online performance is public, however.[3] Instead, to claim this right, the copyright owner (or her PRO) must establish that the work is performed for a substantial number of online viewers and the viewing is not a substitute for an already compensated performance. This Article also explores why PROs today collect millions of dollars for public performances of audio-visual works and concludes that this role is hardly inevitable as is the case with other performances; however, eliminating PRO audio-visual performance collections would be very difficult.

Establishing the public performance right for new media has been both essential and difficult. Motion pictures would have remained an experimental art form without creation of the right to collect when a film is performed for the public.[4] Broadcasters twice sought to declare cable television

1. American Society of Composers, Authors and Publishers; Broadcast Music, Inc.; and Society of European Stage Authors and Composers (although BMI and SESAC's company names were once acronyms, today they are not abbreviations of anything). *About ASCAP*, ASCAP, http://www.ascap.com/about/ (last visited Sept. 30, 2013); *About*, BMI, http://www.bmi.com/about (last visited Sept. 30, 2013); *About SESAC*, SESAC, http://www.sesac.com/About.aspx (last visited Sept. 30, 2013).

2. Owners of audio-visual works suffer when copies are downloaded from unauthorized websites as well as when the works are viewed on a streaming basis from websites where the content has been posted without authorization. The downloading itself does not constitute a performance, however.

3. *See* Part V, *infra*.

4. On August 24, 1912, motion pictures, previously registered as photographs, were added to the class of protected works in the 1909 Copyright Act. Frank Elvina, *Copyright Lore*, *in* COPYRIGHT NOTICES (Oct. 2004), *available at* http://copyright.gov/history/lore/2004/oct04-lore.pdf.

transmissions of their shows public performances and were twice rebuffed by the U.S. Supreme Court.[5] The Copyright Act of 1976 created a compulsory licensing scheme for cable to perform secondary transmissions, allowing both cable operators and TV program owners to prosper.[6] Sound recording owners, denied a royalty for performance of their phonorecords by analog AM and FM radio stations, changed the calculus by obtaining a performance right when digital, largely Internet, radio was introduced.[7]

When a composer's song is performed on a broadcast or cable program, its owner, typically its publisher, collects twice: first when the derivative audio-visual work is created by synchronizing the picture to the music, second when the audio-visual work is publicly performed as a broadcast or cablecast transmission. The first use is directly licensed from the publisher of the composition or through a rights agent.[8] The latter use is typically paid to PROs under a blanket license issued to networks, stations, or local cable operators, with the money pooled and distributed to composers. The PROs collect and pay based on surveys, cue sheets, and internally-developed algorithms.[9]

It is settled that these "linear" transmissions—that is, programs viewed as they are transmitted to all potential viewers at the same time—of TV programs are public performances.[10] And it is equally clear, if not explicitly

5. *See* Teleprompter Corp. v. CBS Inc., 415 U.S. 394, 413–14 n.14 (1974) (finding active importation of a distant signal not a performance); Fortnightly Corp. v. United Artists Television, Inc., 392 U.S. 390, 399–400 (1968) (deeming passive carriage of a retransmitted broadcast signal not a performance).

6. Pub. L. 94-553, tit. I, § 101, 90 Stat. 2541 (codified at 17 U.S.C. § 111(c) (2012)).

7. The Digital Performance Right in Sound Recordings Act of 1995 grants owners of a copyright in sound recordings an exclusive right "to perform the copyrighted work publicly by means of a digital audio transmission." 17 U.S.C. §§ 106, 114–115 (2012). Setting the right rate for online radio services like Pandora has been highly controversial. *See* Katy Bachman, *Lawmakers Ponder Disparity in Internet Radio Fairness Act; Music Performance Rates a Pandora's Box*, http://www.adweek.com/news/technology/lawmakers-ponder-disparity-internet-radio-fairness-act-145488.

8. MARK S. LEE, ENTERTAINMENT AND INTELLECTUAL PROPERTY LAW, § 7:47 (2013).

9. *See, e.g.*, *ASCAP Payment System: Introduction*, THE AMERICAN SOCIETY OF COMPOSERS, AUTHORS AND PUBLISHERS, http://www.ascap.com/members/payment/ (last visited Feb. 2, 2013) (discussing ASCAP's payment system).

10. Early on television broadcasters accepted that an ASCAP license was needed. The Television Music License Committee, LLC, which handles PRO negotiations for TV stations, summarized this history:

> The first ASCAP television licenses were negotiated in the 1940s. ASCAP initially offered free licenses to television broadcasters. In 1948, ASCAP notified the broadcasters that it was terminating the free licenses and the National Association of Broadcasters (NAB) formed a separate committee to negotiate music licenses for television stations. In 1949, the

decided in case law, that playing back a VCR, DVD, or DVR home recording of that same program to a small group of friends or family is not a "public" performance. Routinely, however, viewers watch linear video programs on-demand via a cable operator set-top box or online over the Internet at a site like Hulu or YouTube. The copyright statute does not provide an unequivocal right for composers to collect under the public performance right in these instances. But the definition of "To perform or display a work 'publicly' " in section 101 of the Copyright Act[11] provides a statutory basis to claim a right to collect. Congress can, of course, redefine the public performance right, as it did for broadcasters (in reference to cable retransmissions) in the 1976 statutory rewrite and for record companies in the 1995 Digital Performance Right in Sound Recordings Act.[12] But it probably should not. Given the political challenges to changing such fundamental terms like "publicly perform," it is unlikely that a statutory change would cleanly resolve the issue. In sum, PROs, other rights holders, and licensees should recognize that online public performances exist, but their scope may be more or less broad than either side to this debate may believe.

II. ONLINE DISTRIBUTION AND PERFORMANCE OF AUDIO-VISUAL WORKS

Evolution of Internet availability of audio and audio-visual works has been rapid. Early downloads from unauthorized central file servers like Napster helped propel music use on the Internet. File sharing facilitators like Grokster, which connected two users but did not store content itself, coincided with the rollout of broadband and its wider capacity in the early 2000s.[13] Broadband capacity facilitates video file sharing and thus ensnared

> parties reached agreement on an ASCAP blanket fee of "radio plus 10." This license fee mirrored the radio percentage of revenue license at 2.25% and also included a 10% surcharge. This was the beginning of a long, contentious and litigious relationship between television stations and the music licensing organizations.

History, TELEVISION MUSIC LICENSING COMMITTEE, http://www.televisionmusic.com/Joomla_1.5.15/index.php?option=com_content&view=article&id=4 (last visited Mar. 16, 2013).

11. 17 U.S.C. § 101 (2012).

12. *See* Copyright Act of 1976 § 101, Pub. L. No. 94-553, 90 Stat. 2541 (codified as amended at 17 U.S.C. § 101 (2012)); Digital Performance Right in Sound Recordings Act of 1995, Pub. L. No. 104-39, § 2 (1995) (codified as amended at 17 U.S.C. § 106(6) (2012)).

13. Cable broadband and DSL offered much faster speeds than dial-up Internet service. Downloading and file sharing, legal and illegal, require sufficiently fast upload and download speeds to be practical. *See* DANIEL L. BRENNER, MONROE E. PRICE & MICHAEL I. MEYERSON, CABLE TELEVISION AND OTHER NONBROADCAST VIDEO § 18:5 (2013).

the motion picture and television industry into the piracy problems faced first by the recording industry.

Courts found both Napster and Grokster—which helped to launch debilitating worldwide online piracy—to be infringing services under the Copyright Act,[14] and legitimate downloading sources like iTunes emerged.[15] But the desire to use broadband for free video downloads continued, initially with the near-overnight success of YouTube. While YouTube reformed its early "post-nearly-anything" policy, it still hosts hundreds of millions of videos that contain unlicensed copyrighted material (including musical compositions that are likewise unlicensed for either reproduction or public performance purposes) along with material that is authorized by its owners, at least for purposes of reproduction.[16] It obtained a court-determined public performance license rate from ASCAP in 2009,[17] following licenses sought and obtained by Yahoo! and others.[18]

In the YouTube, Yahoo!, and related cases, the applicant websites sought whatever public performance license was required. The parties and the ASCAP rate court, which sets the rates in the absence of an agreed-to rate, assumed without analysis that on-demand streams of music videos—the primary audio-visual works in the cases—were being "publicly" performed. In particular, the court assumed that licenses were required by relying on the

14. *See* A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004 (9th Cir. 2001); MGM Studios, Inc. v. Grokster, Ltd., 545 U.S. 913 (2005).

15. The iTunes Store, originally the iTunes Music Store, opened on April 28, 2003. Press Release, Apple Inc., Apple Launches the iTunes Music Store (Apr. 28, 2003), *available at* http://www.apple.com/pr/library/2003/04/28Apple-Launches-the-iTunes-Music-Store.html.

16. YouTube originally allowed longer videos to be posted but in 2006 reduced the length to ten minutes to prevent TV episodes and other longer-form video from being illegally posted. Ken Fisher, *YouTube Caps Video Lengths to Reduce Infringement*, ARS TECHNICA (Mar. 29, 2006), http://arstechnica.com/uncategorized/2006/03/6481-2. Current policy allows up to fifteen minutes of uploaded material unless a subscriber is qualified for longer uploads. *Frequently Asked Questions*, YOUTUBE, http://www.youtube.com/t/faq (last visited Mar. 17, 2013) ("why can't I upload videos longer than 15 minutes?"). One licensor explained its view of the situation this way: "[T]he majority of our music appearing in YouTube videos is actually unlicensed. Like many content owners, instead of attempting to block these unauthorized videos or otherwise pursue these users through legal channels, we simply opt to monetize the videos by having ads placed around them." Ron Mendelsohn, *A Look Inside YouTube: Skateboarding Cats, Talking Dogs and Content ID*, MEGATRAX (Oct. 7, 2012), http://www.megatraxblog.com/2012/10/08/a-look-inside-youtube-skateboarding-cats-talking-dogs-and-content-id/.

17. United States v. ASCAP (*In re* YouTube), 616 F. Supp. 2d 447 (S.D.N.Y. 2009).

18. United States v. ASCAP (*In re* Am. Online, Inc., RealNetworks, Inc., and Yahoo! Inc.) (*RealNetworks and Yahoo! II*), 559 F. Supp. 2d 332 (S.D.N.Y. 2008), *rev'd in part sub nom.* United States v. ASCAP (*In re* RealNetworks, Inc., Yahoo! Inc.), 627 F.3d 64 (2d Cir. 2010), *cert denied*, 132 S. Ct. 366 (2011).

websites' voluntary requests for public performance licenses from the PROs.[19] But in other instances, parties raised the basic question of whether a license was even necessary. For instance, the rate court found that a Yahoo! and Real Networks user who merely downloads songs to an MP3 player did not publicly perform the music.[20]

Cable and broadcast television providers have added to the online mix. Local stations may offer a real-time or delayed online feed of their free-to-air programming, such as locally produced news shows.[21] They may also produce web-only versions of their news, sports, or weather coverage. Starting in 2007, owners of other programming, including popular broadcast and cable network shows, made their content available on the web on a host of sites. Fox and NBC launched Hulu, a service to capture online viewers of their programs (some of which were being posted to YouTube, at a time the networks could not monetize YouTube performances), in 2008.[22] Disney joined the venture in 2009 after making some of its network programs available on its own website earlier.[23] Some cable networks offer their programming through Hulu or through sites created by their cable distributors, such as Comcast[24] and Cox.[25]

There are two primary motivations for these video sites. First, it is generally acknowledged (although not proven) that the recording industry

19. *In re YouTube*, 616 F. Supp. 2d at 448 ("On September 25, 2006, YouTube applied to ASCAP for a blanket through-to-the-listener license for a two-year period commencing in May 2004"); *RealNetworks and Yahoo! II*, 559 F. Supp. 2d at 343.

20. United States v. ASCAP (*In re* Am. Online, Inc., RealNetworks, Inc., and Yahoo! Inc.) (*RealNetworks and Yahoo! I*), 485 F. Supp. 2d 438, 446 (S.D.N.Y. 2007), *aff'd sub nom.* United States v. ASCAP (*In re* RealNetworks, Inc., Yahoo! Inc.), 627 F.3d 64 (2d Cir. 2010).

21. Segments are often available on a TV station's "video" tab. *See, e.g.*, *Video*, CBS LOS ANGELES, http://losangeles.cbslocal.com/video/ (last visited Mar. 17, 2013) (providing the local weather reporter's forecast for Los Angeles and other video clips from KCBS's news department).

22. *Company Timeline*, HULU, http://www.hulu.com/about/company_timeline (last visited Apr. 24, 2013).

23. *Id.*; *Disney Joins Hulu Video Site*, BILLBOARD.BIZ, http://www.billboard.com/biz/articles/news/global/1271122/disney-joins-hulu-video-site (last visited Apr. 24, 2013) ("Disney has previously sought to expand viewership of ABC shows offered on its Web site and its local affiliates' sites, on AOL.com and on Comcast Corp's Fancast site.").

24. Comcast offered video content through its Fancast website which was incorporated into its Xfinity brand. *See* Ryan Kim, *Comcast Launches Fancast XFINITY for TV Everywhere*, *The Tech Chronicles*, SAN FRANCISCO CHRONICLE, Dec. 15, 2009, http://blog.sfgate.com/techchron/2009/12/15/comcast-launches-fancast-xfinity-for-tv-everywhere/.

25. The service is marketed as Cox TV Online. *Internet Support*, COX, http://ww2.cox.com/residential/centralflorida/support/internet/article.cox?articleId=ad1297d0-088e-11e0-7ab7-000000000000#typeContent (last visited Apr. 22, 2013).

realized the impact of the Internet too late.[26] Some claim record companies should have developed a legitimate capability for single-song downloads once they perceived the customer demand that Napster and its progeny had identified instead of emphasizing their right to pursue and punish copyright violators.[27] By the early 2000s a customer-friendly "Internet strategy" for video content was imperative to avoid sending viewers into the arms of pirates.[28]

There is a second reason for some free online video: broadcasters often give away their content for free over the air in exchange for advertising (the recording industry does, too, but a listener cannot typically record songs or albums from a radio broadcast unless the playlist is disclosed in advance). Free-to-air viewers constitute only a small portion of a broadcaster's audience; nearly ninety percent of households get broadcast channels from cable or satellite operators who must obtain retransmission consent in order to carry the signals of the most valuable TV stations.[29] Since the underlying off-air audio-visual works generally have been made available free to the public, facilitating their viewing adds to the value of a free-to-air program, so long as the online viewing can be measured for advertisers.

While retransmission payments are growing for some broadcasters, advertising still constitutes the lion's share of revenues associated with broadcasting.[30] From that perspective, whether a viewer sees a show in its first run, as a repeat, on an affiliated cable channel owned by the broadcast network,[31] or online, a broadcaster revenue model seeks to maximize the

26. *See* Eric Pfanner, *Music Industry Sales Rise, and Digital Revenue Gets the Credit*, N.Y. TIMES, Feb. 26, 2013, http://www.nytimes.com/2013/02/27/technology/music-industry-records-first-revenue-increase-since-1999.html?_r=0 (showing that the music industry is posting stronger profits after the rise of digital music services).

27. *See id.*

28. *See id.*

29. *Nielsen: Broadcast-only TV Households to Slip Below 10 Percent*, BROADCAST ENGINEERING (May 4, 2010), http://broadcastengineering.com/hdtv/nielsen-broadcast-only-tv-households-slip-below-10-percent (stating that by 2010 fewer than ten percent of viewers watched TV over the air).

30. For example, LIN TV Corp., a multimedia company serving twenty-three U.S. markets, projected cash payments for retransmission consent for 2012 Q3 at $5.5–6 million versus net advertising revenues between $110.5–116 million. *LIN TV Corp. Announces Second Quarter 2012 Results*, WISHTV.COM (Jul. 31, 2012, 4:55 PM), http://www.wishtv.com/dpp/about_us/lin-media-announces-2nd-quarter-results.

31. ABC, NBC, and Fox have extensive cable network holdings. NBC can offer a show on its broadcast network, run a second play on USA Network or CNBC, and garner advertising support on each showing. NBC used commonly owned NBC Sports Network, Bravo, CNBC, MSNBC, and Telemundo (vertically owned by cable operator Comcast) to present different or replayed 2012 Olympics coverage. Rene Lynch, *Olympics 2012: Opening*

number of viewers seeing the embedded advertising. So online viewing of broadcast TV shows avoids the negatives associated with online music—copying—so long as viewership can be counted.

Cable networks, too, want to maximize their advertisers' audience but the calculation is more complicated. This is because many cable program networks get roughly half of their revenues from program license fees from cable and Direct Broadcast Satellite ("DBS") distributors.[32] Free online distribution can harm those distributors by making the viewer-pay model a less compelling proposition. Posting marquee programming for free on the Internet reduces the value of programming to a cable operator who has paid for such programming based on a subscription-only assumption. If subscribers can get their favorite shows on the Internet for free, some "cord-cutters" will drop, and have dropped, their cable subscriptions.[33] Going forward, cable or DBS distributors do not want to necessarily dictate what model programmers choose, but they do expect to pay lower fees for programming if it is also made available for free online.

These distributors want to have an Internet strategy, too. Cable operators and programmers are convinced that allowing customers to view programming "whenever, wherever"—in real time, on video-on-demand, via DVRs, on the Internet in the home, or by mobile device—is their only sustainable distribution strategy. The goal is the ability to authenticate, measure, and eventually tailor advertising to viewers for online and mobile viewing. Denying Internet availability is not considered a viable long-term strategy.[34] If unauthorized uploaders can post content to YouTube or other

Ceremony—When It Starts, What to Expect, LA TIMES, July 27, 2012, http://articles.latimes.com/2012/jul/27/nation/la-na-nn-olympics-2012-opening-ceremony-20120727. NBC also repurposed *Monk*, a USA Network show, for NBC in 2007. Edward Wyatt, *NBC To Repurpose USA's 'Monk' and 'Psych'*, N.Y. TIMES MEDIA DECODER BLOG (Dec. 18, 2007, 3:11 PM), http://mediadecoder.blogs.nytimes.com/2007/12/18/nbc-to-repurpose-usas-monk-and-psych/.

32. *The State of the News Media 2013: Cable-Glossary*, THE PEW RESEARCH CENTER'S PROJECT FOR EXCELLENCE IN JOURNALISM, http://stateofthemedia.org/2013/cable-a-growing-medium-reaching-its-ceiling/cable-glossary/ (last visited May 20, 2013).

33. "Cord-cutting," which refers to cable or DBS customers who cancel their pay-TV subscriptions, is much discussed, yet accurate statistics on the extent of the practice do not exist because cable and DBS subscribers disconnect for many reasons besides online substitutes.

34. For example, Time Warner, which operates HBO, has promoted TV Everywhere:

> Time Warner is a leader in the next phase of the digital evolution of media: delivering content that consumers love to watch on any device and at any time. As a guiding principle, the company is aggressively pursuing initiatives that give audiences more choice and quality at no additional

sites anyway, it is better to make the content available on a site whose quality and advertising potential can be controlled by the cable operator, who has paid for a license to present the show.[35] Under this view, online distribution complements the pay-TV model rather than competes with it.

Moreover, the ways in which video is performed on the Internet are evolving. How a court characterizes performing methods has decisional significance, and cases often turn on how the court views the distribution technology.[36] PROs would prefer to characterize online viewing as equivalent to broadcast transmissions, long established as public performance regardless of actual proof of viewing or listening. Online providers who wish to avoid paying for music performances under a PRO license would benefit from likening their performances to home-based VCR or DVD plays, unequivocally not "public" and therefore requiring no license.

Realizing the contentiousness of such descriptions, it is useful, though not dispositive, to differentiate the ways that audio-visual works are "performed" on the Internet: downloads, multicast streaming, and unicast streaming (of which there are two varieties). First, an audio-visual work may be downloaded as a digital file from a server to a computer which hosts the file, and then to the client's devices, which receives a copy of the file during the download. Podcasts, ringtones, and iTunes are familiar examples of downloads. Second, real-time multicast streaming provides a performance of the work and does not result in creation of a permanent file on the

> cost to them—while also maintaining or enhancing the economic models of our businesses.

Content Everywhere, TIMEWARNER, http://www.timewarner.com/our-innovations/content-everywhere/ (last visited Mar. 16, 2013).

35. Cable operators are teaming up with networks to control who views content online. For its online viewing site HBO Go, subscription cable channel HBO requires a user to log in to the online account she holds with her cable provider before acquiring access to HBO's programming online, thereby ensuring only those viewers who actually pay for and receive HBO on their televisions can access the programming online. *See What Is HBO GO*, http://www.hbogo.com/ (last visited Mar. 16, 2013) (requiring login). This trend is spreading beyond cable subscription channels. Fox Broadcasting implemented an authentication system which requires a user's cable subscription information in order to watch shows the day after they air; otherwise, viewers without a cable subscriber have to wait eight days to watch programming online. Chloe Albanesius, *Fox Puts Online Content Behind Pay Wall*, PC MAGAZINE (July 27, 2011), http://www.pcmag.com/article2/0,2817,2389211,00.asp.

36. *Compare* Cartoon Network LP, LLLP v. CSC Holdings, Inc. (*Cablevision II*), 536 F.3d 121, 124–25 (2d Cir. 2008), *with* Twentieth Century Fox Film Corp. v. Cablevision Sys. Corp., 478 F. Supp. 2d 607, 610–16 (S.D.N.Y. 2007) (providing different descriptions of remote DVR technology).

customer's computer. It is most like broadcasting but is not commonly used.[37]

In contrast to multicast streaming, unicast streaming, the third form of online performance, establishes session-based one-to-one connections between a customer and the server.[38] For example, when the customer connects to, for example, the Hulu server, it creates a direct relationship that consumes bandwidth on (i) the server, (ii) the network-of-networks that constitute the Internet backbone (or content delivery network ("CDN") if the content is cached closer to the user), and (iii) the customer's internet service provider ("ISP") network. The server must dedicate specific bandwidth to that unicast. Unlike multicast, where the server sends out one stream to all users, if ten users seek a unicast at the same time, separate bandwidth will be needed for each unicast.[39]

This distinction between multicast and unicast would on its surface seem to differentiate Internet video. Watching the simulcast of a video channel (for example, CSPAN-2 on cspan.org) seems like multicast; viewing a YouTube video of the Kennedy-Nixon debate seems like unicast. In practice, however, the public Internet does not support multicasts. Multicasts are possible on a private network: sending a video out to all terminals on a local area network, for example. Instead, a website sets up a number (up to hundreds of thousands) of unicasts, costing a tenth of a cent or less, to serve

37. As Microsoft explains it:

> The multicast source relies on multicast-enabled routers to forward the packets to all client subnets that have clients listening. There is no direct relationship between the clients and Windows Media server. The Windows Media server generates an .nsc (NetShow channel) file when the multicast station is first created. Typically, the .nsc file is delivered to the client from a Web server. This file contains information that the Windows Media Player needs to listen for the multicast. This is similar to tuning into a station on a radio. Each client that listens to the multicast adds no additional overhead on the server. In fact, the server sends out only one stream per multicast station. The same load is experienced on the server whether only one client or 1,000 clients are listening.

Differences Between Multicast and Unicast, MICROSOFT, http://support.microsoft.com/kb/291786 (last updated Nov. 03, 2003) ("The multicast source relies on multicast-enabled routers to forward the packets to all client subnets that have clients listening. There is no direct relationship between the clients and Windows Media server.").

38. *Unicast and Multicast Streaming*, EYEPARTNER (Dec. 26, 2009), http://www.eyepartner.com/tutorials/unicast-and-multicast-streaming/.

39. Shahar Ze'evi, *Multicast Video Transmission vs. Unicast Video Transmission Methods*, AMERICAN DYNAMICS SECURITY BLOG (May 14, 2012, 10:10 AM), http://security.americandynamics.net/blog/bid/56070/Multicast-video-transmission-vs-Unicast-video-transmission-methods.

each customer separately.[40] Or it uses gridcasting technologies, similar to peer-to-peer, to utilize a user's bandwidth to relay the video, reducing the burden on the source server.[41]

The result is that all online video not previously downloaded to the user's device is viewed primarily in a unicast mode. As discussed below, the fact that each transmission is unicast—and therefore not multicast—is no basis to declare all online viewing as non-public.[42] An online video that simulcasts a broadcast or cablecast program is as "public" a performance as the version carried on a traditional video network. The second type of unicast, programming watched on-demand, presents the more difficult problem of when its performance by a website should be deemed "public."

As noted, the methods of distributing video over the Internet are evolving, and all of them have emerged since the 1976 Copyright Act defined public performance. We now turn to the text of the statute to see how its language draws distinctions among these different forms of performance.

III. TO PERFORM A WORK "PUBLICLY": THE COPYRIGHT ACT'S "PUBLIC PLACE" AND "TRANSMIT" CLAUSES

The Copyright Act grants a copyright owner the exclusive right "in the case of literary, musical, dramatic, and choreographic works, pantomimes, and motion pictures and other audiovisual works, to perform the copyrighted work publicly."[43] As the law surely was not meant to require every whistled tune to be licensed, the focus of the statute is determining when a work is "publicly" performed.

---

40. Telephone Interview with Steven Harkness, Internet Dir., C-Span (Aug. 10, 2009).

41. Gridcasting is a streaming system that uses idle bandwidth on a user's computer to deliver large-scale live or on-demand broadcasts. *See* Stephen Alstrup & Theis Rauhe, *Introducing Octoshape—A New Technology for Large-Scale Streaming over the Internet*, EUR. BROADCASTING UNION TECHNICAL REV. (July 2005), *available at* http://www.octoshape.com/press/pdf/papers/0507ebu.pdf. Gridcasting improves performance, scalability, and cost efficiency of delivering files and streams to end users through the use of a media plug-in. *See* National Association of Broadcasters, *Gridcasting*, RADIO TECHCHECK (Dec. 4, 2006), *available at* http://www.nab.org/xert/scitech/pdfs/rd120406.pdf. It is similar to BitTorrent. *See id.* Octoshape is one gridcasting system and has been used for large-scale broadcasts such as the 2009 Presidential Inauguration, which had over 1.3 million simultaneous viewers. *See* Janko Roettgers, *CNN: Inauguration P2P Stream a Success, Despite Backlash*, GIGAOM (Feb. 7, 2009, 12:01 AM) http://gigaom.com/2009/02/07/cnn-inauguration-p2p-stream-a-success-despite-backlash/.

42. *See* Section IV.C, *infra*. As discussed below, the *Cablevision* and *Aereo* decisions consider it significant that the copy that is being performed is created by that user and performed only for that user. In unicast technology, each transmission is separate but the copy on which the transmission is based is reused. *Id.*

43. 17 U.S.C. § 106(4) (2012).

> To perform or display a work "publicly" means—
>
> (1) to perform or display it at a place open to the public or at any place where a substantial number of persons outside of a normal circle of a family and its social acquaintances is gathered; or
>
> (2) to transmit or otherwise communicate a performance or display of the work to a place specified by clause (1) or to the public, by means of any device or process, whether the members of the public capable of receiving the performance or display receive it in the same place or in separate places and at the same time or at different times.[44]

The definition of "publicly" is in two parts.[45] The most pertinent section of the definition focuses on transmitting a performance of the work to the public.[46] This is known in the case law as the "Transmit Clause." It constitutes the second, and less intuitive, delineation of a public performance. The first definition, in the "Public Place Clause," refers to a physical event. It defines "public" performance of a work to be "at a place open to the public or any place where a substantial number of persons outside of a normal circle of a family and its social acquaintances is gathered."[47]

The paradigmatic performance covered by the Transmit Clause is a program carried on traditional over-the-air radio and television broadcasting. The Transmit Clause, added in the 1976 Copyright Act revision[48]—in part in response to *Fortnightly* and *Teleprompter*,[49] which held that cable broadcasting

44. 17 U.S.C. § 101 (2012).

45. *Id.*

46. *Id.*

47. *Id.* The 1976 Act reversed cases that held that a performance was not public if the audience was limited to a particular group rather than the public in general. *See* Metro-Goldwyn-Mayer Distrib. Corp. v. Wyatt, 21 Copyright Off. Bull. 203 (D. Md. 1932).

48. WNET, Thirteen v. AEREO, Inc. (*Aereo II*), 712 F.3d 676, 685, 694–95 (2d Cir. 2013) ("The legislative history shows that the Transmit Clause was intended in part to abrogate *Fortnightly* and *Teleprompter* and bring a cable television system's retransmission of broadcast television programming within the scope of the public performance right."). The public performance clauses were included in a 1967 predecessor bill, H.R. 2512, 90th Cong. (1967), which was an outgrowth of the 1965 general proposed revision to the 1909 Act. *See* STAFF OF H.R. COMM. ON THE JUDICIARY, 89TH CONG., COPYRIGHT LAW REVISION PART 6: SUPPLEMENTARY REGISTER'S REPORT ON THE GENERAL REVISION OF THE U.S. COPYRIGHT LAW: 1965 REVISION BILL at 23 Comm. Print (1965), *available at* http://ipmall.info/hosted_resources/lipa/copyrights/Supplementary%20Register's%20Report%20on%20the%20General%20Revision%20of.pdf. The 1967 House Report would have included transmissions to "the subscribers of a community antenna television service." H.R. REP. NO. 90-83, at 29 (1967).

49. *See supra* note 5 and accompanying text.

was not a public performance—incorporated judicial recognition as early as the 1920s that broadcasting to car radios or home sets was a public performance under the 1909 Copyright Act.[50] A broadcast performance is public even if no person is in fact operating receiving equipment at the time of the transmission.[51] This includes linear cable networks, whose audiences may be small or nonexistent on a particular system.[52] The 1976 Act also defined "public performance" liability as unrelated to whether the performance is for commercial or non-commercial purposes.[53]

Thus, while the Public Place Clause assumes a public gathering, the Transmit Clause does not. Members of the public who receive the transmission need not receive it in the same place. Unlike a traditional performance in public, a broadcast or cablecast performance will be in the home to groups of family or friends who would not otherwise constitute a "substantial" number of persons.[54] But added together, these television households often constitute a very substantial audience; indeed, mass media advertising is premised on reaching an accumulated audience.

Not only does a public performance under the Transmit Clause encompass performances at different locations; the definition allows the public capable of receiving the performance to "receive it at the same time or at different times."[55] The accompanying legislative history does not explain the legislature's intent in not requiring simultaneity.[56] One explanation could be that Congress thought to cover performances by devices like jukeboxes,

50. *See* Jerome H. Remick & Co. v. Am. Auto Accessories Co., 5 F.2d 411 (6th Cir. 1925) (finding that radio station WLW publicly performed music). The case pre-dates the formal licensing of radio stations. *See* Communications Act of 1934, Pub. L. No. 73-416, 48 Stat. 1064 (codified as amended at 47 U.S.C. § 151 et seq. (2012)).

51. H.R. REP. NO. 94-1476, at 64–65 (1976), *reprinted in* 1976 U.S.C.C.A.N. 5659, 5678.

52. *Id.* at 65 ("[A] performance made available by transmission to the public at large is 'public' . . . whenever the potential recipients of the transmission represent a limited segment of the public, such as . . . subscribers of a cable television service."). Indeed, the concept of switch digital video assumes some narrow-taste channels will be tuned in only occasionally and so their transmission occurs only when a subscriber chooses to watch it.

53. The 1909 Act based the definition of a public performance in part on whether the performance was "for profit." *Id.* at 62–63. Congress rejected this approach and rested the applicability of the Act solely on the question of whether a performance is private or public, with no consideration of whether it is commercial or non-commercial. *Id.*

54. Broadcasting to public places, such as restaurants, would qualify under either clause, but 17 U.S.C. § 110(5)(B) (2012) exempts musical broadcasts in public places of limited size or containing a limited number of speakers.

55. 17 U.S.C. § 101 (2012).

56. MELVILLE F. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 8.14[C][3] (2013) ("The Senate and House Reports offer no explanation of this . . . phrase, and it is difficult to believe that it was intended literally.").

where the same recording is performed for one or a few people at one time but the numerous performances should amount to public performance.[57]

This approach, focusing on the performance of the same copy of a work in a public place, accords with earlier case development. For instance, in *Columbia Pictures Industries v. Redd Horne Inc.*, a videocassette shop rented tapes to customers and provided private screening rooms.[58] No more than four persons, who had to be relatives or close acquaintances, could occupy the screening room at one time.[59] The court declared the private screening rooms the equivalent of a movie theater and the film performances public.[60] A similar situation, where the video store rented cassettes and rooms with video players in separate transactions (here accommodating up to twenty-five people in a room) was also deemed to be a public performance when video cassettes were played.[61] But in *Columbia Pictures Industries v. Professional Real Estate Investors, Inc.*, where a hotel rented videodiscs and hotel rooms furnished with do-it-yourself videodisc players, the court found no public performance.[62] Yet in the same circuit, a district court in *On Command Video Corp. v. Columbia Pictures Industries*[63] found a public performance where a hotel used a bank of video cassette players, with each VCR containing a copy of a particular movie. The single copy of the movie was transmitted electronically to a hotel guest's room upon demand via remote control from a list on the guest's TV.[64] Note that the *Professional Real Estate Investors, Inc.* case involved the Public Place Clause, whereas *On Command* interpreted the Transmit Clause.[65]

Attempting to apply these cases (which addressed essentially outdated types of video performances), the Second Circuit in *The Cartoon Network LP,*

57. *See id.* at 8-192.8(1) (citing H.R. REP. NO. 94-1476, at 114).
58. 749 F.2d 154 (3d Cir. 1984).
59. *Id.* at 157.
60. *Id.* at 159. In evaluating the liability of the defendant Maxwell's, the court stated:

> We find it unnecessary to examine the second part of the statutory definition because we agree with the district court's conclusion that Maxwell's was open to the public . . . . Any member of the public can view a motion picture by paying the appropriate fee. The services provided by Maxwell's are essentially the same as a movie theater, with the additional feature of privacy.

Id.

61. Columbia Pictures Indus. v. Aveco, Inc., 612 F. Supp. 315 (M.D. Pa. 1985), *aff'd*, 800 F.2d 59 (3d Cir. 1986); *accord*, Video Views, Inc. v. Studio 21, Ltd., 925 F.2d 1010 (7th Cir. 1991).
62. 866 F.2d 278 (9th Cir. 1989).
63. 777 F. Supp. 787 (N.D. Cal. 1991).
64. *Id.* at 788.
65. *See id.* at 789; *Columbia Pictures Inds.*, 866 F.2d at 280.

LLLP v. CSC Holdings, Inc. (*Cablevision II*)[66] emphasized that decisions that found a public performance involved the use of a single copy over and over. The use of a single copy for each performance in *Cablevision* rather than reusing one copy for performances from a single copy was critical to the decision and the precedent it created. In the case, the cable operator, Cablevision, offered its customers a network remote-storage digital video recorder ("RS-DVR") service. The service made a copy at its network headend[67] of a show on a linear network, at the request of the customer. This differed from DVRs in other cable systems, where the copying was done in the digital set-top box DVR in the customer's home.[68] Content owners sued, arguing that, unlike customer-situated recordings on an in-home DVR or VCR, here Cablevision was doing the recording and was thus liable for making an unauthorized copy and for publicly performing content through the playback of the networked DVR copy.[69]

In deciding that Cablevision did not engage in unauthorized public performances of the plaintiffs' work through the playback of the RS-DVR copies, the circuit court rejected the idea that the cable operator "performed" the copyrighted work.[70] Instead, it concluded that the playback did not involve the transmission of the performance to the public.[71] It reached this conclusion because only the particular customer who had ordered the recording received the transmission.[72] Both the appeals court and the district court offered detailed treatment of the technology of the copying process.[73]

---

66. 536 F.3d 121 (2d Cir. 2008) (referring to appellants, CSC Holdings, Inc. and Cablevision Systems Corporation, as "Cablevision").

67. A headend is the facility where a cable operator assembles all content—such as broadcast, satellite, and on-demand video—going to and from a subscriber's premises. *See* BRENNER ET AL., *supra* note 13, § 1.5.

68. The court of appeals noted:

> As the district court observed, "the RS-DVR is not a single piece of equipment," but rather "a complex system requiring numerous computers, processes, networks of cables, and facilities staffed by personnel twenty four hours a day and seven days a week." To the customer, however, the processes of recording and playback on the RS-DVR are similar to that of a standard set-top DVR.

Cablevision II, 536 F.3d at 125 (citations omitted).

69. *See id.*

70. *Id.* at 134.

71. *Id.*

72. *Id.* at 139 ("Because each RS-DVR playback transmission is made to a single subscriber using a single unique copy produced by that subscriber, we conclude that such transmissions are not performances 'to the public,' and therefore do not infringe any exclusive right of public performance.").

73. 20th Century Fox Film Corp. v. Cablevision Sys. Corp. (*Cablevision I*), 478 F. Supp. 2d 607 (S.D.N.Y. 2007), *rev'd sub nom.* The Cartoon Network LP, LLLP v. CSC Holdings,

The cable operator emphasized that copying occurred at the direction of the subscriber.[74] Although a "buffer" copy was made before (and whether or not) a customer chose to record it, that "copy" remained in the buffer for no more than 1.2 seconds, a time insufficient to meet the requirement that a copy be "fixed" to constitute "copying" under the Copyright Act.[75]

As to the public performance claim, the district court summed all of Cablevision's RS-DVR subscribers who requested a copy of the program and

Inc. (*Cablevision II*), 536 F.3d 121 (2d Cir. 2008). Judge Chin, who authored the district court opinion finding a public performance, was elevated to the Second Circuit Court of Appeals which reversed this opinion. He vigorously dissented to the Second Circuit's subsequent decision in *Aereo II*, 712 F.3d 676, 696 (2d Cir. 2013) (Chin, J., dissenting), where broadcasters unsuccessfully argued that a system that transmitted broadcast signals via the Internet using dedicated antennas for each customer amounted to unlicensed public performances. Judge Chin's dissent called Aereo's platform a "sham . . . a Rube Goldberg-like contrivance, over-engineered in an attempt to avoid the reach of the Copyright Act and to take advantage of a perceived loophole in the law." *Id.* at 697. Judge Chin distinguished the Second Circuit's *Cablevision II* case inter alia by noting that the cable system had paid license fees for the content performed through the RS-DVR system; had viewers watched the real-time transmission, no additional fee would have been owed. *Id.* at 699–700. Aereo paid the content owners no fees, having taken the signals off-air via dedicated antennas. *Id.* Judge Chin emphasized the distinction between the activity of Cablevision as a licensed customer of the plaintiffs whereas Aereo paid the plaintiff broadcasters nothing. *Id.* It should be noted, however, that the over-the-air signals of broadcasters are offered free, without license, to anyone with a receiving antenna. The content of The Cartoon Network and other *Cablevision* plaintiffs is solely available under license. The significant economic issue raised by Aereo is whether it will cause cable and DBS systems, which pay broadcasters retransmission consent fees under 47 U.S.C. § 325(b) (2012) for the right to include their signals as part of a basic cable tier, to try to avoid these payments by switching to an Aereo-like system. *See* Chris Davies, *Aereo in AT&T and DISH Deal Talks amid Broadcaster Fury*, SLASHGEAR (Apr. 1, 2013), http://www.slashgear.com/aereo-in-att-and-dish-deal-talks-amid-broadcaster-fury-01275957/ (discussing possible deal for AT&T U-verse or DishTV to buy Aereo).

74. While the court placed great emphasis on the customer, as opposed to Cablevision, initiating the copying, see *Cablevision II*, 536 F.3d at 139, other courts have not found that to be a significant factor. *See* On Command Video Corp. v. Columbia Pictures Indus., 777 F. Supp. 787, 790 (N.D. Cal. 1991) ("The non-public nature of the place of the performance has no bearing on whether or not those who enjoy the performance constitute 'the public' under the transmit clause.").

75. The court of appeals found the lifespan of the buffer copy to be de minimus and therefore no copy was created for purposes of liability:

> Given that the data reside in no buffer for more than 1.2 seconds before being automatically overwritten, and in the absence of compelling arguments to the contrary, we believe that the copyrighted works here are not "embodied" in the buffers for a period of more than transitory duration, and are therefore not "fixed" in the buffers. Accordingly, the acts of buffering in the operation of the RS-DVR do not create copies, as the Copyright Act defines that term.

Cablevision II, 536 F.3d at 130.

concluded the work was "publicly" performed.[76] The appeals court rejected this conclusion, stating it "makes Cablevision's liability depend, in part, on the actions of legal strangers."[77] The appeals court emphasized that it was the subscriber-initiated copy that was being "transmitted" or "performed."[78] Because the transmission was to one customer's home only, that transmission (and the copy on which it was based) was not capable of being seen by more than the one home. This one-to-one analysis meant that the potential audience for the particular transmission was limited; therefore, there was no public performance under the Transmit Clause.[79] This remained true even though the potential audience for the underlying work—if thousands recorded the program, thousands would see it, at different times and at different locations—was vast.[80]

The Second Circuit followed this thinking in *WNET, Thirteen v. AEREO, Inc.* (*Aereo*),[81] which reiterated in considerable detail the analysis in *Cablevision.* There, the defendant online distributor retrieved broadcast programs off-the-air from an antenna dedicated to the subscriber (even though there was some antenna reuse) and allowed the subscriber to decide whether to watch immediately, pause, or record the programming via their Internet connection.[82] In finding no public performance by this system, *Aereo*

---

76. *Cablevision I*, 478 F. Supp. 2d at 622–23 ("Under the plain language of this clause, a transmission 'to the public' is a public performance, even if members of the public receive the transmission at separate places at different time.").

77. *Cablevision II*, 536 F.3d at 136.

78. *Id.* at 137 ("And because the RS-DVR system, as designed, only makes transmissions to one subscriber using a copy made by that subscriber, we believe that the universe of people capable of receiving an RS-DVR transmission is the single subscriber whose self-made copy is used to create that transmission.").

79. *Cablevision II*, 536 F.3d at 139.

80. *See* Am. Broad. Cos. v. AEREO, Inc. (*Aereo I*), 874 F. Supp. 2d 373, 389 (S.D.N.Y. 2012), *aff'd sub nom.* WNET, Thirteen v. AEREO, Inc. (*Aereo II*), 712 F.3d 676, (2d Cir. 2013).

> Whether a user watches a program through Aereo's service as it is being broadcast or after the initial broadcast ends does not change that the transmission is made from a unique copy, previously created by that user, accessible and transmitted only to that user, the factors *Cablevision* identified as limiting the potential audience.

Id. at 389 (citing *Cablevision II*, 536 F.3d at 134–39).

81. WNET, Thirteen v. AEREO, Inc. (*Aereo II*), 712 F.3d 676, 689, *rehearing en banc denied,* 722 F.3d 500 (2d Cir. 2013), *petition for cert. filed* (Oct. 11, 2013) ("It is therefore irrelevant to the Transmit Clause analysis whether the public is capable of receiving the same underlying work or original performance of the work by means of many transmissions."). Injunction was also denied against Aereo in *Hearst Stations v. Aereo*, Civ. A. No. 13-11649-NMG, 2013 WL 5604284 (D. Mass. Oct. 8, 2013).

82. *Aereo I*, 874 F. Supp. 2d at 376–81.

emphasized—perhaps created—a limitation to *Cablevision*'s non-aggregation of one-to-one transmissions: if private copies are generated from the *same* copy, then private transmissions should be aggregated; "and if these aggregated transmissions from a single copy enable the public to view that copy, the transmissions are public performances."[83]

Both *Cablevision* and *Aereo* concluded that the requirement of the Transmit Clause—"transmit . . . a performance . . . of the work"—was not met.[84] *Cablevision* did so by focusing on two different meanings of the word "performance." The Copyright Act defines "perform"[85] but not "performance." "Performance" can refer to the audio-visual presentation of the work, for example, actors playing and speaking the lines in a movie script (the "work"). Or "performance" can refer the act of *transmission*; for instance, a company can perform delivery of a film reel to a theater in the physical world, or send the work (as in *Cablevision*) via a cable system's network to a subscriber.[86] In the case of the RS-DVR, the system sends electronic impulses to the customer's TV. Thus, the "transmittal of a work" is not the same as transmittal of a "performance."[87] A content owner, like plaintiff Cartoon Network, transmitted a work that was captured by the customer's RS-DVR. The RS-DVR then transmitted the performance of the recorded work. As explained by the court of appeals in a subsequent decision, "the former [is] a transmittal of the underlying work and the latter [is] a transmittal that is itself a performance of the underlying work."[88] The Second Circuit in *Aereo* makes the same point: "*Cablevision* . . . decided that 'capable of receiving

83. *Id.* This statement comes close to the proposal in Part V of this Article. The critical question is not whether the transmissions are made from a single copy or distinct copies but whether the number of transmissions would be considered publicly performed.

84. 536 F.3d at 134; *Aereo II*, 712 F.3d at 696.

85. 17 U.S.C. § 101 (2012).

> To "perform" a work means to recite, render, play, dance, or act it, either directly or by means of any device or process or, in the case of a motion picture or other audiovisual work, to show its images in any sequence or to make the sounds accompanying it audible.

Id.

86. United States v. ASCAP (*In re* RealNetworks, Inc. Yahoo! Inc.) (*RealNetworks and Yahoo! III*), 627 F.3d 64, 73 (2d Cir. 2010) (" '[W]hen Congress speaks of transmitting a performance to the public, it refers to the performance created by the act of transmission,' not simply to transmitting a recording of a performance." (quoting *Cablevision II*, 536 F.3d 121, 136)).

87. The court concluded that "the transmit clause directs us to examine who precisely is 'capable of receiving' a particular transmission of a performance." *Cablevision II*, 536 F.3d at 135.

88. *ASCAP (In re RealNetworks, Inc., Yahoo! Inc.)*, 627 F.3d at 74.

the performance' refers not to the performance of the underlying work being transmitted but rather to the transmission itself."[89]

Thus, using the second meaning of the word "performance," *Cablevision* concluded that "a transmission of a performance is itself a performance."[90] But it is not necessarily a *public* one. To use the language of the Transmit Clause: although the public can receive the transmission, it is not the same as the statutory requirement that the public is "capable of receiving the performance"[91]—that is, the public is capable of receiving a particular *transmission*.[92] Another court characterized this formulation this way: courts "are to look to the transmission being made as the performance at issue, rather than simply to whether the public receives the underlying work."[93]

89. *Aereo II*, 712 F.3d 676, 687.

90. *Cablevision II*, 536 F.3d at 134.

91. *Cablevision II* in this regard turns what might be considered ways the statutory language qualifies performances as public—that they are received at different times or different places—to be less helpful to plaintiffs: "[I]t is of no moment that the potential recipients of the transmission are in different places, or that they may receive the transmission at different times." *Id.* Rather the critical question is to "discern who is 'capable of receiving' the performance being transmitted." *Id.* (quoting 17 U.S.C. § 101 (2012)).

92. Congress has used the concept of transmissions to define copyright liability in the cable and satellite context. So-called secondary "transmissions" of primary transmissions of TV signals by a cable or direct broadcast system ("DBS") are subject to compensation to program owners under a compulsory license, first established in 1976 for cable. 17 U.S.C. §§ 111, 119(a). A cable or DBS system can carry a local or distant TV signal on its channel lineup subject to payments. *See* BRENNER, PRICE & MEYERSON, *supra* note 13, §§ 9:6, 9:30, 9:31, 15:27. What is not certain is which right under 17 U.S.C. § 106 (2012) the cable/DBS compulsory copyright is intended to address: the right to copy, to distribute, or to publicly perform the primary transmission. Most likely, the secondary transmission by the cable system or DBS operator falls under the copyright owner's right to distribute copies, 17 U.S.C. § 106(4).

93. *Aereo I*, 874 F. Supp. 2d 373, 384. Professor Goldstein believes this construction was error. He emphasizes certain words in the Transmit Clause—"to transmit . . . a performance or display of the work . . . to the public, by means of any device or process, whether the members of the public capable of receiving the *performance or display* receive *it* in the" same or different places or times. 17 U.S.C. § 101 (2012) (emphasis added). He concludes: "There can be little doubt that the italicized word *it* in the definition refers to 'performance or display,' not transmission, which in fact appears only as a verb, and not as a noun, in the definition." PAUL GOLDSTEIN, GOLDSTEIN ON COPYRIGHT § 7.7.2, at 7:168 (3d ed. Supp. 2012). However, while this reading makes sense, it does not address what the phrase "to the public" means. Surely, not every transmitted performance is "to the public"; there must be something more. Otherwise, every email sent to one person that may have an audio accompaniment is a public performance. So it is hard to conclude that the *Cablevision II* court was wrong to focus on the nature of the transmission, particularly where a broader reading would have imposed liability.

So the definition of a public performance must inevitably return to the word "public." Courts must give meaning to the size of the actual audience for a work in the on-demand world. Dual use of the term "performance" to cover both the underlying work and the act of transmission allows courts (and those negotiating rights) some flexibility to focus on economically significant numbers of transmissions, discussed below, rather than counting every transmission as a public performance.

How to decide what is the "public," that is, those "capable of receiving" a performance, also arises in differentiating a download from streaming in the ringtone context, although this line-drawing may not fully explain the Transmit Clause. In determining that the act of downloading a ringtone from a vendor is not a public performance, the Second Circuit Court of Appeals distinguished streaming transmissions that render a musical work audible as it is received by the client's computer's temporary memory ("public" performance) from a downloading a musical work that is "transmitted at one point in time and performed at another" (not a performance, public or otherwise).[94] Under this rationale, so long as the performance leads to a contemporaneous delivery to a broad public—even if the challenged performance itself is not seen by anyone[95]—the performance is public. On the other hand, millions of downloads of the same digital ringtone file do not amount to a public performance because the ringtone is not rendered audible during any transmission.[96]

IV. WHEN A TRANSMISSION IS "PUBLICLY" PERFORMED

The questions posed by the Copyright Act, related cases, and commentary beg for a useful, credible criterion to draw the line between public and non-public performance when a "transmission" of a performance occurs. Does simultaneity of watching or listening matter? Does the fact that

94. *RealNetworks and Yahoo! III*, 627 F.3d at 74.

95. *Id.* (citing *NFL v. PrimeTime 24 Joint Venture*, 211 F.3d 10 (2d Cir. 2000), as an example of meeting this contemporaneous public performance standard). In *NFL*, the NFL sought to enjoin transmissions sent to Canada by a satellite uplink. *NFL*, 211 F.3d at 11. The problem for the copyright owners was that the uplink provided by defendant Prime Time 24 itself could not be perceived by viewers; the performance was only viewable in Canada, which meant bringing the infringement action under Canadian copyright law. *Id.* at 12. The appeals court determined that the uplink transmission captured in the United States amounted to a public performance because it was an integral part of the process by which the NFL's work was inarguably delivered to the public. *Id.* at 13. Had the uplink transmission only led to downloads of the NFL's content, to be played back on a device at the viewer's choosing, the theory would break down.

96. *RealNetworks and Yahoo! III*, 627 F.3d at 74–75.

one copy is used for one performance matter? Can we use an economic test to separate which performances should be "public" and therefore compensable from those that should not? This Part considers the different criteria in turn.

A. RECEIVED AT "SEPARATE LOCATIONS": A DISTINCTION WITHOUT A DIFFERENCE?

As noted in Part III, *supra*, the Copyright Act statute and legislative history leave no room for doubt that a Transmit Clause performance need not be received at a single location to qualify as "public."[97] This approach differs from the requirement of reaching "substantial persons" under the Public Place Clause. No single location receiving the transmission must have a substantial number of persons under the Transmit Clause. By stating the condition in the negative (i.e., it need not be received at a single location), the Transmit Clause criterion is less helpful than a rule explaining when a performance *is* public. At most it is a clarifying point but does little to prove a positive statement about whether a performance is public under the Transmit Clause.[98]

An early case under the 1909 Act, decided well before the Transmit Clause was enacted, illustrates why its application is not bounded by a single location (as the Public Place clause is).[99] Radio station transmissions in 1925 were received by individuals in separate cars and homes.[100] Still the court in *Jerome Remick & Co.* did not hesitate to describe the performance as "public."[101] The statute is meant to cover such broadcasts, even if delivered to separate locations.[102]

On the other hand, *Cablevision* and *Aereo* indicate that courts cannot simply sum up different transmissions to different locations, however widespread and frequent the performances, and conclude that a performance is public. This principle is not limited to the facts in these cases: determining that a transmission occurs in geographically distinct locations does not decide

97. *See* 17 U.S.C. § 101 (2012) (including in the Transmit Clause "whether the members of the public capable of receiving the performance or display receive it in the same place or in separate places").

98. In terms of logic, it is why neither the converse nor the inverse of a true statement are necessarily true (i.e., if $A > B$, it is not the case that $-A > -B$, or that $B > A$).

99. Jerome H. Remick & Co. v. Am. Auto. Accessories Co., 5 F.2d 411, 412 (6th Cir. 1925); *see supra* note 48 and accompanying text.

100. *See Jerome H. Remick & Co.*, 5 F.2d at 412.

101. *See id.*; *see also* Note, *Copyright Law: Existence of a Second Legitimate Use Held Ineffective to Cure An Infringing Public Performance for Profit*, 1965 DUKE L.J. 404, 406 (1965).

102. *See* 17 U.S.C. § 101 (2012).

whether the performance is public. In traditional linear transmissions,[103] such dispersed performances will be public. However, in the RS-DVR, or even non-networked DVR case, the sum of several individual performances does not equate to a finding of a public performance.[104] If this were not the case, separate performances of, say, a DVD on DVD players would require a public performance license by either the customer or the equipment supplier, a result no less absurd than if the publishers of *Goodnight Moon* could claim the right to a public performance license of iterative plays of an audio download version (or bedside reading) of that book.[105]

B. RECEIVED AT "DIFFERENT TIMES": A SIMILARLY LIMITED CRITERION

Recall that the Transmit Clause indicates that a public performance can occur not only when it is disbursed as to location but also when disbursed as to time. The previously mentioned jukebox example illustrates when "chronologically disbursed"[106] performances amount to public ones.[107] But not all performances, disbursed over time, amount to a public performance. Again, it would be bizarre to conclude that multiple plays of a CD or a child's "Barney" DVD require either the manufacturer or the home purchaser to obtain a performance license. So this "different times" criterion, too, has limited utility in drawing a distinct line between public and nonpublic performances.[108]

---

103. Linear transmission refers to broadcasts or cablecasts of programming at a scheduled time and delivered to all viewers eligible to receive it. It differs from on-demand programming, which is viewed when and how an eligible viewer chooses to watch it.

104. In *Cablevision II*, broadcasters argued that the court should look "upstream," that is, at the original cablecast of the programs that had been recorded, to decide whether they were publicly performed—not at the transmission to the single user through the subscriber's RS-DVR. The court rejected this approach. *Cablevision II*, 536 F.3d 121, 136 (2d Cir. 2008) ("Furthermore, we believe it would be inconsistent with our own transmit clause jurisprudence to consider the potential audience of an upstream transmission by a third party when determining whether a defendant's own subsequent transmission of a performance is 'to the public.' ").

105. The cassette and book are available at *Goodnight Moon (Book and Cassette)*, BARNESANDNOBLE.COM, http://www.barnesandnoble.com/w/goodnight-moon-margaret-wise-brown/1100337988 (last visited Mar. 27, 2013).

106. This is Nimmer & Nimmer's phrase. NIMMER & NIMMER, *supra* note 56.

107. *See* Section III, *supra*.

108. NIMMER & NIMMER, *supra* note 56. As Nimmer & Nimmer point out, the "at different times" language was not necessary to account for the different time zones associated with, say, a single network broadcast. The broadcast in each time zone already constitutes a "public" performance. *Id.*

C. ONE COPY VERSUS MULTIPLE COPIES

In determining that there was no public performance of an audio-visual work, the Second Circuit in *Cablevision*,[109] the *Aereo* district court,[110] and the Second Circuit in *Aereo*[111] all placed great emphasis on the presence of a dedicated copy, created by the viewer, that was performed. Had Cablevision used a single copy of a program or had Aereo streamed all broadcast programming[112] not at the direction of the subscriber, the performance might possibly have been deemed public and resulted in liability, although the Second Circuit did not affirmatively conclude this in *Cablevision*.[113]

In analyzing the video viewing booth[114] and hotel cases,[115] Nimmer & Nimmer suggest that the "different times" criteria can be explained by focusing on whether the *same copy* of a work is repeatedly played or received by different members of the public at different times.[116] Nimmer & Nimmer identified motion picture peep shows and video jukeboxes[117] as technologies through which performances should be considered public even if only displayed to one viewer at a time. The same copy is used over and over again. Summing these "non-substantial" crowds leads to the result that the peep show device or jukebox performance is "public." In contrast, the *Cablevision* court stressed that the performance involved a unique copy for each viewer in finding no performance "to the public."[118] It is, however, worth noting

109. *Cablevision II*, 536 F.3d at 138.

110. *Aereo I*, 874 F. Supp. 2d 373, 399 (S.D.N.Y. 2012), *aff'd*, *Aereo II*, 712 F.3d 676 (2d Cir. 2013).

111. *Aereo II*, 712 F.3d at 696 (2d Cir. 2013).

112. The online offering of broadcast services without the permission of the TV station was deemed to be a violation of broadcasters' copyright in *WPIX, Inc. v. ivi, Inc.*, 691 F.3d 275, 284 (2d Cir. 2012). In that case the online service claimed the right to the Copyright Act's compulsory copyright for secondary transmissions by a cable system, 17 U.S.C. § 111(c)(1) (2006). The court relied on the Copyright Office's analysis of the statute to conclude that the ivi system did not constitute a "cable system" under Section 111(f)(3) of the Act. Ivi had also not obtained retransmission consent of the stations it carried. *WPIX*, 691 F.3d at 285.

113. *Aereo II*, 712 F.3d at 689 (noting "there is an exception to this no-aggregation rule when private transmissions are generated form the same copy of the work." (citing *Cablevision II*, 536 F.3d at 135–37)). But as *Aereo II* itself notes, the aggregated transmissions from a single copy must enable "the public" to view that copy. *Id.*

114. Columbia Pictures Indus., Inc. v. Redd Horne, Inc., 749 F.2d 154 (3d Cir. 1984).

115. On Command Video Corp. v. Columbia Pictures Indus., Inc., 777 F. Supp. 787, 790 (N.D. Cal. 1991).

116. NIMMER & NIMMER, *supra* note 56.

117. *Id.*

118. *See Aereo I*, 874 F. Supp. 2d 373, 384 (citing *Cablevision II*, 536 F.3d at 125, 135, 139).

that the plaintiffs in that case did not think that the number of copies of a work mattered.[119]

The Transmit Clause makes no mention of how many copies are involved in determining whether a performance is public.[120] In the case of live radio or television, there is no copy to even consider beyond the copy that is transmitted linearly. Apart from the Copyright Act's silence, this suggested criterion, i.e., relying on single-copy-to-many as a way to distinguish public from non-public performances, is also likely to produce a false positive, as Nimmer & Nimmer point out.[121] It would mean that renting a videocassette would give rise to a public performance in homes because the same copy gives rise to numerous performances "at different times."

Recognizing that the motion picture studios have not made this argument and may indeed have conceded that in-home use of cassettes or DVDs is non-infringing, Nimmer & Nimmer draw a distinction between private screening rooms where the viewing activity "is a substitute for a theater" and liability attaches; and a screening in a dwelling place, where there is no liability.[122] This distinction would also explain why the video store operator in *Redd Horne* was liable, but the hotel that furnished video discs to individual rooms was not.[123]

This "public space / private space" distinction is of decreasing utility because digital technology and the Internet make it easier to view content in private spaces than in the now nearly extinct world of private screening rooms.[124] More significantly, relying on one copy rather than many copies would recreate the videocassette rental paradigm in cyberspace. It elevates a formality over efficiency and freezes copyright law based on mid-1970s technology. Thus, the courts in *Cablevision* and *Aereo* found no liability under

119. *Cablevision II*, 536 F.3d 121, 137 (2d Cir. 2008) ("Plaintiffs contend that it is 'wholly irrelevant, in determining the existence of a public performance, whether "unique" copies of the same work are used to make the transmissions.' ").

120. The definitions are silent on this point and this issue forms the crux of the *Cablevision* and *Aereo* holdings—that where the performance is to a single customer, using a single copy, there is no public performance. Where a single copy is used to generate several private transmissions, and the aggregated transmissions "enable the public to view that copy, the transmissions are public performances." *Aereo II*, 712 F.3d 676, 689 (2d Cir. 2013) (citing *Cablevision II*, 536 F.3d at 137–38).

121. NIMMER & NIMMER, *supra* note 56.

122. *Id.*

123. *See supra* notes 58–62 and accompanying text.

124. Private screening rooms were often peep show parlors carrying adult films, famously on 42nd Street in New York but common in major cities. The last such theater in Toronto was chronicled in Dimitrios Otis, *The Last Peep Show*, VANCOUVER NEON (Aug. 31, 2005), http://www.vancouverneon.com/page_q/q_arcade.htm.

the public performance clause because each performance was tied to a dedicated copy. The price to avoid copyright liability is creation of a separate electronic copy for each user—"a Rube Goldberg-like contrivance"[125] in the words of the dissent in *Aereo*.

This artificial, and non-enumerated, distinction[126] runs counter to the Constitution's patent and copyright clause itself: "To promote the Progress of Science and Useful Arts."[127] The advent of digital technology—surely a significant scientific development—allows a single server copy to suffice for unlimited performances whereas a VCR copy can only be used on one machine at a time. It would be strange indeed, then, to base the "public performance" definition on a criterion that does not appear in the statute and which has a backwards look to it.

The court in *Aereo* recognizes a limit to the "same copy equals public performance" approach under the Transmit Clause. The "aggregated transmissions from a single copy" must "enable the public to view that copy."[128] While a useful limitation, it begs the question of what constitutes the "public," which this Article addresses in Part V.

In its defense, the one-copy, one-user approach does resemble the reasoning behind the first sale doctrine. While an owner of a copy of a work may not make copies without a license, she may generally sell that copy freely without paying the owner any gain from the sale.[129] The analogy breaks down, however, when one considers that the content at issue under the

125. *Aereo II*, 712 F.3d at 697 (Chin, J., dissenting).

126. James Grimmelman explains the problems with this distinction:

> Making separate copies for each user is a massive waste of storage. Systems engineers would say that *Cablevision* should make only as many copies as it needs to meet demand. Making more does nothing to improve the experience for users; it does nothing to change the impact on copyright owners. All it does is drive up costs. But courts have to play the hands they're dealt, and the *Cablevision* court was working with precedents that made the use of individual copies highly significant. If it is fair to say that Cablevision won on a technicality . . . [t]hese are precisely the kinds of technicalities that matter in modern copyright law.

James Grimmelmann, *Why Johnny Can't Stream: How Video Copyright Went Insane*, ARS TECHNICA (Aug. 30, 2012), http://arstechnica.com/tech-policy/2012/08/why-johnny-cant-stream-how-video-copyright-went-insane/.

127. U.S. CONST. art. I, § 8, cl. 8.

128. *Aereo II*, 712 F.3d at 689.

129. 17 U.S.C. § 109(a) (2012) ("Notwithstanding the provisions of section 106(3), the owner of a particular copy or phonorecord lawfully made under this title, or any person authorized by such owner, is entitled, without the authority of the copyright owner, to sell or otherwise dispose of the possession of that copy or phonorecord."). There are exceptions for rentals of recordings and computer programs. *Id.* § 109(b).

Transmit Clause is not a purchased copy but is licensed for viewing by the subscriber or online buyer.

D. "SUBSTITUTE FOR OTHERWISE PUBLIC THEATER VIEWING"

Nimmer & Nimmer's distinction as to when a single-copy performance is public boils down to this: the location of the transmitted performance is public if that location is really a substitute for a concededly public performance space.[130] Display of a single copy is considered a public performance if the space in which it is shown is a theater; but not if the space in which it is shown is a dwelling.[131]

However, this test too may exclude much that the statute intends to cover. The Transmit Clause was certainly meant to cover traditional broadcasting. This use typically involves a performance at the network studio of one phonorecord (in the case of radio) or one copy of an audiovisual work (in the case of filmed or video entertainment carried by a broadcaster or cable operator), and it is viewed overwhelmingly in private dwellings. In other words, the Transmit Clause should classify performances as public even if they are received in private spaces and even if they use but one copy.

Use of this substitution test would define too many intended public performances as nonpublic. While it may decide the issue of public-versus-private performance of a video rental, it does not serve as a defining criterion in a broader context. Yet the concept of "substitution" appears to be on the right track analytically in trying to apply the Transmit Clause. It invokes the economic consequences that should guide which performances are compensable as public.

V. ADAPTING THE "TRANSMIT" CLAUSE TO THE ONLINE CONTEXT

A. CRITERION ONE: "SUBSTITUTE FOR" VERSUS "ORIGINATE" A PUBLIC PERFORMANCE

Running through the public performance infringement cases, and the system of copyright generally, is the proposition that liability exists when economic benefits accrue to the user of content. Exceptions abound to this proposition,[132] but copyright considers explicitly economic trade-offs in

130. NIMMER & NIMMER, *supra* note 56, § 8.14[C][3].

131. *Id.*

132. A common example is the first sale doctrine, 17 U.S.C. § 109(a) (2012), which states that "the owner of a particular copy or phonorecord lawfully made under this title, or any person authorized by such owner, is entitled, without the authority of the copyright

determining the scope of the owner's exclusivity. For example, the fair use statute considers the economic impact of a use in deciding whether that use infringes.[133] And the exclusion of a performance right in non-digital sound recordings was based, in part, on the view that traditional radio airplay provided free publicity for the sound recording.[134]

How might copyright law's economic trade-offs be considered in determining when a transmission is "publicly" performed? The rewrite of the public performance clauses in 1976 rejected the proposal that any transmission for "commercial purposes" in and of itself is a public performance.[135] However, the concept of "substitution" introduced in Nimmer & Nimmer's formulation above is a useful tool. The inquiry would be two-fold: does the transmission "substitute" for an already likely compensated public performance, or does it originate a performance that has not yet occurred and been paid for? This test looks to whether a prior public performance that has been properly licensed would cover the subsequent performance in question. The subsequent performance then is—or should be—a performance covered by the initial license, though possibly received in different places and nearly always at different times. This attention to the economic consequences formed the basis of the dissent in *Aereo* as well.[136]

---

owner, to sell or otherwise dispose of the possession of that copy or phonorecord." The purchaser of a first edition of a book that becomes a classic gets all the benefits for holding and reselling the book; the creator of the work will get nothing on resale. And fair use under section 107 of the Act allows some uses to be non-infringing. *Id.* § 107.

133. *Id.* § 107 ("In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include . . . (4) the effect of the use upon the potential market for or value of the copyrighted work.").

134. Broadcasters have successfully fended off fees for performances of phonorecords on AM and FM radio, although bills to create a performance right have been frequently introduced. In June 2012, a major broadcaster agreed to pay a record company a royalty for use of its library. *Big Machine Label Group and Clear Channel Announce Groundbreaking Agreement to Enable Record Company and Its Artists to Participate in All Radio Revenue Streams and Accelerate Growth of Digital Radio*, CLEAR CHANNEL, http://www.clearchannel.com/Pages/Big-Machine-Label-Group-and-Clear-Channel-Announce-Groundbreaking-Agreement-to-Enable-Record-Company-and-Its-Artists-to-Par.aspx.

135. *Cablevision II*, 536 F.3d 121, 139 (2d Cir. 2008).

136. Judge Chin, in his dissent in *Aereo II*, attempted to distinguish that case from *Cablevision* (in which he was reversed while sitting as the trial judge). He observed that there were "critical differences between Cablevision and this case. Most significantly, Cablevision involved a cable company that paid statutory licensing and retransmission consent fees for the content it retransmitted, while Aereo pays no such fees." *Aereo II*, 712 F.3d 676, 697 (2d Cir. 2013) (Chin, J., dissenting). Another case has also recognized this distinction:

> If Defendants can transmit Plaintiffs' content without paying a fee, Plaintiffs' existing and prospective licensees will demand concessions to make up the loss of viewership to non-paying alternatives, and may push

If the subsequent use is a substitute for an otherwise compensated performance, however, the subsequent use is not considered a public performance. The paradigmatic example involves a VCR or home-DVR recording of a broadcast television program. The linear network originally licensed the transmitted public performance. The user playing the recorded version is not separately publicly performing the work but getting the value of the content for which the subscriber has already paid the licensee through subscription or downloading charge.[137] The initial public performance license paid by the distributor would cover these time-shifted performances. Thus, the "substitution/origination" distinction would treat a straight-to-online audio-visual work differently than a broadcast or cablecast episode that is also available online. And it should make no difference whether the customer sets the DVR at home, sets a remotely located DVR (as in *Cablevision*), or obtains the program on an online service. Nor should it matter if the "copy" used as a substitute is a digital version of the copy sent to the distributor-licensee.

Online viewing of broadcasted programs operates as a form of place-shifting (if viewed simultaneously with the distributor's transmission) or time-shifting/place-shifting (if not simultaneous) for much of the public.[138] Online viewing constitutes a separate, subsequent performance. But it makes little sense to require distributors to pay additional fees because some viewers forget to set a VCR or DVR to record or decide not to pay the additional fees imposed by the cable operators for its DVR service.

For the major broadcast networks whose programming is free to watch (and record, via VCR, DVD, or DVR) over the air, making content available online allows these networks to extend their audiences to the DVR-less and

> additional players away from license-fee paying technologies and toward free technologies like Defendants'. The availability of Plaintiffs' content from sources other than Plaintiffs also damages Plaintiffs' goodwill with their licensees.

Fox Television Stations, Inc. v. BarryDriller Content Sys., PLC, 915 F. Supp. 2d 1138, 1147 (C.D. Cal. 2012), *appeal docketed sub nom.* Fox Television Stations, Inc. v. Aereokiller, LLC. Nos. 13-55156, 13-55157 (9th Cir. Jan. 25, 2013). A preliminary injunction was also granted against a service similar to Aereo in Fox Television Stations, Inc. v. FilmOn X LLC, No. CV 13-758, F. Supp. 2d, 2013 WL 4763414 (D.D.C. 2013), *appeal docketed,* No. 13-7146 (D.C. Cir. Sept. 17, 2013).

137. If the user performed the recording for a substantial number of persons, it is possible the first clause would trigger, but there could be a fair use defense. For example, a bar that tapes an Olympic event that is on linear TV that airs in the middle of the night for replay when the bar is open, and makes no separate charge, should be excused from liability.

138. The U.S. Supreme Court has stated that since time-shifting expands public access to freely broadcast programming, the popular practice yields societal benefits. Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 454 (1984).

those with poor television reception (as in rural areas), which is part of Aereo's strategy and was the basis for cable television's creation in the 1940s.[139] Moreover, as cable systems increasingly move to Internet protocol ("IP") format for distribution—AT&T's U-Verse is already an all-IP video provider—distinguishing between VCR, DVD, DVR, and IP-delivered content on specialized networks or via the public Internet makes decreasing sense, just as any distinction between digital and analog delivery did prior to over-the-air TV's transition to digital in 2009.[140]

Courts have explicitly prohibited requiring two licenses for the same use of a work. The principle formed the basis of the "through-to-the-viewer" (also referred to as "through-to-the-audience") license; the cable industry established that BMI was not entitled to treat the transmission from the uplink to the satellite and down to the cable headend as a public performance separately compensable from the performance received by the viewer at the end of the transmission.[141] As a practical matter, this means that HBO's PRO licenses cover the cable operator's performance through to the viewer's receiving device (e.g., TV set, iPad, or computer).[142] This principle should also apply to the purely online world,[143] with the added complication that

139. Note that cable's compulsory copyright license regime only charges for distant signals, not for carriage of local signals. *See* BRENNER ET AL., *supra* note 13, § 9:15, :17. Local broadcasters do collect under the separate retransmission consent right, created in the 1992 Cable Act, 47 U.S.C. § 325(a) (2012).

140. *See* Amy Schatz & Christopher Rhoads, *Shift to Digital TV Sends Late Adapters Scrambling*, WALL ST. J., June 13, 2009, http://online.wsj.com/article/SB124480704993709781.html.

141. Nat'l Cable Television Ass'n v. Broad. Music, Inc., 772 F. Supp. 614, 650 (D.D.C. 1991) (deciding that BMI's practice of issuing split licenses was incompatible with the BMI Consent Decree).

142. In setting the through-to-the-audience rate in *United States v. ASCAP* (*In re Turner Broad., Inc.*), 782 F. Supp. 778, 794 (S.D.N.Y. 1991), *aff'd*, 956 F.2d 21 (2d Cir. 1992), the court found that ASCAP's Consent Decree prohibits collecting multiple fees per music use to "all industries in which it was potentially applicable." Whether receiving devices like tablets are covered by current licensing agreements was disputed by Viacom in its distribution agreement with Time Warner Cable in 2011 but was settled the following year. *See* Don Jeffrey & Edmund Lee, *Viacom Settles Dispute with Time Warner Cable over IPad Viewing*, BLOOMBERG.COM (May 16, 2012), http://www.bloomberg.com/news/2012-05-16/viacom-settles-with-time-warner-cable-over-viewing-on-ipads-1-.html.

143. This is essentially the result in the *Cablevision* case. *Cablevision II*, 536 F.3d 121, 138–39. In *ASCAP v. MobiTV*, 681 F.3d 76, 78 (2d Cir. 2012), the appeals court upheld the district court's use of wholesale rather than retail revenues as the basis for computing a public performance license. The through-to-the-viewer concept in cable does not use the larger retail revenue number in computing the value of the music. This is because the purchaser of the music rights is the network (e.g., HBO), not the retail customer. The price paid by the customer includes fees paid to the distributing cable operator. And in *RealNetworks and Yahoo! III*, 627 F.3d 64, 75 (2d Cir. 2010), the court found that the digital

performances may be through-to-the-viewer on the viewer's schedule. Is a performance still "through-to-the-viewer" even if the viewer watches it as a video on demand ("VOD") offering on the cable system or as an online offering on a TV Everywhere platform? The answer should be yes.

Two principles emerge from the foregoing discussion. First, it is artificial at best and rearward at worst to base the public performance license on whether the performance emanates from a single digital copy or a dedicated digital copy assigned to each performance. To do so is to tether the physical copy world to the digital one. The policy may make sense when deciding whether the right to copy has been infringed.[144] Thus, we expect that a copy downloaded on a Kindle is subject to the owner's right to control copies. But that relatively straightforward application of copyright to digital content to determine the right to copy does not explain why a performance that uses a dedicated digital copy (such as in *Cablevision)* is not a public performance but a performance that uses a common copy (such as in *YouTube*) is.[145] For statutory support for this principle, one can look to the fair use provisions in sections 107(1) and (4) of the Copyright Act.[146] A fair use inquiry implicitly includes a consideration of whether a subsequent performance is or is not a substitute for an already-paid-for-performance and is explicitly called out in those subsections.

Second, it does not make sense to differentiate downloaded-pause-perform copies differently from downloaded-no-pause-perform copies. Courts have not addressed this distinction precisely, and semantics add to

download of a song does not constitute a compensable public performance of that song. *See* Stephen Kramarsky, *Public Performance in the Digital Age*, *Law Technology News*, LAW.COM (Nov. 19, 2009), http://www.law.com/jsp/lawtechnologynews/PubArticleLTN.jsp?id=1202435609971 (noting the trend of unwillingness to charge twice for both downloading and performance rights).

144. In the physical world, we do not allow a person who buys a book to make complete photocopies of it without obtaining a license. *See* NIMMER & NIMMER, *supra* note 56, § 8.02. At the same time, the single copy of a book may be passed around and read without additional license. The latter use of the book is not likely to be a satisfactory substitute for the former.

145. *Cf.* Capitol Records, Inc. v. MP3Tunes, LLC, 821 F. Supp. 2d 627, 649–50 (S.D.N.Y. 2011) (finding no "master copy" of songs from which other online copies are made but approving a system that "eliminates redundant digital data").

146. The Copyright Act provides the fair use provisions:

> In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include—(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; . . . and (4) the effect of the use upon the potential market for or value of the copyrighted work.

17 U.S.C. § 107 (2012).

some of the problem. For instance, courts have used "streaming" to convey two quite distinct online events: sometimes streaming refers to a real-time feed of a linear channel, like a broadcast service; other times it refers to downloading a YouTube clip seen only by one user at a time.[147] Instead of using the pause/no-pause distinction, courts should focus on the differences between the audiences that determine whether a performance is "public." This brings the analysis back to the Public Place Clause as a way to apply the Transmit Clause in the streaming context, in addition to the "substitution" test just discussed.

B. CRITERION TWO: THE PUBLIC PLACE CLAUSE'S "SUBSTANTIALITY" REQUIREMENT

The definition of public performance in the statute is not based on a line between substitution and origination, although the statute supports this interpretation because Congress considered both the Public Place Clause and the Transmit Clause as ways to identify public performances. But when are performances wholly originated as VOD or online "public"? The Copyright Act's legislative history suggests that "public" in public performance has meaning in the Transmit Clause, just as it does in the Public Place Clause:

> Under the bill, as under the present law, a performance made available *by transmission to the public at large* is "public" even though the recipients are not gathered in a single place, and even if there is no proof that any of the potential recipients was operating his receiving apparatus at the time of the transmission. The same principles apply whenever the potential recipients of the transmission represent a limited segment of the public, such as . . . subscribers of a cable television service.[148]

The House Report on the Act makes it clear that whether any particular number of viewers actually watches a transmitted performance is immaterial.[149] But the highlighted clause requires that the transmission is made available "to the public at large" to qualify as a public performance. In

147. *Compare RealNetworks and Yahoo! III*, 627 F.3d at 74 ("A stream is an electronic transmission that renders the musical work audible as it is received by the client-computer's temporary memory. This transmission, like a television or radio broadcast, is a performance because there is a playing of the song that is perceived simultaneously with the transmission."), *with id.* at 69 ("For example, a user can enjoy the specific song or music video he desires from an 'on-demand' stream in Yahoo! Search.").

148. H.R. REP. NO. 94-1476, at 64–65 (1976), *reprinted in* 1976 U.S.C.C.A.N. 5659, 5677–78 (emphasis added).

149. *Id.*

a broadcast or cablecast linear performance, the transmission is made to the entire viewer or subscriber base; the potential audience is always more than a few.

In VOD, the path to the viewer differs. A transmission is made to the cable system headend's server from a program source, to await a customer's possible request to view. But unless and until a customer orders that VOD content, it never reaches the public.[150] The same principle applies to online content—if no one ever visits the webpage or chooses to view the posted video, the content never reaches the public.[151] In short, public performance requires either a measurable audience or one that is predicted to be substantial. If neither occurs, the performance should not be considered public for copyright purposes.

This interpretation of the Transmit Clause—that the actual number of viewers is immaterial, but only so long as the transmission is actually to a predicted substantial public—gains support by the words of that Clause used to define "publicly." In the "separate places / different times" clauses, the statute requires that "*the public capable* of receiving the performance . . . *receive* it."[152] If no one orders a VOD program, no one actually "receives" it, and the work would not be performed publicly.

This condition precedent—that a copyrighted audio-visual work be distributed to and perceived simultaneously by the public—is the working assumption behind a linear channel, whether broadcast over the air or streamed live online. It is why it makes sense that the words of the Transmit Clause classify radio and broadcast/cablecast linear transmissions as public performances. Even if no one winds up watching a particular program, the distributor assumes that the service will be viewed by a substantial segment

150. *See* Carrie Nation, *What is VOD Technology?*, HOUS. CHRON., http://smallbusiness.chron.com/vod-technology-14311.html (last visited Oct. 5, 2013) ("Before it is delivered instantly to your home, the show or movie is first put into digital format, and stored on a video server. When a video on demand request is made, the movie is compressed and transmitted through the cable or broadband connection."). Until a subscriber orders content, the file sits at the server. *Id.*

151. The United States District Court of the Southern District of New York reached a similar conclusion as to why ringtone downloads do not constitute a public performance: the ringtones once downloaded may never actually be played (the user may choose another ringtone) and even if played, they may never be played in public. *In re* Cellco Partnership, 663 F. Supp. 2d 363, 377 (S.D.N.Y. 2009). A TV receiver is similarly "capable of receiving" a performance. But no one would seriously conclude that TV receivers in a show room not wired for cable are capable of receiving a transmission; or that a TV receiver outside of a TV station's coverage area is "capable" of receiving a performance from that station. To be "capable" must mean that the performance can be and at some point is received.

152. 17 U.S.C. § 101 (2012) (emphasis added).

of the public, to generate an audience for the accompanying advertising or to justify the subscription price. If this assumption proves wrong, at some point a lightly viewed program is cancelled. Otherwise the expenditure for its programming and channel capacity would make no sense.

But that is not the working assumption behind every audio-visual work on the Internet. As the cost of distribution is nearly zero to post a video to YouTube, the myriad economic considerations that go into linear distribution, including production, distribution, and marketing costs, do not apply. It is therefore logical to differentiate between direct-to-VOD or direct-to-online programming (i.e., content not also available on a separately licensed linear network) that is significantly viewed by set-top-box-equipped viewers or those viewing on the web; and such programming that might be viewed only by a few customers or perhaps none at all.

Looked at this way, the Public Place Clause in the Act's definition of public performance informs its companion Transmit Clause. The first clause declares a public performance when the performance is at a place where a "substantial number of persons . . . is gathered."[153] This language does not modify the Transmit Clause, to be sure. But it suggests a common-sense requirement of predicted substantiality be imposed on liability under the Transmit Clause, where a linear broadcast is not involved. Absent evidence that a "substantial number" of viewers are likely to view a particular VOD or nonlinear online offering, it should not be deemed to be publicly performed.

How does this second criterion square with *Cablevision*? The short answer is that the first test (substitution), not the second (substantiality), would govern. It is worth recalling that *Cablevision* rejected the plaintiffs' view that the *potential* recipients of the linear transmission should be counted in determining whether the one-to-one RS-DVR playback transmission was publicly performed.[154] Instead, the Second Circuit found that the RS-DVR "transmission" went to one recipient: "the universe of people capable of receiving an RS-DVR transmission is the single subscriber whose self-made copy is used to create that transmission."[155]

This seems correct in the RS-DVR situation, but it creates a large exception in the online context. As noted, as a technical matter, every unicast stream is but a single one-to-one transmission. Even a simulcast of an admittedly publicly performed linear network, on the Internet, is a single transmission, separately established for each person seeking access to the file.

153. *Id.*
154. *Cablevision II*, 536 F.3d 121, 137 (2d Cir. 2008).
155. *Id.*

For example, the simulcast of C-SPAN.org is a unicast, in reality. If the test is whether an online performance is a single, one-to-one transmission, then every performance is not "public": there can never be "public" performances on the Internet.

Cablevision reached its conclusion because it was concerned with the "hapless customer" whose liability (or Cablevision's) turned on whether some other unknown party had "transmitted" the same program by ordering and playing a copy.[156] The court focused on a "potential" audience that could, in its view, unfairly trigger liability.[157]

Thus, the correct way to analyze the *Cablevision* case is under the first, not second, criteria: the RS-DVR performance merely substituted for a public performance *already under license* to Cablevision for use by its subscribers on the plaintiff linear networks. The transmission system in *Cablevision* was not the Internet but the closed cable system available only to the operator's subscribers. *Cablevision* should not be read to exclude "public" out of any online transmission; unicasts of one-to-one transmissions to a substantial audience are publicly performed. Web-only clips or segments made for on-demand cable viewers, if actually viewed in large numbers, should be covered by the Transmit Clause.[158]

C. APPLYING THE "SUBSTITUTE" AND "SUBSTANTIAL AUDIENCE" TESTS

1. DVR and RS-DVR

Cablevision illustrates how the first criterion works. Assume a program that had been recorded on a network DVR had been under a performance license when it appeared on the original linear broadcast or cablecast

156. *Id.* at 136.

157. *Id.* at 135 ("[The Transmit Clause] speaks of people capable of receiving a particular 'transmission' or performance,' and not of the potential audience of a particular 'work.' "). The traditional focus of the "transmit" clause—broadcasting—did not consider the actual numbers of people listening or viewing a work, only the number capable of doing so. Only the one customer who ordered an RS-DVR is capable of receiving a transmission of that copy, so the court did not undo the traditional reading of the clause.

158. *Cf.* H.R. REP. NO. 90-83, at 29 (1967) (A public performance occurs "where the transmission is capable of reaching different recipients at different times, as in the case of sounds or images stored in an information system and capable of being performed or displayed *at the initiative of individual members of the public*.") (emphasis added). While the 1967 report contemplated computers and other information retrieval systems, the Internet—a combination of interlinked computers—was first introduced in the late 1960s and could not have been in the contemplation of the report writers in 1967. *Internet History*, COMPUTER HISTORY MUSEUM, http://www.computerhistory.org/internet_history/ (last visited Oct. 5, 2013) (describing the first host-to-host connection on October 29, 1969). In addition, this House Report did not accompany the 1976 Copyright Act.

transmission. Its replay on the RS-DVR would substitute for a transmission that had already been licensed for public performance. There is no reason that the PRO should collect an additional public performance fee for a VOD performance rendered at "different times."

2. Video on Demand ("VOD")

Increasingly, cable operators also offer linear content on demand (e.g., HBO on Demand features the network's series and movies). Indeed, from a capital expenditure viewpoint, VOD plays can be viewed as a more economical version of RS-DVR, since such a network-based system does not require the operator to furnish DVR-capable, set-top box equipment in the home. Instead of waiting for the customer to push the "record" button, everything authorized by the content owner is, or eventually could be, put into the on-demand listings.[159]

Assume that the cable operator already has a reproduction license to make the VOD server copy.[160] The underlying linear content that is available on a particular cable system channel but also available on demand is subject to a public performance license. Adding VOD use will likely not increase the total amount of public performance of the work beyond what the linear audience could generate. Calling up the shows on VOD is simply a substitute for linear (or DVR/VCR) viewing.[161] The network performance license should be viewed as covering these additional on-demand performances under the first "substitution" criterion.

However, not all performances of VOD content substitute for a linear or DVR-recorded version. Cable companies also post VOD-only programming. In particular, recent first-run releases are frequently available on demand at about the same time that the film is released to DVD or, in some cases, in

159. Cable operators vary in the amount of content offered to customers, with Comcast's Xfinity platform being among the most robust. *See* Brian Stelter & Amy Chozick, *Viewers Start to Embrace Television on Demand*, N.Y. TIMES, May 20, 2013, http://www.nytimes.com/2013/05/21/business/media/video-on-demand-viewing-is-gaining-popularity.html. But not all non-premium linear content is available on the on-demand platform. For instance, ABC's *Modern Family* was not available as VOD in 2011. Stuart Miller, *On-Demand Viewing Poses a Test for Broadcasters*, N.Y. TIMES, May 1, 2011, http://www.nytimes.com/2011/05/02/business/media/02episodes.html.

160. Under section 102's exclusive rights to copy and distribute, VOD copies must be authorized by license, and not all content owners choose to license content for VOD. *See* 17 U.S.C. § 102 (2012). Much of the *Cablevision* opinions concern whether the customer or the cable operator created the playback copy. The Second Circuit, reversing the District Court on this finding, held that the customer, not Cablevision, made the fixed copy. *Cablevision II*, 536 F.3d 121, 139. In VOD, the playback copy is made under license by the cable operator.

161. Indeed, there may be less value to the operator where the VOD performance allows skipping commercials or promotional spots that are part of the VOD play.

theaters.[162] Additionally, operators (and program networks) make thousands of older films and VOD-only programs available to the VOD (or online) platform. There is no pre-existing relevant linear performance of these programs, and so there is no pre-existing license covering the VOD performance. When these programs are transmitted on VOD, it is necessary to decide whether their performance is public.

This issue, which arises even more prominently in online video (discussed *infra*, Section V.C.3), is how to distinguish VOD programming that may be infrequently or never actually played from VOD content where the use is significant. For instance, a cable operator may agree to post public service videos, which may be viewed only rarely, as part of a franchise agreement or community service commitment.[163] On the other hand, an older holiday children's film, not seen on a linear network, and directly licensed from the copyright owner, may receive thousands of requests during certain periods. Here the second criterion would apply if the VOD performances involve a "substantial" number of transmissions. Thus some, but not all possible, works offered on the VOD are publicly performed and require separate licensing. The VOD distributor's PRO license should reflect this distinction between substantial and non-substantial use of a particular program.

3. Online Video Streaming

The two criteria can also determine whether an online video streaming performance requires a public performance license. Where viewing the

162. Content owners have changed the distribution patterns for post-theatrical runs of feature films; much of this strategy is influenced by the worldwide illegal copying on the Internet. The highest margins are associated with purchases of DVDs, but in many cases, prices for new DVD releases have dropped to the $10–20 range from the $70–90 range ten years ago. Wal-Mart and Amazon dropped prices in 2009. Michelle Chapman, *DVD Prices Drop at Online Giants in New Retail War*, BOSTON.COM (Nov. 7, 2009), http://www.boston.com/business/technology/articles/2009/11/07/dvd_prices_drop_at_online_giants_in_new_retail_war/. The decline in retail rentals and the expansion of TV homes equipped with digital boxes to purchase on-demand movies has made cable and online VOD (such as Vudu and Amazon Prime) increasingly important distribution channels. *See* Mike Isaac, *From Apple to Vudu: 8 Netflix Alternatives Compared*, WIRED (Sept. 21, 2011, 6:30 AM), http://www.wired.com/gadgetlab/2011/09/netflix-alternatives/all/.

163. Public access programming, much of which receives few viewers, may be more conveniently found on VOD than on a linear channel because VOD is searchable whereas linear access programming is not always on a regular schedule. A sample agreement for VOD carriage can be found at Brian T. Grogan, Negotiating PEG Channel Carriage Lessons from Retransmission Content, 2012 National Association of Telecommunications Officers and Advisors Annual Conference (Sept. 27–29, 2012), http://www.natoa.org/events/PEGChannelCarriageGrogan.pdf (last visited Oct. 6, 2013).

programming online is a substitute for the linear broadcast or cablecast version, the linear performance license should cover the online viewing. There is no policy reason to require a second payment for the online viewing generally.[164]

Where online video content is not part of a linear network offering (by far the largest share of online content, if assuredly not always the most watched),[165] the "substantial," not the "substitute," criterion makes the most sense to apply. Many, if not most, online videos are not substantially watched. For instance, one unofficial source suggests there are over 1.3 billion YouTube videos,[166] and it is unlikely that the majority are performed substantially (and there are many that are not performed on devices in the United States).[167] The same viewing pattern may be true of videos on other sites populated by user-generated content.

It will be necessary to define "substantial" in this context, as well as made-for-VOD. One way to approach this is to translate what the words "members of the public" in the Transmit Clause mean, as a practical matter, in the broadcasting context. A TV broadcast that regularly reaches only hundreds or even a few thousand listeners or viewers, outside of the smallest markets, is unlikely to remain viable. An accurate test would quantify the online equivalent of a viable linear program audience. In other words, how many views per month would a video require before that video, in the linear context, would be deemed to have reached a substantial audience? In the broadcasting context, program ratings that exceed failing or cancelled linear programs would amount to "substantial." The online equivalent of "too small to succeed" for a linear channel would fail to qualify the program as "publicly performed" under the Transmit Clause.[168] On the other hand, a

164. One exception would be an online version of a linear program that is likely to be significantly viewed over and over (e.g., a famous blooper, the final episode of a popular program). In such cases, it is likely the program will not be the online version of a current linear program.

165. YouTube established early limits to the length of posted programming to avoid unlicensed carriage of TV episodes and movies. *See supra* note 16.

166. *How Many Videos Are There on YouTube?*, HOWMANYARETHERE.NET, http://howmanyarethere.net/how-many-videos-are-there-o-youtube/ (last visited Oct. 6, 2013).

167. *See Statistics*, YOUTUBE, http://www.youtube.com/yt/press/statistics.html (last visited Oct. 6, 2013) ("YouTube is localized in 56 countries and across 61 languages").

168. For example, the trombone fanfare music sheet for "Low Rider" was viewed 216 times before the Author's view. *Low Rider Trombone Fanfare Sheet Music*, METACAFE.COM, http://www.metacafe.com/watch/yt-ib2W1WSrJ70/low_rider_trombone_fanfare_sheet_music/ (last visited Aug. 13, 2009).

popular music video, like *Harlem Shake Miami HEAT Edition*, has been performed over 40 million times, and its performance online is public.[169]

This chart summarizes when an online or VOD performance would require a license:

Table 1: VOD Public Performance and Licensing Requirements

NO LICENSE REQUIRED; NO PUBLIC PERFORMANCE	LICENSE REQUIRED; PUBLIC PERFORMANCE
Performance substitutes for already licensed linear performance	
Original performance with no substantial number of persons in audience	Original performance with substantial number of persons downloading/streaming

How might licensees and the PROs attempt to establish a cut-off? Access to ratings from online measurement services (or the website's own records of VOD sales) would help to develop a formula, but several factors to consider are apparent. Take the example of an online feature film not part of an otherwise licensed linear service. A distributor could tally how many times a month a direct-to-VOD or direct-to-online work is performed per month. Say the film is viewed 5,000 times on the distributor's platform. If this film were carried by a linear network, would that number of views be viable to remain on the network schedule? And given the potential audience for online viewing—that is, all those with access to the program's service—what should be the reasonable cut-off to determine substantiality as percentage of customers? For Netflix or Hulu Plus, the denominator might be the number of subscribers. For YouTube, Hulu or other non-subscription websites, the denominator might be the average number of viewers of video content per month.

The main point of the test proposed here is that some, but not all, online audio-visual works should be deemed "publicly" performed. The net result is that in rate-setting by negotiation or rate court, PROs and their distributing licensees should agree that less than all "streaming" needs a license. As "non-experimental" licenses develop,[170] this Article recommends use of the

169. *Harlem Shake Miami HEAT Edition*, YOUTUBE, http://www.youtube.com/watch?v=Ir2TdfSwH8g (last visited Mar. 23, 2013).

170. PROs have offered experimental licenses for online uses, recognizing that nascent online businesses will try different revenue models. *See, e.g.*, *ASCAP Experimental License Agreement for Non-Interactive Services—Release 5.2*, AM. SOC'Y COMPOSERS, AUTHORS AND PUBLISHERS, http://www.ascap.com/~/media/files/pdf/licensing/digital/non-interactive licenseagreementr5_2.pdf (last visited Oct. 6, 2013); *SESAC Internet License Agreement*, SESAC, http://www.sesac.com/pdf/internet_ATH_2008_click.pdf (last visited Oct. 6, 2013).

"substitutions" or "substantiality" criteria in applying the Transmit Clause, rather than relying on the "single copy" test of *Cablevision* and *Aereo* or the hard-to-apply "separate places" / "different times" language of the statute.

VI. PUBLIC PERFORMANCE AND PROS: ARE THERE ALTERNATIVES FOR AUDIO-VISUAL WORKS?

It is worth considering whether there are other ways of thinking about paying for a public performance, in the context of online audio-visual works, that more directly tie into the value that the owners of a film or television show place on the synchronized music. In other words, does the concept of a separate public performance license for audio-visual works really make sense anymore? Or should these rights be part of the parcel of rights obtained by content owners in order to benefit composers when the audio-visual work is first created, causing PROs to cease collecting for audio-visual performances?

The role of PROs in collecting for public performances of audio-visual works is well established.[171] But, this role was hardly inevitable when necessity led to ASCAP's creation in 1914, before the existence of audio-visual works. ASCAP was founded to correct an economic injustice experienced by composing giants of early 20th century popular music like Victor Herbert, Irving Berlin, and John Philip Sousa.[172] Having written

171. The first television public performance license was negotiated in 1949. *History*, *supra* note 10. Motion picture theater public performance licenses for compositions existed starting in the 1920s, but litigation led to their incorporation in the synchronization licenses when the movies were created:

> ASCAP had begun licensing motion picture theater exhibitors in the 1920s during the "silent movie" era, when the only music performed in a theater was played live (such as by a piano player). Because theaters did not know in advance what music was going to be played, it made sense to cover these performances under a blanket license in order to avoid any question of copyright liability. Even after the creation of "talking pictures," in which music was pre-recorded with the motion picture, ASCAP continued to license performance rights to the motion picture theater exhibitors. Thus, when a motion picture theater exhibitor received a movie from a producer, all of the rights needed for that exhibitor to display the film came "in the can" of film, except for the music performing rights.

Id.

172. It is interesting to speculate on how composers of music that was publicly performed before the 20th century were compensated, if at all. Congress established the copyright owner's exclusive right to publicly perform music in 1897. Act of Jan. 6, 1897, ch. 4, Stat. 481 (codified at 17 U.S.C. § 106(4) (2012)). Sales of piano rolls and sheet music existed in the 19th century, and musicians who wanted to play a composer's work would have to acquire copies of the orchestration, which the composer could authorize to be

popular compositions, their publishers could generally control the reproduction of their creations in piano rolls or sheet music. And, they could control, more or less, the licensing of dramatic rights if stage shows were involved—a Broadway version of a Berlin song would quickly be detected.

But an orchestra might publicly perform a song in a restaurant with no payment to the composer for the use of his work. This occurred in a dinner room of Shanley's Restaurant in New York. The restaurant owner did not obtain rights from the composers, who only by accident would learn about the use of their compositions.[173] The need to detect and collect for such uses led to creating a system so that composers could benefit when their songs were performed. As Oliver Wendell Holmes wrote, "If music did not pay, it would be given up."[174]

This early case demonstrated the need to license performance rights separately and led to the creation of ASCAP, formed and continually governed by composers and their representatives.[175] BMI, formed in 1939, is owned by broadcasters,[176] and SESAC is privately owned.[177] It could be

copied and distributed by a publisher. But it is also worth noting that some of humankind's greatest compositions occurred without the competitive spur of copyright, although religious and royal commissions provided incentive to compose in many instances.

173. Herbert v. Shanley Co., 242 U.S. 591, 594 (1917). John Church Company's comedy march, "From Maine to Oregon," was performed in the dining room of the Vanderbilt Hotel for dinner guests; music from Victor Herbert's "Sweethearts" was performed by singers at Shanley's Restaurant on Broadway. *See id.*; Leonard Allen, *The Battle of Tin Pan Alley*, 181 HARPER'S MAG. 514, 516 (1940).

174. *Herbert*, 242 U.S. at 594.

175. ASCAP was accused of monopolizing performance rights and entered into several consent decrees. *See* United States v. ASCAP, 1940–43 Trade Cas. (CCH) ¶ 56,104 (S.D.N.Y. 1941), *amended*, No. 42-245, 1950 WL 42273, 1950–51 Trade Cas. (CCH) ¶ 62,595 (S.D.N.Y. July 17, 1950), *amended*, No. 41-1395, 2001 WL 1589999, 2001–02 Trade Cas. (CCH) ¶ 73,474 (S.D.N.Y. June 11, 2001).

176. Formed as a non-profit-making performing rights organization, BMI was founded by radio executives to provide competition to ASCAP in the field of performing rights for music writers and publishers. *History*, *supra* note 10. Accused of monopolizing the licensing of performance rights by creating an illegal copyright pool, BMI agreed to a consent decree similar to ASCAP's 1941 decree. United States v. BMI, 1966 Trade Cas. (CCH) ¶ 71,941 (S.D.N.Y. 1966), *amended*, No. 64-CIV-3787, 1994 WL 901652, 1996-1 Trade Cas. (CCH) ¶ 71,378 (S.D.N.Y. Nov. 18, 1994) [hereinafter BMI Decree]. It too has a rate court in the event that licensees and BMI cannot agree to terms.

177. Representing a small minority of the royalty pool owners (between three and ten percent: a disputed number that itself makes it harder to determine its licensing fees—BMI and ASCAP control the rest), SESAC was founded in 1930 to assist European composers with securing rights in the United States. It was purchased by private equity in 1992 and signed marquee writers like Bob Dylan and Neil Diamond. Unlike ASCAP and BMI it is not subject to Department of Justice Consent Decree. *About SESAC*, SESAC, http://www.sesac.com/About/History.aspx (last visited May 9, 2013).

argued that the orchestras that bought a composer's sheet music could have purchased at the same time a performance right to allow them to play the songs publicly. But such rights bundling would have disserved those who only wanted sheet music for personal use, a common home use of music in the early twentieth century (along with player piano rolls).

Similarly, high transaction costs of licensing each radio or Internet performances of compositions make PROs and their offer of a nonexclusive blanket license an essential part of rewarding composers when their songs are played. Detection, licensing, collection, and disbursement would likely be impossible for individual composers in many contexts, although PROs do not nominally have exclusivity to license a composer's work.[178] A non-collective approach seems unworkable in these cases,[179] and blanket licensing has been upheld for use by nightclubs and bars,[180] radio stations,[181] television networks,[182] and local stations, including their syndicated programming.[183]

The advent of separate public performance licenses of music in audio-visual works, distinct from the right to synchronize the composition to a visual work, became an important turning point for PROs. It expanded their licensing domain—and enhanced the justification for PROs' looking back,

178. Movie theater owners successfully challenged the ability of ASCAP and BMI to obtain exclusive licenses for music performances from their members and affiliates, preventing theaters from negotiating directly with composers for rights to individual compositions. *See* Alden-Rochelle, Inc. v. ASCAP, 80 F. Supp. 888 (S.D.N.Y. 1948). The ASCAP consent decree was amended in 1951 to require ASCAP to grant a blanket license to anyone requesting it, but prohibiting ASCAP from acquiring exclusive music performing rights and from interfering with the right of its members to issue a performance license. United States v. ASCAP, 1950–51 Trade Cas. (CCH) ¶ 62,595 (S.D.N.Y. 1950). The BMI consent decree was similarly amended in 1966. United States v. BMI, 1966 Trade Cas. (CCH) ¶ 71,941 (S.D.N.Y. 1966).

179. As Professor Lionel Sobel observed:

> ASCAP's enforcement activities have never been criticized. The Justice Department and the courts always have recognized that it would be impossible for individual composers and music publishers to police the public performance of their works. Thus, in this regard, there seems to be a consensus that ASCAP performs an essential service.

Lionel S. Sobel, *The Music Business and the Sherman Act: An Analysis of the 'Economic Realities' of Blanket Licensing*, 3 LOY. L.A. ENT. L. J. 1, 3–4 (1983).

180. BMI v. Moor-Law, Inc., 527 F. Supp. 758 (D. Del. 1981), *aff'd mem.*, 691 F.2d 490 (3d Cir. 1982); BMI v. Grant's Cabin, Inc., No. 77-1192C(1), 1979 WL 1063 (E.D. Mo. 1979).

181. K–91, Inc. v. Gershwin Publ'g Corp., 372 F.2d 1, 7 (9th Cir. 1967).

182. CBS v. ASCAP (*CBS-remand*), 620 F.2d 930 (2d Cir. 1980).

183. Buffalo Broad. Co. v. ASCAP, 744 F.2d 917 (2d Cir. 1984); United States v. ASCAP (*In re* Shenandoah Valley Broad., Inc.), 208 F. Supp. 896, 897–98 (S.D.N.Y. 1962), *aff'd*, 331 F.2d 117 (2d Cir. 1964).

the split of public performance from synchronization rights seems unnecessary. Why, other than bolstering ASCAP's *raison d'etre*, did it make sense to split synchronization and public performance licensing when they could have been granted at the same time?[184] The origin of that split as an exercise in turf-building is less mischievous than might first appear, however.

As chronicled in the *Alden-Rochelle* case,[185] theaters began obtaining blanket licenses from ASCAP in 1923 to cover the music played by pianists, organists and orchestras that accompanied films.[186] With the advent of sound movies, ASCAP negotiated public performance licenses based on seating capacity.[187] In 1947, ASCAP proposed a new formula that would have hiked the theater license fees by as much as 1500%.[188] This threat led theater owners to seek a new rate, with fee increases of 25–30%.[189] The fee dispute led non-signatory movie theater owners to allege that ASCAP's licensing terms violated the antitrust laws.[190] In 1948, the theater owners in *Alden-Rochelle v. ASCAP* successfully prevented ASCAP from issuing performance licenses on the original 1947 terms and restrained ASCAP members from refusing to grant joint performance-synchronization rights licenses to movie producers.[191]

The same year, a court denied ASCAP members recovery against theaters that had publicly performed music without a license.[192] More substantially,

---

184. *See supra* note 171.

185. Alden-Rochelle v. ASCAP, 80 F. Supp. 888 (S.D.N.Y. 1948), *amended*, 80 F. Supp. 900 (S.D.N.Y. 1948).

186. *Alden-Rochelle*, 80 F. Supp. at 892–93.

187. *Id.* at 892.

188. *Id.* at 895.

189. *Id.*

190. *Id.* at 896.

191. *See id.* at 894–96. As the Second Circuit explained in *CBS v. ASCAP*, 562 F.2d 130, 133 (2d Cir. 1977):

> The problem was special to the theatre exhibition industry which was required at that time to take an ASCAP blanket performance license in order to exhibit motion pictures, the synchronized music of which had already been licensed to the motion picture producer. The specific holding . . . in *Alden-Rochelle, Inc.* . . . was that it was unlawful for ASCAP to require the motion picture producer to contract with distributors that the film would be shown only in theatres having an ASCAP *performance* license. In broader terms, the decision held that ASCAP was a combination in restraint of trade because the members had transferred all their non-dramatic performing rights to ASCAP and were barred from individually assigning such rights to motion picture producers.

Id. (emphasis added).

192. M. Witmark & Sons v. Jensen, 80 F. Supp. 843 (D. Minn. 1948), *appeal dismissed mem. sub nom.* M. Witmark & Sons v. Berger Amusement Co., 177 F.2d 515 (8th Cir. 1949).

Alden-Rochelle pointed to a loophole in the 1941 ASCAP consent decree, which allowed ASCAP to require its members to pool all of their licensing revenues, including revenues received by members who directly licensed music.[193] That rule made it unlikely that any composer or publisher would ever grant separate public performance licenses to producers.[194] A 1950 amendment to the Consent Decree prohibited ASCAP from requiring members to pool fees they received when directly issuing their own performance licenses.[195]

The economic consequence for music publishers was incorporation of the expected value of U.S. domestic theatrical public performance rights into the value of the synch license granted to the movie producer. The audio-visual work, which is a derivative work of the composition, cannot be created without the synchronization license. So, the creation of that license also presents the opportunity to assign a value for a movie theater public performance license. In other words, the price of the license paid by the firm producer to create the audio-visual work also included the public performance license needed to perform the finished work in movie houses. As a general result, then, movie theaters do not obtain public performance licenses from PROs when exhibiting motion pictures.[196]

There are three observations about this late-1940s litigation that bear on the public performance licenses in audio-visual works outside of motion picture theaters. First, music was initially unattached to the visual work being created by silent movie producers. Music performed in theaters by pianists or organists fit snugly into the justification that led to ASCAP's creation, that is, that composers could not easily license or monitor performances of music in thousands of theaters. So, pre-talkie motion picture theaters using live accompanists were unquestionably suitable for ASCAP licenses. But, that suitability did not make ASCAP's later audio-visual licensing regime inevitable or necessarily desirable. Even when talkies developed, the audio portion was provided by publicly performing sound from a separate disc.[197] The early commercial soundtracks were not printed on the film print itself (which had been tried unsuccessfully by Western Electric in the early 1920s),

193. *Alden-Rochelle*, 80 F. Supp. at 892.

194. *See also* United States v. ASCAP, 1950 Trade Cas. No. 62,594 at ¶ 63,752 (S.D.N.Y. 1950) (incorporating the *Alden-Rochelle* holding that prohibited ASCAP from issuing performance licenses to movie theaters).

195. *Id.* at ¶ 63,754 (creating, with the 1950 amendment, the rate court process when a user is dissatisfied with the fee that ASCAP demanded).

196. RON SOBEL & DICK WEISSMAN, MUSIC PUBLISHING 40 (2008).

197. DONALD CRAFTON, THE TALKIES: AMERICAN CINEMA'S TRANSITION TO SOUND, 1926–1931, 59–60 (1997).

but issued on separate phonograph records to exhibitors.[198] Warner Brothers' Vitaphone system, used from 1926 to 1931, was the only commercially successful sound-on-disc system; the discs were recorded at thirty-three and one-third rpm and played on a turntable, physically coupled to the projector motor of the film being projected.[199]

Second, from a copyright standpoint, because of the early separation of sound and pictures, there was no synchronization license needed to create the first sound pictures. Synchronization of sound to picture occurred only upon performance in the theater, not when the film print was created, as was the case with live organ or piano performances in silent movie theaters.[200] So, ASCAP's traditional role collecting for performances heard in a public venue fit well with the advent of early talkies like *Don Juan* (1926) and *The Jazz Singer* (1927).

Third, the requirement that writers and publishers license the public performance of their music in movies coincident with the granting of synchronization rights is inextricably tied up in the resolution of complaints with ASCAP's practices identified in *Alden-Rochelle* and the resulting, binding changes to the ASCAP Consent Decree.[201] In other words, ASCAP agreed to

198. *Id.*

199. *Id.*

200. The transition was opposed by the musicians who were losing their jobs:

> The evil face of that campaign was the dastardly, maniacal robot. The Music Defense League spent over $500,000, running ads in newspapers throughout the United States and Canada. The ads pleaded with the public to demand humans play their music (be it in movie or stage theaters), rather than some cold, unseen machine. A typical ad read like this one from the September 2, 1930 *Syracuse Herald* in New York:
>
> > Tho' the Robot can make no music of himself, he can and does arrest the efforts of those who can.
> >
> > Manners mean nothing to this monstrous offspring of modern industrialism, as IT crowds Living Music out of the theatre spotlight.
> >
> > Though "music has charms to soothe the savage beast, to soften rocks or bend a knotted oak," it has no power to appease the Robot of Canned Music. Only the theatre-going public can do that.

Matt Novak, *Musicians Wage War Against Evil Robots*, SMITHSONIAN.COM (Feb. 10, 2012), http://blogs.smithsonianmag.com/paleofuture/2012/02/musicians-wage-war-against-evil-robots/.

201. Films released to theaters before *Alden-Rochelle* and the Consent Decree modification were exhibited under ASCAP licenses paid by theaters and presumably had no public performance license granted should the movie be re-released to theaters after 1950. Outside of a few classics, however, pre-1950s films were not re-released to theaters. While *Alden-Rochelle* did not directly outlaw ASCAP's separate licensing of performance rights, the court viewed the practice with skepticism and led to the combined synch-public performance license, which was common practice prior to the case for producer's licensed musical compositions that were not in the ASCAP repertoire:

cover performances of film music in theaters to settle other liability questions.

As is often the case with consent decrees, however, what is not excluded is arguably permitted; television in particular was not covered. ASCAP began to license broadcast networks and local TV stations as a result.[202] One might ask why broadcasters did not seek to extend the *Alden-Rochelle* concept of requiring performance licenses to be issued at the time the synchronization license is granted. There may have been several reasons. For one thing, a lot of early television was live and there was no significant output of audio-visual works where synchronization licenses were required. Much of the available filmed programming was low-budget and did not rely on well-known (and pricey) musical accompaniment.[203] For another, major TV networks and station owners were radio station owners as well[204] and were used to paying

> Unquestionably it would be a simpler and a proper arrangement for the owner of the copyright to deal directly with the producer on both the synchronization rights and the performing rights, and thus have the motion picture producer acquire both rights at the same time, so that he in turn could rent the film without requiring the exhibitor to obtain the performance rights from Ascap. But that in some way the value of the performing rights would be claimed by the copyright owner and eventually would be passed on to the exhibitor, I have no doubt at all. The ultimate result would be that the exhibitor would not be separately charged for the performance rights, as he now is through Ascap, but he would be charged for those rights in the total rental he would pay for the film.

Alden-Rochelle v. ASCAP, 80 F. Supp. 888 (S.D.N.Y. 1948), *amended*, 80 F. Supp. 900 (S.D.N.Y. 1948).

202. *History*, *supra* note 10 ("The first ASCAP television licenses were negotiated in the 1940s. ASCAP initially offered free licenses to television broadcasters.").

203. *See* Jim Davidson, *Television Transmission Methods for Dummies or "Is It Live or Is It Kinescope?,"* *CLASSIC TV* INFO (May 15, 2005), http://www.classictvinfo.com/Essays/TVTransmission.htm. Davidson states:

> A live broadcast was, of course, the most basic method of disseminating a television broadcast. It required nothing more than putting actors in front of a camera and transmitting their images to viewers' TV sets. Post-production was non-existent because there was no film to edit. The lack of post-production made live cheaper than film, which was important in the days when budgets were low because the bulk of advertising dollars hadn't yet migrated from radio to TV.

Id.

204. While not all television pioneers came from radio, NBC and CBS—the leading radio networks and station owners—were among the first to experiment with the service. Christopher Anderson, *National Broadcasting Company*, MUSEUM BROADCAST COMM., http://www.museum.tv/archives/etv/N/htmlN/nationalbroa/nationalbroa.htm (last visited Mar. 16, 2013).

for public performance licenses for these transmissions. Furthermore, film library content made available to broadcasters early on was often old as the movie industry viewed television as a threat to its core theatre business rather than a part of a sequential distribution chain that today includes the Internet as well. The music rights granted by publishers to film producers did not include public performance licenses to their synchronized compositions for distribution by technologies "whether now known, or hereinafter invented," a phrase that is often included as boilerplate in grant of creative rights today.[205] From the producers' side, there was no incentive to reopen old agreements and agree to pay performance rights unless broadcasters refused to license their works.

Perhaps more importantly, the transaction costs that favor purchasing a blanket license, if reasonably priced, outweighed the years of litigation which would have slowed down the business plans of the new television medium. Broadcasters were more interested in priming the pump, with content that would attract viewers and TV set buyers, than dickering out the terms of a public performance license that would also cover live performances already suited to ASCAP's blanket license. Then, too, BMI was owned by radio broadcasters; its owners could hardly be blamed for consigning *Alden-Rochelle* to its unique time and history and promoting blanket licenses for television. The same consideration may have influenced online providers like YouTube and Yahoo, nascent technologies in this century, whose initial rate disputes with ASCAP were not over whether rates applied, but instead how much needed to be paid.[206]

> NBC began experimental broadcasts from New York's Empire State building as early as 1932. By 1935 the company was spending millions of dollars annually to fund television research. Profits from the lucrative NBC radio networks were routinely channeled into television research. In 1939 NBC became the first network to introduce regular television broadcasts with its inaugural telecast of the opening day ceremonies at the New York World's Fair of 1939. RCA's goal was to produce and market receivers and programs, to become the driving force in the emerging industry.

Id.

205. *See* Frederick C. Boucher, *Blanket Music Licensing and Local Television: An Historical Accident in Need of Reform*, 44 WASH. & LEE L. REV. 1157, 1166 (1987).

206. United States v. ASCAP (*In re* YouTube), 616 F. Supp. 2d 447 (S.D.N.Y. 2009); *RealNetworks and Yahoo! II*, 559 F. Supp. 2d 332 (S.D.N.Y. 2008), *rev'd in part RealNetworks and Yahoo! III*, 627 F.3d 64 (2d Cir. 2010), *cert denied*, 132 S. Ct. 366 (2011); *see also supra* text accompanying notes 18, 19.

For these reasons—and perhaps others[207]—when movies were distributed to television stations in the 1950s, agreements for synchronization licenses did not include a public performance license for television exhibition. PROs treated these transmissions (eventually extending to cable networks and now online) as public performances outside the grant of theater performance and sought an additional performance license payment, subject to the protection of the rate court.[208]

Once a subsequent method of exhibition of an audio-visual work is established and underway, however, there is no obvious reason why the performance rights of composers necessarily need to be licensed separately by the exhibiting medium. In other words, now that linear, VOD, and online performances are known methods of distribution, owners of audio-visual works can determine the value of performance and pay composers as part of the initial licensing process. This forward-looking, producer-centric scheme to determine economic benefits applies to other creative contributions to a film. A film is a derivative work of many separate copyrighted (and non-copyrighted, though creative) elements: the screenwriter's script, choreography, music, sound recordings and some visual elements. Film owners obtain rights to exhibit and perform these works in subsequent markets through guild agreements or bi-lateral negotiations. Payments for acting performances, while not separately copyrightable from their fixation in films, are also dealt with at the outset of a work with expected residuals or other income depending on the work's success.

In short, there are no PROs for script writers, directors, or other above-the-line talent[209] who rely on guild agreements on dickered terms, or any other contributors to an audio-visual work. If the distributor of an audio-

207. The ASCAP rate court, in place today to determine a reasonable price of a license in the absence of a negotiated rate, did not exist prior to the 1950 Consent Decree modification. Recall, it was ASCAP's proposed steep rate increase that caused theater owners to file an antitrust lawsuit against ASCAP. *See supra* notes 165–69.

208. While the first television public performance licenses were granted in the 1940s, the amount to be paid, and whether the license should be per-program or a blanket, has been much disputed by TV stations and networks. The broadcast industry created its "All-Industry Television Station Music License Committee" in 1949 to devise a licensing regime. For a discussion of the Committee's history and litigation, see *History*, *supra* note 10. Cable operators pay for public performances of content they locally perform (e.g., inserted advertising, locally produced programs) while public performance of music in content from satellite delivered networks are paid "through to the viewer" by the networks. *See also* BRENNER ET AL., *supra* note 13, § 9:34.

209. "Above the line" refers to costs associated with major creative talent, for example, lead actors, directors, producers, and writers. Films with expensive special effects (and few stars) have more "above the line" budget costs for technical aspects. "Below the line" refers to other production costs.

visual work (such as a studio) has the right to license content to a particular medium, its licensee (such as HBO) does not have to separately negotiate with all of the creative contributors, (save for the one exception of composers). Instead it looks to the film owner to make those payment arrangements, and its license fee will cover those payments. The ASCAP rate court recognized that the PRO approach is an exception to production compensation schemes in *In re MobiTV*, noting that "pricing the public performance right at the time the content is first sold gives direct and immediate feedback to content producers about the value of a component of their product."[210]

But as observed, when films move to television, there are significant transition issues that make it difficult to implement the rate court's reasonable suggestion. First, there are disincentives for the composer/publisher's side and the producer's side to add public performance rights to a synchronization agreement. The owner of a composition in high demand for a film can command a higher synchronization license and, through PROs, carve out the performance license for later payments. Plus, it will not have to determine the present discounted value of a performance license that might be needed for media yet to be invented. It is not surprising that publishers and composers would rather rely on the PRO collection and distribution process, whatever its shortcomings in identifying and rewarding particular music use, than to try to do it as a standalone publisher. For its part, the producer has no interest in paying for public performance licenses if those costs can be sloughed off to future distributors like broadcast, cable, and online. So, no one has much incentive to determine the public performance license fee for compositions used in audio-visual works at the time all other rights are negotiated.

Second, restructuring the process for reasons short of a government fiat is unlikely. It took a court decision and a Consent Decree modification to bring the movie theater public performance rights into the production budget. That result covered one kind of exhibition for what parties must have believed would not have been, in most cases, a lengthy term. The duration of theatrical exhibition of a movie is minute compared to the possibly perpetual, so-called long tail availability and performance (public or not) of that content online.[211]

210. *In re* MobiTV, 712 F. Supp. 2d 206, 246 (S.D.N.Y. 2010).

211. Chris Anderson, THE LONG TAIL: WHY THE FUTURE OF BUSINESS IS SELLING LESS OF MORE (2006).

VII. CONCLUSION

This brief history points to two observations about public performance rights and online video. First, PROs are beneficial to creating efficiencies in some aspects of public performances. PROs filled an inevitable need to deal with myriad music performances that occur publicly in clubs, theaters, restaurants, and other venues for which the transaction costs of obtaining a license are prohibitive. And blanket licenses work for linear transmissions. Radio or Internet airplay of compositions could be tied only to songs for which a direct license is obtained. But, the utility of dealing with three PROs (on a blanket or per-song basis) instead of thousands of publisher contacts is desirable.

Second, logic would suggest a different path for public performance rights for audio-visual works. When an audio-visual work is created, all other intellectual property interests for their subsequent public performances seem to be managed by licensing. When new, not contemplated uses arise, as with video cassettes or web players, guilds like the Writers Guild of America set fees with producers, not with each sequential distributor. Only composers, whose music is part of the production, create a separate performance pay window which every distributor must face.

No doubt, the current arrangement of separately-negotiated public performance rights through PROs is likely to benefit a publisher. Having a separate, sequential negotiation for each public performance enhances a composer's ability to analyze values and to collect more on subsequent uses, relying on the expertise and heft of PROs. Every new distribution window is a source of new performance revenues. Guild agreements for writers, actors, or directors may not cover such performances or lag behind new media. The net result for song owners may exceed what would be obtained through the kinds of front-loaded arrangements other creative contributors like screenwriters make at the beginning of a project. It also allows composers to test their entitlement to public performance payments in new, uncharted rights areas, such as streaming, that may not be spelled out at the time the audio-visual work is created. This is the season of litigation we are currently in. As the lyric in this Article's title advises, when traveling by boat downstream, anything more than gentle rowing can upset the enterprise.

A change to an ex ante system—where a producer negotiates all public performance licenses up front—would be difficult, unlikely to occur voluntarily, and perhaps not worth the commotion to a market beset with technological disruptions. And, all of the extant millions of audio-visual works for which no combined synchronization/public performance fees have been set would still need PRO licenses. Plus, PROs are here to stay, necessary whenever public performances occur or audio works are

performed on the radio or Internet. So it may simply be too much to suggest that we transition to a license-at-the-outset regime, however logical it may be. The cost of transition is too formidable.

But it is worth observing, in trying to calculate what the performance rights are worth in streaming, that there is a less contentious, alternative scheme: namely, spelling out at the time the audio-visual work is created the payments due to publishers when that derivative work is publicly performed. PROs are now part of cyberspace, and the benefits of blanket licensing help promote new technology. But courts should be careful not to assume every performance online is public.

FINDING THE POINT OF NOVELTY IN SOFTWARE PATENTS

Bernard Chao[†]

ABSTRACT

The issue of patentable subject matter eligibility is in considerable flux. In 2012, the Supreme Court set forth a confusing new framework for determining patent eligibility. The decision in *Mayo v. Prometheus* cast serious doubt on the continued viability of many software patents. Indeed, a split quickly emerged in the Federal Circuit. As a result, it was unclear whether adding computer limitations to an otherwise unpatentable concept somehow renders the concept patent-eligible. In an attempt to settle this question, the Federal Circuit granted a petition to rehear the issue en banc. But in *CLS Bank Int'l v. Alice Corp.*, the judges could not find common ground and the decision contained seven separate opinions reflecting at least three distinct approaches. Thus, there remains a pressing need to find a common analytical framework for deciding software patent eligibility questions.

There is a way out of the current morass without departing from precedent. In *Mayo*, the Supreme Court implicitly revived long rejected point-of-novelty thinking. In an earlier essay, I expanded on that approach and offered a general framework for making patentable subject matter eligibility determinations. This Article applies this approach to software patents. Specifically, it explains that the key to determining whether a software patent covers eligible subject matter is assessing the strength of the connection between the patent's point of novelty and physical devices found in the other claim limitations. This test serves to rein in harmful business method software patents without affecting more deserving industrial patents. Thus, the test is justified from both doctrinal and policy perspectives.

---

† Assistant Professor, University of Denver Sturm College of Law. I would like to thank Brian Love, Kevin Emerson Collins, Viva Moffat, Dmitry Karshtedt, Harry Surden, Justin Pidot, John Soma, and Ian Farrell for their comments on earlier drafts of this Article.

TABLE OF CONTENTS

I. INTRODUCTION

Last year, in *Mayo v. Prometheus*, the Supreme Court made another attempt to define the scope of patentable subject matter.[1] A unanimous Supreme Court held that the personalized medicine dosing process invented by Prometheus Laboratories was not eligible for patent protection because the process was *effectively* an unpatentable law of nature.[2] Although the decision did not directly address software patents, it set forth a confusing framework for subject-matter patent eligibility that will apply to software patents.

In the wake of *Mayo*, the Federal Circuit has already issued two conflicting decisions on the eligibility of software patents.[3] Although both cases involved patents on business concepts implemented through software, the two decisions applied different approaches to patentability and arrived at different outcomes. On July 9, 2012, in *CLS Bank International v. Alice Corp.*, a

1. Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289 (2012). Just two years earlier, the Supreme Court decided *Bilski v. Kappos*, 130 S. Ct. 3218 (2010), a decision that ostensibly set out the rules for determining subject matter eligibility under § 101. *See also Mayo*, 132 S. Ct. at 1293 ("[L]aws of nature . . . are not patentable.") (internal quotation marks omitted) (citing *Bilski*, 130 S. Ct. at 3233–34). Section 101 provides that "[w]hoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title." 35 U.S.C. § 101 (2012).

2. *Mayo*, 132 S. Ct. at 1294.

3. CLS Bank Int'l v. Alice Corp. (*CLS Bank I*), 685 F.3d 1341 (Fed. Cir. 2012), *reh'g en banc granted, opinion vacated*, 484 F. App'x 559 (Fed. Cir. 2012), *aff'd en banc*, 717 F.3d 1269 (Fed. Cir. 2013); Bancorp Servs. v. Sun Life Assurance Co., 687 F.3d 1266 (Fed. Cir. 2012).

panel of the Federal Circuit found that patents covering a trading system platform for exchanging obligations contained patent-eligible subject matter.[4] Less than a month later, in *Bancorp Services v. Sun Life Assurance Co.*, a different panel found that patents covering a system for administering and tracking life insurance values were invalid because they covered an unpatentable abstract idea.[5] Unsurprisingly, the Federal Circuit decided to resolve this split and granted a petition for an en banc rehearing in *CLS Bank*.[6] Specifically, the court asked the parties: "[W]hat test should the court adopt to determine whether a computer-implemented invention is a patent ineligible 'abstract idea' . . . ?"[7]

This Article originally set out to respond to this question. However, shortly before it went to press, the Federal Circuit issued its en banc decision in *CLS Bank*.[8] Unfortunately, the seven separate opinions found in this split decision only added to the confusion.[9] Part V of this Article is a "postscript" describing these different views. Although a majority of seven judges found that the method and computer-readable claims at issue were not patent-eligible, the court split evenly (5-5) on the eligibility of the system claims.[10] Moreover, no majority could agree on a common analytical approach. Instead, the 135-page decision reflected at least three distinct analytical approaches.[11] The judges themselves characterized the decision as "irreconcilably fractured"[12] and "devoid of consensus."[13] Thus, there continues to be a pressing need to find a "consistent, cohesive, and accessible" framework for determining when software patents cover patent-eligible subject matter under § 101 of the Patent Act.[14]

4. *CLS Bank I*, 685 F.3d at 1356 (Linn, Prost & O'Malley, JJ.).

5. *Bancorp*, 687 F.3d at 1281 (Lourie, Prost & Wallach, JJ.).

6. CLS Bank Int'l v. Alice Corp. (*CLS Bank II*), 484 F. App'x 559 (order granting hearing en banc).

7. *Id.*

8. CLS Bank Int'l v. Alice Corp. (*CLS Bank III*), 717 F.3d 1269 (Fed. Cir. 2013) (en banc).

9. *See id.*

10. The result was that the district court's holding that none of the claims were drawn to eligible subject matter was affirmed. *CLS Bank III*, 717 F.3d at 1273 (per curiam).

11. *See* discussion *infra* Part V (discussing the opinions of Judge Lourie (concurring), Chief Judge Rader (concurring in part and dissenting in part), and Judge Newman (concurring in part and dissenting in part)).

12. *CLS Bank III*, 717 F.3d at 1314. (Moore, J., dissenting in part).

13. *Id.* at 1321 (Newman, J., concurring in part and dissenting in part).

14. *Id.* at 1277 (Lourie, J., concurring) (discussing the need for a workable approach in § 101 jurisprudence).

Building on earlier work,[15] this Article attempts to provide that framework. The suggested approach does not attempt to make any fundamental changes to § 101 in hopes of finding the "best" solution for patent law. Rather, the proposed approach seeks to provide a practical and coherent framework that sensibly brings the doctrine of subject-matter patentability as applied to software patents in line with the *Mayo* decision.

Laws of nature, natural phenomena, and abstract ideas are not patentable.[16] But identifying when a patent covers one of these unpatentable concepts (as opposed to an application of such a concept—which is patentable[17]) has proven to be quite difficult. The Federal Circuit panel in *Mayo* dissected the claims of Prometheus' personalized medicine patent and held that they did not add "enough" to an unpatentable law of nature to gain patent protection.[18] Unfortunately, the court never explained what, exactly, would be "enough." Since many patents involve unpatentable concepts to some extent, *Mayo* exposed a host of seemingly uncontroversial patents, including many software patents, to attacks on patent-eligibility grounds.[19]

The Supreme Court used a kind of point-of-novelty analysis in *Mayo* by focusing on what limitations were added to the law of nature at the heart of Prometheus' patents.[20] This hearkens back to the reasoning used years ago in *Parker v. Flook*.[21] In *Flook*, the Supreme Court treated the unpatentable formula that lay at the heart of Flook's patent as if it were in the prior art.[22] But once that determination had been made, the application did not contain any patentable invention and the Court concluded that it did not cover patent-eligible subject matter.[23] However, the Court later rejected this approach in *Diamond v. Diehr* when it said that "[i]n determining the eligibility

15. Bernard Chao, *Moderating* Mayo, 107 NW. U. L. REV. 423 (2012).

16. Diamond v. Diehr, 450 U.S. 175, 185 (1981).

17. Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289, 1293–94 (2012).

18. *Id.* at 1297 (emphasis omitted).

19. Michael J. Malecek & Kenneth M. Maikish, *The* Prometheus *Effect on Software Patents*, 24 NO. 6 INTELL. PROP. & TECH. L.J. 3, 3, 7 (2012) (arguing that the reasoning in *Mayo* suggests that software patents containing a mental step are not directed towards patentable subject matter); Tony Dutra, *Computer, Medical Diagnostics, Gene Patents At Risk in Light of* Mayo*, Panelists Contend*, PAT., TRADEMARK & COPYRIGHT L. DAILY (Apr. 4, 2012) ("[Intel's Tina] Chappell predicted that the court would view the algorithms that are typically cited in software patents in the same way that it analyzed the law of nature in medical diagnostics in *Mayo*.").

20. *See infra* text accompanying notes 113–23.

21. Parker v. Flook, 437 U.S. 584, 590 (1978).

22. *Id.* at 594.

23. *Id.* at 594–95.

. . . for patent protection under § 101, . . . claims must be considered as a whole."[24]

In an earlier essay, *Moderating* Mayo, I argued that *Mayo* should be interpreted as reviving a point-of-novelty approach.[25] Although the *Mayo* decision clearly reflected this perspective, the Supreme Court did not provide any test for lower courts to apply. My essay filled this void by offering a point-of-novelty test different from the rejected *Flook* test. This new point-of-novelty test follows from both *Diehr* and *Mayo* by considering the point of novelty in the context of the claim as a whole.[26] Assuming that an otherwise unpatentable concept lies at the patent's point of novelty, this two-part test explains when other claim limitations add "enough" to the unpatentable concept to make it patent-eligible.[27]

Although I previously explained how my test applied to different variations of the medical diagnostics technology found in *Mayo*,[28] I have not explained how it would apply to software patents. The revised point-of-novelty approach can also work in this context. The point of novelty of many software patents is a mathematical formula or abstract idea.[29] In an attempt to minimize patent eligibility concerns, patent attorneys typically draft software claims so that the idea is connected to a physical device.[30] Under the new point-of-novelty approach, that tactic should only be effective for certain kinds of patents. Some of the ideas underlying software patents are bound together with the physical components; for example, when a patent claims a novel algorithm for curing rubber products, both the formula and the physical components are necessary to accomplish the invention's goals.[31] Without the physical device, the formula could not achieve the goal of the invention. Moreover, it makes no sense to discuss the formula apart from the physical devices used to implement it. Thus, there is a sufficiently strong

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

25. Chao, *supra* note 15, at 432–33.

26. *Id.* at 436.

27. *Id.*; *see infra* text accompanying note 182 (discussing the basic modified point-of-novelty test).

28. Chao, *supra* note 15, at 436–40.

29. Mathematical formulas are a type of law of nature that has frequently arisen in software patent cases. *See, e.g.*, *Diehr*, 450 U.S. at 187; Parker v. Flook, 437 U.S. 584, 590 (1978); Gottschalk v. Benson, 409 U.S. 63, 71 (1972). In addition, many of the recent disputes before the Federal Circuit have centered on abstract ideas. *See infra* text accompanying notes 93–98 and 139–63.

30. *See infra* note 186.

31. *See infra* Part II.

nexus between the idea and the device such that the subject matter should be patent-eligible.[32]

However, in many other cases, the idea underlying the software patent lacks a strong nexus to the device.[33] One example of such a patent is the method for administering life insurance values claimed in *Bancorp Services*.[34] Although the drafting of the patent attempts to establish a connection to a computer, one is not actually required; the idea of administering life insurance policy values makes perfect sense standing alone.[35] Thus, the nexus between the physical components and the idea is weak. In such cases, attaching the idea to a machine should not be enough to make the concept patentable.

The rubber-curing and life insurance policy administration examples illustrate that the key to determining whether a software patent covers eligible subject matter is assessing the strength of the nexus between the patent's point of novelty and the physical devices found in the other claim limitations. Bits and pieces of this theory are scattered throughout both Supreme Court and Federal Circuit precedent,[36] but neither court has fully appreciated the point-of-novelty approach to subject matter patent eligibility. When the approach is finally appreciated, patent law will finally have a practical tool for distinguishing questionable business method patents from other kinds of more deserving industrial software patents.[37]

In Part II, this Article describes the different types of software patents. At one end of the spectrum are software patents that are little more than business method patents. An example is the controversial Amazon one-click patent (click only once to buy).[38] Business method software patents have been the subject of intense criticism and are often thought to burden innovation. At the other end of the spectrum are industrial software patents. From a policy perspective, these patents are indistinguishable from other

32. *See Diehr*, 450 U.S. at 191–93.

33. *See infra* Sections III.A, III.C.

34. Bancorp Servs. v. Sun Life Assurance Co., 687 F.3d 1266, 1279 (Fed. Cir. 2012) ("[T]he claims merely employ computers to track, reconcile, and administer a life insurance policy with a stable value component—*i.e.*, the computer simply performs more efficiently what could otherwise be accomplished manually.").

35. *See id.* at 1275.

36. *See infra* text accompanying notes 187–88.

37. As Brian Love suggests, this may be the "least bad" option for dealing with problematic software patents. Brian J. Love, *Why Patentable Subject Matter Matters for Software*, 81 GEO. WASH. L. REV. ARGUENDO 1, 8–11 (2012).

38. David Orozco, *Administrative Patent Levers*, 117 PENN. ST. L. REV. 1, 22–23 n.119 (2012) (discussing Free Software Foundation's boycott in response to Amazon's assertion of the one-click patent).

industrial patents that are not implemented through software and these patents have not been subject to the same criticism as their business method cousins.

In Part III, this Article describes the two most recent Supreme Court decisions on patent-eligible subject matter, *Bilski*[39] and *Mayo*,[40] and the subsequent Federal Circuit decisions on software patents. *Bilski* is important for two reasons. First, it endorsed (for the most part) the machine-or-transformation test the Federal Circuit had previously adopted.[41] Second, although *Bilski* did not categorically reject business method patents, the decision demonstrated a strong hostility towards them.[42] *Mayo* is important because the Supreme Court took a new tack and looked at subject matter eligibility determinations from a point-of-novelty perspective.[43] Unfortunately, these decisions have not yielded any clarity for software patents. As Part III describes, different Federal Circuit panels have applied different tests for determining when software patents are drawn to eligible subject matter.

In Part IV, this Article describes the new point-of-novelty test that I offered in my earlier work and explains how the test can be applied to software patents. More specifically, my test concludes that software patents are drawn to patent-eligible subject matter when the physical limitations (e.g., rubber molding machines) are bound together with—and necessary to—the unpatentable concepts that lie at the patent's point of novelty. To be clear, I do not suggest that this view is the one I would take if I were given a clean slate. Others have already proposed idealized solutions.[44] But patent applicants, examiners, litigants, and the lower courts need greater clarity now.[45] The goal of this Article is to find a realistic path out of the current morass. Thus, the test described here is intended to be a practical solution that works within the constraints of current Supreme Court jurisprudence, satisfiying that jurisprudence in two important respects. First, the test should

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

40. Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289 (2012).

41. *See Bilski*, 130 S. Ct. at 3221.

42. *See id.* at 3257.

43. *See Mayo*, 132 S. Ct. at 1304.

44. *See, e.g.*, Mark A. Lemley, Michael Risch, Ted M. Sichelman & R. Polk Wagner, *Life After* Bilski, 63 STAN. L. REV. 1315, 1339–41 (2011) (proposing five factors for a scope-based § 101 determination); Kevin Emerson Collins, Bilski *and the Ambiguity of "An Unpatentable Abstract Idea,"* 15 LEWIS & CLARK L. REV. 37 (2011) (proposing a more precise framework for the exclusion of abstract ideas from patent eligibility).

45. *See CLS Bank III*, 717 F.3d 1269, 1314 (Fed. Cir. 2013) (Moore, J., dissenting in part) (noting that many other cases dealing with the patent eligibility of software patents are pending in the Federal Circuit and district courts).

achieve the Supreme Court's desired results by eliminating most business method software patents while retaining industrial software patents. Second, the test also operates within the Supreme Court's theoretical framework by applying a point-of-novelty approach.

Finally, Part V is a "postscript" that describes the latest morass created by the Federal Circuit in CLS Bank. The decision contains at least three different analytical approaches. This Article labels the two primary approaches as the "strong view" and "weak view" of § 101's patent eligibility requirement. There is also a third approach advocated by Judge Newman alone that would substantially abandon the use of § 101 to determine patent eligibility. Part V explains how the strong view is the only approach that faithfully follows the recent Supreme Court decisions. The two other approaches either implicitly (in the case of the weak view) or explicitly (in Judge Newman's opinion) reject the path the Supreme Court has taken. Assuming that the Supreme Court will not suddenly reverse itself, Part V explains why the point-of-novelty approach described here provides more clarity than the strong view.

II. CATEGORIES OF SOFTWARE PATENTS

Numerous commentators have been critical of software patents, arguing that software patents discourage innovation,[46] have unclear boundaries,[47] and are of low quality.[48] According to a recent empirical study, software patents include some of the most litigated patents, but on the whole are much less

46. *E.g.*, FEDERAL TRADE COMMISSION, TO PROMOTE INNOVATION: THE PROPER BALANCE OF COMPETITION AND PATENT LAW AND POLICY 56 (2003), *available at* http://www.ftc.gov/os/2003/10/innovationrpt.pdf ("Many panelists and participants expressed the view that software and Internet patents are impeding innovation."); *cf.* Stuart J.H. Graham, Robert P. Merges, Pam Samuelson & Ted Sichelman, *High Technology Entrepreneurs and the Patent System: Results of the 2008 Berkeley Patent Survey*, 24 BERKELEY TECH. L.J. 1255, 1262, 1289–90 (2009) (finding in a survey of start-up companies that (1) the first mover advantage, not patent protection, was the most "important" means to "capture competitive advantage" in the software industry; and (2) the majority of start-up companies in the software industry hold no patents at all).

47. *E.g.*, Peter S. Menell, *A Method for Reforming the Patent System*, 13 MICH. TELECOMM. & TECH. L. REV. 487, 505–06 (2007) ("The boundaries of software and business method patents are inherently ambiguous.").

48. *E.g.*, Love, *supra* note 37, at 8–9 (arguing that examiners allow "many overbroad software applications to issue as patents"). *But see* John R. Allison & Ronald J. Mann, *Disputed Quality of Software Patents*, 85 WASH. U. L. REV. 297 (2007) (discussing an empirical study which suggests that, with respect to disclosure of prior art, the quality of software patents is not worse than the quality of patents in other fields).

likely than other types of patents to be found valid and infringed.[49] This suggests that software patents take up disproportionate litigation resources, while offering only a slight benefit in return.[50] This may be because software patents are one of the weapons of choice for non-practicing patent entities.[51] Commentators have widely criticized the many patent lawsuits brought by non-practicing entities for focusing resources on litigation instead of innovation.[52] In 2011, Congress even enacted a temporary program for challenging the validity of business method patents.[53]

While these criticisms of software patents as a category are valid in some cases, there are many different kinds of software patents.[54] One kind is the business method patent. Business method patents like those in *CLS Bank* (covering a trading system platform for exchanging obligations)[55] and *Bancorp Services* (covering a system for administering and tracking life insurance values) lie at the heart of the controversy.[56] Many of the claims in both cases explicitly contained computer-based limitations or have been interpreted to

49. John R. Allison, Mark A. Lemley & Joshua Walker, *Patent Quality and Settlement Among Repeat Patent Litigants*, 99 GEO. L.J. 677, 707–08 (2011).

50. *Id.* at 708.

51. Brian J. Love, *An Empirical Study of Patent Litigation Timing: Could a Patent Term Reduction Decimate Trolls Without Harming Innovators?*, 161 U. PA. L. REV. 1309, 1343 (2013) (finding that about 40% of assertions by non-practicing entities were brought to enforce software patents).

52. *See, e.g.*, President's Council of Economic Advisers, the National Economic Council, and the Office of Science & Technology Policy, *Patent Assertion and U.S. Innovation* 5 (2013), *available at* http://www.whitehouse.gov/sites/default/files/docs/patent_report.pdf (discussing the relationship between patent assertion entities and software patents); Ted Sichelman, *Commercializing Patents*, 62 STAN. L. REV. 341, 368 (2010) (saying that non-practicing entities stifle commercialization of patented inventions by "exploit[ing] litigation and licensing market defects to extract unwarranted rents").

53. Pub. L. No. 112-29, § 18, 125 Stat. 284, 329–31 (2011).

54. *Compare* Allison & Mann, *supra* note 48, at 308–09 (defining a software patent broadly as "one in which at least one claim element covers data processing—that is, the act of manipulating data—regardless of whether the code carrying out that data processing is on a magnetic storage medium or embedded in a chip"), *with* Stuart J.H. Graham & David C. Mowery, *Intellectual Property Protection in the U.S. Software Industry*, *in* PATENTS IN THE KNOWLEDGE-BASED ECONOMY 219, 232 (Wesley M. Cohen & Stephen A. Merrill, eds., Committee on Intellectual Property Rights in the Knowledge-Based Economy, National Research Council, 2003) (identifying software patents based on the industry characteristics; this results in a narrower set).

55. Terry Baynes, *Federal Circuit finds business method patentable*, THOMSON REUTERS NEWS & INSIGHT, July 9, 2012, *available at* http://newsandinsight.thomsonreuters.com/New_York/News/2012/07_-_July/Federal_Circuit_finds_business_method_patentable (describing the patents in CLS Bank as business method patents).

56. *CLS Bank I*, 685 F.3d 1341, 1343 (Fed. Cir. 2012), *reh'g en banc granted, opinion vacated*, 484 F. App'x 559 (Fed. Cir. 2012), *aff'd en banc*, 717 F.3d 1269 (Fed. Cir. 2013); Bancorp Servs. v. Sun Life Assurance Co., 687 F.3d 1266, 1269 (Fed. Cir. 2012).

require implementation on a computer.[57] The true focus of these patents, however, is on the business method itself.

Another kind of software patent is the industrial method patent.[58] For example, the patents in both *Parker v. Flook*[59] and *Diamond v. Diehr*[60] used new algorithms in industrial applications. In *Flook*, the claims involved a formula for calculating an alarm limit for a catalytic chemical conversion of hydrocarbons.[61] The patent in *Diehr* used a mathematical equation to develop a new process for molding and curing raw rubber into products.[62] Both patents were implemented with software.[63]

Of course, not every software patent can be easily classified as either a business method patent or an industrial patent. Some patents, for instance, operate solely on computers, like most business method software patents, but also improve the performance of a physical machine (i.e., the computer), like most industrial software patents. The patents in *Research Corporation Technologies, Inc. v. Microsoft* fall within this middle ground.[64] The patents in that case covered a particular method of digital image halftoning, which allowed computers to present many shades and color tones using only a limited number of pixel colors.[65] The claims did not require any physical device other than a computer. This might suggest that the patents were business method software patents. However, the technology was used to improve images displayed on printers and displays.[66] Thus, they could also be thought of as industrial software patents. Ultimately, any rule the Federal Circuit issues concerning patentable subject matter must be able to address all types of software patents, including those that do not fit easily into a discrete category.

---

57. *CLS Bank I*, 685 F.3d at 1344; *Bancorp Servs.*, 687 F.3d at 1271.

58. There is similar concept in Europe where such software patents are said to have "technical effect." *See* Patrick E. King, Ryan M. Roberts & Andrew V. Moshirnia, *The Confluence of European Activism and American Minimalism: Patentable Subject Matter After* Bilski, 27 SANTA CLARA COMPUTER & HIGH TECH. L.J. 247, 258–59 (2011) (discussing interpretations of patentability requirements under Article 52 of the European Patent Convention).

59. Parker v. Flook, 437 U.S. 584, 585 (1978).

60. Diamond v. Diehr, 450 U.S. 175, 187 (1981).

61. *Flook*, 437 U.S. at 585.

62. *Diehr*, 450 U.S. at 187.

63. *Flook*, 437 U.S. at 586; *Diehr*, 450 U.S. at 177.

64. Research Corp. Techs. Inc., v. Microsoft Corp., 627 F.3d 859, 862–63 (Fed. Cir. 2010).

65. *Id.* This technology is used to enhance the images found on computer displays and printers.

66. Notably, the claims at issue did not actually require the printers or display devices. See, for example, claim 1 of the '310 patent. *Id.* at 865.

There are a variety of reasons why many of the criticisms of software patents generally do not apply well to industrial software patents. First, the boundaries of these patents are not as amorphous as those of many business method software patents because the scopes of most industrial software patents are tied to particular applications.[67] Second, conventional wisdom suggests that the non-practicing entity problem is greater with respect to business method patents than to industrial software patents.[68] Third, many of the critiques of software patents focus on patents held by software companies; these companies, however, typically do not hold industrial software patents.[69]

As a practical matter, the Federal Circuit was unlikely to declare in its en banc review of *CLS Bank* that industrial software patents are per se ineligible. Such a rule would disturb the settled expectations of too many industries that rely on industrial software patents to protect their intellectual property rights.[70] Additionally, a decision eliminating industrial software patents could violate the United States' obligations under international law.[71] Thus, the only real question on the table for the Federal Circuit in *CLS Bank* was whether to rule that some software patents are ineligible even when the claims tie the

67. Even the Supreme Court has expressed concern about "vagueness and suspect validity of some of these [business method] patents." eBay Inc. v. MercExchange LLC, 547 U.S. 388, 396 (2006) (Kennedy, J., concurring).

68. *See* Orozco, *supra* note 38, at 15–23 (discussing the problem of business method patent assertions by non-practicing entities). *But see* Michael Risch, *Patent Troll Myths*, SETON HALL L. REV. 457, 477 (2012) (suggesting that "business methods are a relatively small part of NPE litigation, perhaps smaller than conventional wisdom might assume").

69. This is not surprising given how software patents are classified. *See* Graham et al., *supra* note 46, at 1268–69 n.41, 1271 n.46 (selecting primarily software companies for the authors' survey sample). In a different article, Graham and Mowery define software patents to actually exclude "embedded software" that is directly incorporated into a product and whose operation is typically not controlled by the user. Graham & Mowery, *supra* note 54, at 235–36. Of course these are typically industrial software patents.

70. Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co., 535 U.S. 722, 739 (2002) (explaining that a fundamental change to patent law could "risk destroying the legitimate expectations of inventors in their property").

71. *See* Eric Keller, *Time-Varying Compulsory License: Facilitating License Negotiation for Efficient Post-Verdict Patent Infringement*, 16 TEX. INTELL. PROP. L.J. 427, 439 ("Solutions that discriminate in the protection of patent rights based on 'field of technology' may also run afoul of treaty obligations under TRIPS Article 27.1."); Agreement on Trade-Related Aspects of Intellectual Property Rights, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, Legal Instruments-Results of the Uruguay Round vol. 31, 33 I.L.M. 1197 (1994) (subject to certain permissible exceptions, "patents shall be available for any inventions, whether products or processes, in all fields of technology, provided that they are new, involve an inventive step, and are capable of industrial application").

software to a computer. Unfortunately, the splintered decision provided no helpful guidance whatsoever.[72]

The point-of-novelty approach described in Part IV naturally distinguishes between industrial and business method patents and offers a framework that suggests that most business method software patents are ineligible for patent protection. The proposal is not intended to be an ideal solution divorced from reality. Rather, it draws upon existing concepts found in both Supreme Court and Federal Circuit precedent to create a test for patent eligibility that meets the goals of the courts and commentators alike. But before providing the details of this approach, Part III describes how recent case law has addressed subject matter patent eligibility, particularly as applied to software patents.

III. THE FRACTURED JURISPRUDENCE

The two most recent Supreme Court cases on subject matter patent eligibility place very different constraints on how the Federal Circuit must think about software patents. *Bilski* offers the machine-or-transformation test as one possible test for analyzing subject matter patent eligibility.[73] But just as importantly, *Bilski* shows a strong hostility towards business method patents, albeit for different reasons.[74] In contrast, *Mayo* says nothing about business method patents, but offers an entirely different analytical approach, which implicitly requires a point-of-novelty framework.[75] The decision from the Federal Circuit in *CLS Bank* had to account for these two different strands of thinking.

A. *BILSKI V. KAPPOS*

In *Bilski*, the Supreme Court evaluated the patent eligibility of a procedure for instructing buyers and sellers on how to protect against the risk of price fluctuations in a discrete section of the economy.[76] Although the Court unanimously concluded that Bilski's claims did not cover patent-eligible subject matter, there was significant disagreement about how to reach that conclusion.

The Court considered two proposed limitations under § 101: the machine-or-transformation test and the categorical exclusion of business method patents. Writing the opinion of the Court, Justice Kennedy first

72. *See infra* Part V (discussing the different opinions from *CLS Bank Int'l v. Alice Corp.*).
73. Bilski v. Kappos, 130 S. Ct. 3218, 3227 (2010).
74. *See id.* at 3228–29.
75. Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289 (2012).
76. *Bilski*, 130 S. Ct. at 3223–24.

addressed the machine-or-transformation test.[77] The underlying Federal Circuit decision had held that the machine-or-transformation test was the sole test for determining the patentability of a "process" under § 101.[78] In other words, a process was only patentable if it was tied to a particular machine or transformed an article to another state.[79] The Supreme Court decision modified that holding, finding that the machine-or-transformation test may be "a useful and important clue" or "investigative tool," but it is "not the sole test for deciding whether an invention is a patent-eligible 'process' " under § 101.[80]

Speaking for only four members of the Court, Justice Kennedy's opinion recognized that it was unclear how the machine-or-transformation test might apply to software patents.[81] On the one hand, "[t]he machine-or-transformation test may well provide a sufficient basis for evaluating processes similar to those in the Industrial Age—for example, inventions grounded in a physical or other tangible form."[82] On the other hand, Justice Kennedy recognized that the machine-or-transformation test "would create uncertainty as to the patentability of software, advanced diagnostic medicine techniques, and inventions based on linear programming, data compression, and the manipulation of digital signals."[83] In fact, Justice Kennedy went out of his way to say that he was "not commenting on the patentability of any particular invention, let alone holding that any of the above-mentioned technologies from the Information Age should or should not receive patent protection."[84]

While the entire Court agreed that the machine-or-transformation test was not an exclusive test, the justices differed sharply on the eligibility of business method patents. The majority held that § 101 does not categorically exclude business method patents.[85] Indeed, the opinion questioned whether there was even a common understanding of the term "business method patents."[86] Four members of the Court disagreed, arguing that business

77. Justices Roberts, Thomas, and Alito joined the opinion in full while Justice Scalia only joined part of the opinion. *Id.* at 3223.

78. *In re* Bilski, 545 F.3d 943, 954 (Fed. Cir. 2008) (en banc).

79. *Id.* at 956.

80. *Bilski*, 130 S. Ct. at 3227.

81. Justice Scalia did not join in this part of Justice Kennedy's opinion. *Id.* at 3223.

82. *Id.* at 3227.

83. *Id.*

84. *Id.* at 3228.

85. *Id.* at 3227. Justices Roberts, Thomas, Alito and Scalia joined Justice Kennedy in this part of the opinion. *Id.* at 3223.

86. *See id.* at 3228 ("Nor is it clear how far a prohibition on business method patents would reach, and whether it would exclude technologies for conducting a business more

methods are categorically unpatentable. Relying chiefly on a lengthy historical analysis, Justice Stevens, joined by Justices Ginsburg, Breyer, and Sotomayor, argued that Bilski's "method is not a 'process' [under § 101] because it describes only a general method of engaging in business transactions—and business methods are not patentable."[87] Although Stevens did not define what a business method patent was, his opinion provided plenty of examples, including insuring against loss by bad debt, a method of abbreviating rail tariff schedules, the cafeteria system for transacting a restaurant business, and a diaper service.[88]

After *Bilski*, it appeared that business method patents had survived, but just barely. Four Justices would have categorically excluded business method patents.[89] Moreover, even Justice Kennedy's opinion explicitly left the door open for further restrictions on business method patents.[90] Since most business method patents fall within the category of software patents,[91] it is not surprising that the Federal Circuit has given software patents more scrutiny.

Soon after *Bilski,* the Federal Circuit issued three inconsistent decisions on the patent eligibility of business method software patents. In *Ultramercial v. Hulu*, the court faced a § 101 challenge to a patent claiming "a method for distributing copyrighted products (e.g., songs, movies, books) over the Internet."[92] The decision characterized the underlying idea as using advertising as currency.[93] This idea was admittedly abstract, but the Federal Circuit noted that the claimed steps were "likely to require intricate and

efficiently."); *see, e.g.*, Bronwyn H. Hall, *Business and Financial Method Patents, Innovation, and Policy*, 56 SCOT. J. POL. ECON. 443, 445 (2009) ("There is no precise definition of business method patents.").

87. *Bilski*, 130 S. Ct. at 3232.

88. *Id.* at 3246–48.

89. *Id.* at 3257.

90. *Id.* at 3231. The Court stated:

> It may be that the Court of Appeals thought it needed to make the machine-or-transformation test exclusive precisely because its case law had not adequately identified less extreme means of restricting business method patents . . . . In disapproving an exclusive machine-or-transformation test, we by no means foreclose the Federal Circuit's development of other limiting criteria that further the purposes of the Patent Act and are not inconsistent with its text.

Id.

91. JAMES BESSEN & MICHAEL J. MEURER, PATENT FAILURE: HOW JUDGES, BUREAUCRATS, AND LAWYERS PUT INNOVATORS AT RISK 187 (2008).

92. Ultramercial, LLC v. Hulu, LLC, 657 F.3d 1323 (Fed. Cir. 2011), *vacated sub nom.*, Wildtangent, Inc. v. Ultramercial, LLC, 132 S. Ct. 2431 (2012).

93. *Id.* at 1328, 1330.

complex computer programming" and could only be performed on the Internet.[94] The court held that because all the claims connected the underlying concept to a computer or the Internet, they were patent eligible.[95]

But two other decisions arrived at very different results. In *CyberSource v. Retail Decisions*, the Federal Circuit found that a patent related to a "method and system for detecting fraud in a credit card transaction between [a] consumer and a merchant over the internet" was not patent eligible.[96] The Federal Circuit reasoned that the ideas underlying the software claims were not sufficiently connected with their computer-based limitations to satisfy the machine-or-transformation test.[97] The Federal Circuit in *Dealertrack v. Huber* also suggested that adding computer limitations to a claim would not render every concept patent eligible.[98] In *Dealertrack*, the patents related to a computer-aided method and system for processing credit applications over electronic networks.[99] Even though some of the claims explicitly required the Internet, the Federal Circuit found that this recitation was insufficient because "the claims . . . recite[d] only that the method is 'computer aided' without specifying any level of involvement or detail."[100] Moreover, the court concluded that the claims preempted a fundamental concept.[101] Thus, *Dealertrack* appeared to expand the potential grounds for rejecting a software patent on subject matter patent eligibility grounds.

These decisions could be interpreted to suggest that software patents are drawn to eligible subject matter when the computer limitations are more complex, or simply have more steps. Indeed, a comparison of the central claims in these cases shows that *CyberSource* struck down a claim with three computer steps,[102] while *Ultramercial* upheld a claim with eleven computer

94. *Id.* at 1328.

95. *Id.*

96. CyberSource Corp. v. Retail Decisions, Inc., 654 F.3d 1366, 1367 (Fed. Cir. 2011).

97. *Id.* at 1375.

98. *See* Dealertrack, Inc. v. Huber, 674 F.3d 1315, 1333 (Fed. Cir. 2012).

99. *Id.* at 1317.

100. *Id.* at 1334.

101. *See id.* at 1333 ("Neither Dealertrack nor any other entity is entitled to wholly preempt the clearinghouse concept.").

102. CyberSource Corp. v. Retail Decisions, Inc., 654 F.3d 1366, 1370 (Fed. Cir. 2011). Claim 3 recites:

> A method for verifying the validity of a credit card transaction over the Internet comprising the steps of: a) obtaining information about other transactions that have utilized an Internet address that is identified with the [] credit card transaction; b) constructing a map of credit card numbers based upon the other transactions and; c) utilizing the map of credit card numbers to determine if the credit card transaction is valid.

Id.

steps.[103] But making subject matter eligibility determinations based on how many detailed (and often inconsequential) steps a patent attorney can draft makes little sense. Most observers simply viewed the decisions as inconsistent.[104] Thus, *Bilski* did not result in any clarity for standards of patentability for software.[105] *Mayo* was decided immediately on the heels of

---

103. Ultramercial, LLC v. Hulu, LLC, 657 F.3d 1323, 1324–25 (Fed. Cir. 2011) (discussing claim 1 of the '545 patent). Claim 1 recites:

> A method for distribution of products over the Internet via a facilitator, said method comprising the steps of: a first step of receiving, from a content provider, media products that are covered by intellectual property rights protection and are available for purchase, wherein each said media product being comprised of at least one of text data, music data, and video data; a second step of selecting a sponsor message to be associated with the media product, said sponsor message being selected from a plurality of sponsor messages, said second step including accessing an activity log to verify that the total number of times which the sponsor message has been previously presented is less than the number of transaction cycles contracted by the sponsor of the sponsor message; a third step of providing the media product for sale at an Internet website; a fourth step of restricting general public access to said media product; a fifth step of offering to a consumer access to the media product without charge to the consumer on the precondition that the consumer views the sponsor message; a sixth step of receiving from the consumer a request to view the sponsor message, wherein the consumer submits said request in response to being offered access to the media product; a seventh step of, in response to receiving the request from the consumer, facilitating the display of a sponsor message to the consumer; an eighth step of, if the sponsor message is not an interactive message, allowing said consumer access to said media product after said step of facilitating the display of said sponsor message; a ninth step of, if the sponsor message is an interactive message, presenting at least one query to the consumer and allowing said consumer access to said media product after receiving a response to said at least one query; a tenth step of recording the transaction event to the activity log, said tenth step including updating the total number of times the sponsor message has been presented; and an eleventh step of receiving payment from the sponsor of the sponsor message displayed.

Id.

104. *See, e.g.*, Recent Case, CyberSource Corp. v. Retail Decisions, Inc.*, 654 F.3d 1366 (Fed. Cir. 2011)*, 125 HARV. L. REV. 851, 857 (2012) (noting that the *Ultramercial* court's attempt to distinguish *Cybersource* seems forced); Kelly J. Kubasta, *Litigation Affecting Five Key Patent Law Areas*, *in* INTELLECTUAL PROPERTY LAW 2012 57, 62 (Eddie Fournier ed., 2012) ("Unfortunately, the result of *CyberSource* and *Ultramercial* is quite unclear and it remains uncertain as to where the line will be drawn as to software and methods as patent-eligible subject matter.").

105. *See* N. Scott Pierce, *A Great Invisible Crashing: The Rise And Fall Of Patent Eligibility Through* Mayo v. Prometheus, 23 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 186, 189–90 (2012) ("[T]here is little to guide . . . Federal Circuit cases that have issued since *Bilski* and

CyberSource, *Ultramercial*, and *Dealertrack*. Ideally, the Supreme Court would have addressed the Federal Circuit's conflicting case law and provided a framework for making subject matter eligibility determinations, particularly in the area of software patents. Unfortunately, *Mayo* just created more confusion.

B. *MAYO V. PROMETHEUS*

Although the technology in *Mayo* related to medical diagnostic testing,[106] the approach the Court laid out has significant implications for software patents.[107] The inventors of Prometheus' claimed methods discovered a specific correlation between the levels of metabolized drug in the body and the optimal drug dosage.[108] Two patents were issued on this discovery, both claiming a method for determining the level of the metabolized drug in a subject and informing a doctor to adjust the dosage within specific parameters.[109] The defendants argued that the claims were not drawn to patent-eligible subject matter as required by § 101, and this issue eventually made its way to the Supreme Court.[110]

A unanimous Court held that Prometheus' claims were not patent eligible.[111] The decision first noted that "Prometheus' patents set forth laws of nature—namely, relationships between concentrations of certain metabolites in the blood and the likelihood that a dosage of a thiopurine drug will prove ineffective or cause harm."[112] Accordingly, the Supreme Court framed the question by asking, "do the patent claims add enough to their statement of the correlations to allow the processes they describe to qualify as patent-eligible processes that apply natural laws?"[113]

Relying on an examination of each of the claimed limitations, the Supreme Court answered its own question by saying "no."[114] The decision first examined a step of "administering" the particular drug.[115] According to the Court, this step simply limited the use of the correlation to the relevant

---

the Supreme Court's apparent encouragement to lower courts to continue to develop new tests of patent eligibility reflect a continuing potential for confusion.").

106. Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289, 1293 (2012).

107. Indeed, the Supreme Court vacated *Ultramercial* and remanded it to the Federal Circuit in light of *Mayo*. WildTangent, Inc. v. Ultramercial, LLC, 132 S. Ct. 2431 (2012).

108. *Mayo*, 132 S. Ct. at 1295.

109. *Id.*

110. *Id.*

111. *Id.* at 1304.

112. *Id.* at 1296.

113. *Id.* at 1297.

114. *Id.*

115. *Id.*

audience: doctors.[116] Since limiting the use of an abstract idea to a particular technological environment cannot circumvent the prohibition against patenting abstract ideas, that step did not render the claims patentable.[117] Second, the Court examined two "wherein" limitations that noted the correlation between particular drug metabolite levels and a need to change the dosage.[118] The Court characterized these limitations as "simply tell[ing] a doctor about the relevant laws, at most adding a suggestion that he should take those laws into account when treating his patient."[119] The decision said nothing more on the topic, apparently indicating that these limitations clearly could not change an unpatentable concept into a patentable application. Third, the decision turned to the step of "determining" the level of the drug's metabolite in the body.[120] This step was well known in the prior art.[121] Since conventional or obvious pre-solution activity is not normally sufficient to transform an unpatentable law of nature into a patent-eligible application, the Court disregarded this step as well.[122] The Court concluded that none of the limitations, individually or in combination, were sufficient "to transform the nature of the claim."[123] In short, the Supreme Court determined that three types of limitations do not make an unpatentable idea patent eligible: (1) limiting an unpatentable concept to a particular audience, (2) telling someone about the concept, or (3) adding a conventional or obvious pre-solution activity.[124]

After determining that the claims did not add "enough" to the unpatentable idea at the heart of the invention,[125] the Supreme Court pursued three additional lines of analysis that ostensibly corroborated its conclusion. I previously criticized these dicta because they are analytically weak and

116. *Id.*
117. *Id.* (citing *In re* Bilski, 545 F.3d 943, 962 (Fed. Cir. 2008)).
118. The Court quoted the following limitations:

> [W]herein the level of 6-thioguanine less than about 230 pmol per $8x10^8$ red blood cells indicates a need to increase the amount of said drug subsequently administered to said subject and wherein the level of 6-thioguanine greater than about 400 pmol per $8x10^8$ red blood cells indicates a need to decrease the amount of said drug subsequently administered to said subject.

Id. at 1295 (citing U.S. Patent No. 6,355,623 col.20 ll.10–20 (filed Apr. 8, 1999)).
119. *Id.* at 1297.
120. *Id.* at 1297–98.
121. *Id.*
122. *Id.*
123. *Id.*
124. *Id.*
125. *Id.*

difficult to apply.[126] I provide an abbreviated version of my critique here with a few refinements. First, the decision used two earlier Supreme Court decisions as guideposts, *Parker v. Flook*[127] and *Diamond v. Diehr*,[128] and suggested that Prometheus' claims were closer to the ineligible claims in *Flook* than the eligible claims in *Diehr*.[129] These guideposts are problematic because—as many commentators have noted—the distinction between *Flook* and *Diehr* is unclear, and may be nonexistent.[130] Both cases appear to apply a new formula to an industrial process. Even the *Mayo* Court appeared to have trouble explaining just why the additional steps in *Diehr* rendered its claims patent eligible.[131]

Second, the Court said that simply appending general limitations to a concept is just like saying "apply it," and clearly is insufficient to render an unpatentable law of nature patent eligible.[132] This line of inquiry is also troublesome. The complaint about *general* limitations appears disingenuous. The claims in *Mayo* were quite specific; they identified particular levels of particular drug metabolites that would indicate when the dosing should change. Thus, this line of inquiry really appears to be an unhelpful "know it when you see it" kind of analysis.[133]

Third, the Court said that Prometheus' claims were too broad and impermissibly tied up the future use of a law of nature.[134] This issue is often framed as a question of preemption: Does the claim preempt all uses of the unpatentable concept (e.g., law of nature or abstract idea)?[135] But this preemption test can easily be manipulated. Almost any claim can be

126. Chao, *supra* note 15, at 429–32.

127. Parker v. Flook, 437 U.S. 584 (1978).

128. Diamond v. Diehr, 450 U.S. 175 (1981).

129. *Mayo*, 132 S. Ct. at 1298–1300.

130. *See, e.g.*, Mark A. Lemley, *Point of Novelty*, 105 NW. U. L. REV. 1253, 1278 (2011) (characterizing the claims in *Diehr* and *Flook* as "almost exactly parallel"); Kevin Emerson Collins, *Propertizing Thought*, 60 S.M.U. L. REV. 317, 349 (2007) ("*Flook* and *Diehr* are difficult to reconcile.").

131. *Mayo*, 132 S. Ct. at 1299 (only saying that "[t]hese other steps *apparently* added to the formula *something* that in terms of patent law's objectives had significance—they transformed the process into an inventive application of the formula." (emphasis added)).

132. *Id.* at 1300.

133. *See* CLS Bank Int'l v. Alice Corp., 685 F.3d 1341, 1348–52 (Prost, J., dissenting); *see also infra* text accompanying notes 148–49.

134. *Mayo*, 132 S. Ct. at 1301–02.

135. *See id.* at 1301; *see, e.g.*, Bilski v. Kappos, 130 S. Ct. 3218, 3231 (2010) ("Allowing petitioners to patent risk hedging would pre-empt use of this approach in all fields, and would effectively grant a monopoly over an abstract idea."); Diamond v. Diehr, 450 U.S. 175, 185 (1981) (stating that no one can patent "laws of nature, natural phenomena, and abstract ideas").

characterized as too broad if the concept is defined narrowly.[136] *Mayo* provides a good example of this problem. The Supreme Court defined the natural laws at issue as "the relationships between the concentration in the blood of certain thiopurine metabolites and the likelihood that the drug dosage will be ineffective or induce harmful side-effects."[137] However, the Court could have just as easily said that the natural law was the effect that thiopurine had on humans. Alternatively, the natural law could have been the understanding that a reduction in *any* drug dose leads to lower levels of the corresponding metabolite in the body. If the natural law were characterized in either of these fashions, the claims would have been drawn to one narrow application. In other words, the claims would not have preempted all applications of the natural law, suggesting that they were drawn to eligible subject matter.

Importantly, *Mayo* did not apply (or reject) the machine-or-transformation test, which had effectively emerged as the only test for determining patent eligibility after *Bilski*.[138] Instead, the Supreme Court merely said "we have neither said nor implied that the [machine-or-transformation] test trumps the 'law of nature' exclusion. That being so, the test fails here."[139] The Court then proceeded to apply its new approach. By assessing whether a claim's limitations added "enough" to the law of nature that Prometheus' inventors had discovered, *Mayo* outlined another line of inquiry to examine. But that was not all. The three corroborating justifications provided even more fodder for the lower courts to chew upon. Given all these varied and difficult ways to assess patent eligibility, it is not surprising that there continued to be disagreement within the Federal Circuit.

C. THE POST-*MAYO* SPLIT

CLS Bank International v. Alice Corp.[140] was the first Federal Circuit decision to address the eligibility of software patents after *Mayo*. The central idea underlying the patents in *CLS Bank* related to exchanging obligations

136. *See* Tun-Jen Chiang, *The Rules and Standards of Patentable Subject Matter*, 2010 WIS. L. REV. 1353, 1369–71 (2010) (explaining how claims can be viewed at different levels of abstraction).

137. *Mayo*, 132 S. Ct. at 1294.

138. *See* Lemley et al., *supra* note 44, at 1316 ("[T]he U.S. Patent and Trademark Office (PTO), patent litigants, and district courts have all continued to rely on the machine-or-transformation test in the wake of Bilski: no longer as the sole rule, but as a presumptive starting point that threatens to become effectively mandatory.").

139. *Mayo*, 132 S. Ct. at 1303 (internal citations omitted).

140. *CLS Bank I*, 685 F.3d 1341 (Fed. Cir. 2012), *reh'g en banc granted, opinion vacated*, 484 F. App'x 559 (Fed. Cir. 2012), *aff'd en banc*, 717 F.3d 1269 (Fed. Cir. 2013).

using a third party to eliminate risk.[141] Although the patents contained both method and system claims, they all used a computer and the court said that the "form of the claim" did not change the patent eligibility analysis.[142]

Although the majority opinion in *CLS Bank* discussed the recent Supreme Court decision in *Mayo*, the decision had little impact on the rule the majority announced.[143] Judge Linn, joined by Judge O'Malley, wrote: "[T]his court holds that when—after taking all of the claim recitations into consideration—it is not *manifestly evident* that a claim is directed to a patent ineligible abstract idea, that claim must not be deemed for that reason to be inadequate under § 101."[144] But the "*manifestly evident*" language is not from *Mayo* or even from *Bilski*. It comes from *Research Corp. Technologies v. Microsoft*[145] and *Ultramercial v. Hulu*,[146] two Federal Circuit decisions that were decided prior to *Mayo*. The majority in *CLS Bank* interpreted the "*manifestly evident*" rule in an even more patent friendly way, stating that a claim is only drawn to unpatentable subject matter if "the single most reasonable understanding is that a claim is directed to nothing more than a fundamental truth or disembodied concept . . . ."[147] Under this standard, it was easy for the Federal Circuit to find that Alice's patents covered patent-eligible subject matter.

CLS Bank does not follow *Mayo*'s approach, which focuses on whether certain claim limitations add "enough" to the unpatentable abstract concept to render it patent eligible.[148] To be fair, at the very end of the opinion, the majority paid lip service to *Mayo* by characterizing some claim limitations as being "integral" to the method, "playing a significant part in permitting the method to be performed."[149] Judge Prost's dissent challenged these statements and argued that the majority did "not explain whether [the

141. *Id.* at 1343.

142. *Id.* at 1353.

143. *See id.* at 1348, 1350–51.

144. *Id.* at 1352 (emphasis added).

145. *See* Research Corp. Techs., Inc. v. Microsoft Corp., 627 F.3d 859, 868 (Fed. Cir. 2010) (stating that a "disqualifying characteristic should exhibit itself so manifestly as to override the broad statutory categories of eligible subject matter").

146. *See* Ultramercial, LLC v. Hulu, LLC, 657 F.3d 1323, 1327 (Fed. Cir. 2011) (citing *Research Corp.*, 627 F.3d at 868), *vacated sub nom.* WildTangent, Inc. v. Ultramercial, LLC, 132 S. Ct. 2431 (2012).

147. *CLS Bank I*, 685 F.3d at 1352.

148. Judge Prost's dissent also noticed this problem and criticized the majority for failing to follow the Supreme Court's approach. *CLS Bank I*, 685 F.3d at 1357 (Prost, J., dissenting); *see also* Robert D. Swanson, Note, *Section 101 and Computer-Implemented Inventions*, 16 STAN. TECH. L. REV. 161, 166 (2012) (noting that a "manifestly abstract" test is inconsistent with both *Bilski* and *Mayo*).

149. *CLS Bank I*, 685 F.3d at 1355.

additional limitations] should be characterized as such, and what 'integral' means in the context of § 101 in the first place."[150]

Although the majority's statements were conclusory, the dissent's approach was hardly more illuminating. It provided a simplified description of the claims and found: "The claim in effect presents an abstract idea and then says 'apply it.' That is not enough."[151] Although this construct is clearly found in *Mayo*, it is unhelpful. Both the majority and dissent believe that they know a claim directed at an unpatentable abstract idea when they see it, but they clearly see particular claims differently.

Two weeks after *CLS Bank*, the Federal Circuit decided *Bancorp Services v. Sun Life*,[152] which only added to the confusion. In *Bancorp*, the patents covered both system and method claims for administering and tracking the values of life insurance policies in separate accounts.[153] Some of the method claims did not have to be implemented on a computer while all the remaining claims did. Despite these differences, the Federal Circuit treated all the claims "as equivalent for purposes of patent eligibility under § 101."[154]

This time the Federal Circuit applied a variation of the *Mayo* approach that was specifically tailored to software patents. After reviewing the § 101 jurisprudence, the court in *Bancorp Services* declared, "To salvage an otherwise patent-ineligible process, a computer must be integral to the claimed invention, facilitating the process in a way that a person making calculations or computations could not."[155]

Applying this test, the Federal Circuit identified an unpatentable abstract idea underlying the claims—"managing a stable value protected life insurance policy and then instructing the use of well-known calculations to help establish some of the inputs into the equation."[156] Even though many of the claims also required a computer, the court found that "[t]he computer required by . . . Bancorp's claims is employed only for its most basic function, the performance of repetitive calculations, and as such does not impose meaningful limits on the scope of those claims."[157] The Federal Circuit emphasized the limited role computers played in Bancorp's claim,

150. *Id.* at 1357 (emphasis omitted).

151. *Id.* at 1358 (citing Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289, 1294 (2012)).

152. Bancorp Servs. v. Sun Life Assurance Co., 687 F.3d 1266 (Fed. Cir. 2012).

153. *Id.* at 1270–72.

154. *Id.* at 1277.

155. *Id.* at 1278 (citing SiRF Tech., Inc. v. Int'l Trade Comm'n, 601 F.3d 1319, 1333 (Fed. Cir. 2010)).

156. *Id.* (citing Bilski v. Kappos, 130 S. Ct. 3218, 3231 (2010)) (brackets omitted).

157. *Id.*

finding that "[i]t is the management of the life insurance policy that is 'integral to each of Bancorp's claims at issue,' not the computer machinery that may be used to accomplish it"; the determination of the values in the claims was "a matter of mere mathematical computation."[158]

Notably, *Bancorp Services* did not explicitly reject, nor did it discuss, *CLS Bank*'s "manifestly evident" rule.[159] Rather, *Bancorp Services* distinguished the outcome in *CLS Bank* by saying that the computer limitations in *CLS Bank* played a "*significant part* in the performance of [that] invention or that the claims were limited to a *very specific application* . . . ."[160] Even though *Bancorp Services* attempted to reconcile its holding with *CLS Bank*, these cases took fundamentally different approaches to analyzing the patent eligibility of software patents.[161]

Under the *Bancorp Services* approach, a court dissects a claim to determine whether there is an unpatentable abstract idea at its core.[162] If there is, the court then determines whether any computer limitations are "integral" to the claimed invention.[163] In contrast, the *CLS Bank* approach looks at a claim as a whole and seeks to determine whether it is "manifestly evident that [the] claim is directed to a patent ineligible abstract idea."[164]

Given this disagreement, it is not surprising that the Federal Circuit decided to rehear *CLS Bank* en banc. The order granting the petition asked, "What test should the court adopt to determine whether a computer-implemented invention is a patent ineligible 'abstract idea'; and when, if ever, does the presence of a computer in a claim lend patent eligibility to an

158. *Id.* at 1279–80.

159. *See id.*; *CLS Bank I*, 685 F.3d 1341, 1352 (Fed. Cir. 2012), *reh'g en banc granted, opinion vacated*, 484 F. App'x 559 (Fed. Cir. 2012), *aff'd en banc*, 717 F.3d 1269 (Fed. Cir. 2013).

160. *Bancorp*, 687 F.3d at 1280 (citing *CLS Bank I*, 685 F.3d at 1355).

161. Dennis Crouch, *Ongoing Debate: Is Software Patentable?*, PATENTLY-O (July 27 2012, 3:53 PM), http://www.patentlyo.com/patent/2012/07/ongoing-debate-is-software-patentable.html ("Despite this attempted reconciliation, it is clear that the CLS majority has a different approach to subject matter eligibility questions [than *Bancorp*]."); *see also* Eric Guttag, Bancorp Services*: Further Fracturing of the Patent Eligibility Landscape for Business Methods and Systems*, IPWATCHDOG (July 27, 2012, 11:15 AM), http://www.ipwatchdog.com/2012/07/27/bancorp-services-further-fracturing-of-the-patent-eligibility-landscape-for-business-methods-and-systems/id=26881/ (characterizing *Bancorp* as "yet more evidence of the further fracturing of the patent-eligibility landscape").

162. *Bancorp*, 687 F.3d at 1276.

163. *Id.* at 1278.

164. *CLS Bank I*, 685 F.3d 1341, 1352 (Fed. Cir. 2012), *reh'g en banc granted, opinion vacated*, 484 F. App'x 559 (Fed. Cir. 2012), *aff'd en banc*, 717 F.3d 1269 (Fed. Cir. 2013) (emphasis added).

otherwise patent-ineligible idea?"[165] In the following Part, I argue that the point-of-novelty approach answers the question the Federal Circuit presented.

IV. A POINT-OF-NOVELTY RESOLUTION

Because *Mayo* only identified categories of claim limitations that failed to render an unpatentable concept patent eligible, many worried that *Mayo* might radically limit patent-eligible subject matter.[166] One commentator went so far as to say that the decision "creates a framework for patent eligibility in which almost any method claim can be invalidated."[167] My earlier essay, *Moderating* Mayo, offered a more restrained interpretation of the Supreme Court decision. I argued that *Mayo* implicitly adopted a point-of-novelty approach, and that this approach did not need to radically limit patent eligibility in the way many feared.[168]

The point of novelty is the claim limitation or limitations that correspond to the heart or gist of the invention.[169] Historically, patent law has refused to consider a patent's point of novelty in a wide-ranging number of doctrines.[170] Both Mark Lemley and I have separately criticized that jurisprudence. As Lemley said, "It makes little sense for a law focused on invention to pay no attention to what is inventive about the patentee's technology."[171] More

165. *CLS Bank II*, 484 F. App'x 559, 559 (Fed. Cir. 2012) (order granting hearing en banc). The order also asked whether the form of the claim matters (i.e., method, system, or storage medium claims). *Id.* at 559–60.

166. *See, e.g.*, Gene Quinn, *Killing Industry: The Supreme Court Blows* Mayo v. Prometheus, IPWATCHDOG (Mar. 20, 2012, 1:44 PM), http://www.ipwatchdog.com/2012/03/20/supreme-court-mayo-v-prometheus/id=22920/ ("The sky is falling! . . . Those in the biotech, medical diagnostics and pharmaceutical industries have just been taken out behind the woodshed and summarily executed . . . ."); *see also supra* note 19.

167. Robert R. Sachs, *Punishing Prometheus: The Supreme Court's Blunders in* Mayo v. Prometheus, PATENTLY-O (Mar. 26, 2002, 8:10 AM), http://www.patentlyo.com/patent/2012/03/punishing-prometheus-the-supreme-courts-blunders-in-mayo-v-prometheus.html.

168. Chao, *supra* note 15, at 425.

169. *See* Aro Mfg. Co. v. Convertible Top Replacement Co. 365 U.S. 336, 344–45 (1961) ("[T]here is no legally recognizable or protected 'essential' element, 'gist' or 'heart' of the invention in a combination patent.").

170. Bernard Chao, *Breaking* Aro's *Commandment: Recognizing that Inventions Have Heart*, 20 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 1183, 1192 (explaining that the point of novelty has been rejected in assessing direct infringement, anticipation, obviousness, the written description requirement, and repair and reconstruction).

171. Lemley, *Point of Novelty*, *supra* note 130, at 1274–75; *see also* Kevin Emerson Collins, *Getting into the "Spirit" of Innovative Things: Looking to Complementary and Substitute Properties to Shape Patent Protection for Improvement*, 26 BERKELEY TECH. L.J. 1217, 1237 (2011) (arguing that the failure to consider the point of novelty is "highly problematic in the context of patent protection for improvements.").

specifically, I argued that the point of novelty "should play an important role" in subject matter patent eligibility determinations.[172] The *Mayo* Court appears to implicitly embrace this view. *Mayo*'s approach assessed whether a patent's point of novelty was an unpatentable concept (i.e., law of nature, natural phenomenon, or abstract idea) and then determined whether any claim limitations somehow transformed that concept into a patent-eligible application.[173]

To be clear, *Mayo*'s approach is very different from the point-of-novelty approach the Supreme Court applied years ago in *Parker v. Flook*.[174] The claims in *Flook* involved a new formula for calculating an alarm limit for a catalytic chemical conversion of hydrocarbons.[175] The claimed process contained three steps: "an initial step which merely measures the present value of the process variable (e.g., the temperature); an intermediate step which uses an algorithm to calculate an updated alarm-limit value; and a final step in which the actual alarm limit is adjusted to the updated value."[176] The Supreme Court found that the invention was not patent-eligible "because once [the] algorithm is assumed to be within the prior art, the application, considered as a whole, contains no patentable invention."[177] Thus, *Flook* suggested that a claim's point of novelty could not be based on an unpatentable concept.[178]

Diehr appears to have later rejected the *Flook* point-of-novelty analysis,[179] and *Mayo* did not revive it. Although *Mayo* also focused on a patent's point of novelty, the decision did not assume that the point of novelty was in the prior art and require that other limitations be "new." Rather, *Mayo* focused on the natural law that corresponded to the claimed point of novelty and asked if the other limitations added enough to that concept to render the

172. Chao, *supra* note 170, at 1220. *But see* Lemley, *supra* note 130, at 1278–79 (expressing concern about relying on the point of novelty when making subject matter eligibility determinations).

173. Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289, 1297–98 (2012).

174. *Compare id.*, *with* Parker v. Flook, 437 U.S. 584 (1978).

175. *Flook*, 437 U.S. at 586.

176. *Id.* at 585 (footnotes omitted).

177. *Id.* at 594.

178. *See id.*

179. *Compare id.*, *with* Diamond v. Diehr, 450 U.S. 175, 188 (1981) ("In determining the eligibility of [a] claimed process for patent protection under § 101, [the] claims must be considered as a whole. It is inappropriate to dissect the claims into old and new elements and then to ignore the presence of the old elements in the analysis."), *and* Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289 (2012).

claim patent eligible.[180] That inquiry went far beyond whether the other limitations simply added something new and non-obvious.[181] Thus, there is no reason to believe that an application based on a newly discovered formula or natural law is categorically unpatentable under *Mayo*.

Unfortunately, the Supreme Court only explained when certain limitations failed to add enough to an unpatentable concept. *Mayo* did not explain what kind of limitation could be added to an unpatentable concept to render it patent-eligible. My earlier proposal attempted to fill in that gap and offer a new point-of-novelty framework (i.e., one different from *Flook*). Relying on concepts already found in existing § 101 jurisprudence, I offered a two-part test for determining when patents cover subject matter that should be patent eligible.[182] Courts should first examine the limitation that embodies the point of novelty to determine whether it describes an unpatentable concept (i.e., a law of nature, a natural phenomenon, or an abstract idea).[183] If it does, the court should then determine whether the other limitations bring the concept into the realm of patentable subject matter.[184] This occurs when the other limitations are both concrete and strongly connected to the point of novelty.[185]

The point of novelty of many software patents is often an abstract idea. It could be a mathematical formula or a new way of doing business. Understanding that ideas themselves cannot be claimed, patent attorneys typically draft software patent claims to include a concrete physical device like a computer or the Internet.[186] They hope that by adding these limitations, an otherwise unpatentable abstract idea may be rendered patentable. Under the point-of-novelty approach, this tactic would only work for certain kinds of patents.

---

180. *Mayo*, 132 S. Ct. at 1297–98.

181. *See supra* Section III.B.

182. Chao, *supra* note 15, at 426, 436.

183. *Id.* at 436.

184. *Id.*

185. *Id.*

186. *See* Robert A. King, *Developing A Successful Intellectual Property Program*, *in* DEVELOPING A PATENT STRATEGY 117, 2011 WL 1120279 (Aspatore 2011) ("The machine or transformation test represents a 'safe harbor' for claim drafting. Many patent practitioners draft claims to meet this test, and, in many cases, starting with the minimum amount of machine-related references in the claims."); John R. Allison & Starling D. Hunter, *On the Feasibility of Improving Patent Quality One Technology at a Time: The Case of Business Method*s, 21 BERKELEY TECH. L.J. 729, 736 (2006) ("[A]ttorneys had little difficulty drafting patent applications on software as though they claimed machines and devices of a more traditional physical nature."); David R. Heckadon, *Six Months After* Bilski*: Practical Claim Drafting Tips for Software and Business Method Patents*, GORDON & REES LLP (Nov. 2010), http://www.gordonrees.com/publications/viewPublication.cfm?contentID=1705.

The key is to test the strength of the nexus between the point of novelty and any additional concrete claim limitations. There may be a number of ways of characterizing such a test. I suggest that when the unpatentable concept and the additional limitations are *bound together* and *necessary* to achieve the goal of the claimed invention, the nexus is sufficiently strong to be patent eligible. One way to assess whether the concept is sufficiently bound together with the other limitations is to determine if the concept stands by itself. For example, consider Bilski's idea of hedging risk in a particular industry.[187] Computer limitations could be added to the concept, but the concept would make sense standing by itself and therefore remain ineligible for patenting. Moreover, just because a computer or other physical device may be useful or practically necessary does not mean that the claim should be patent eligible. This means that a patent's goal cannot be characterized as merely applying a concept in a manner that is more efficient, faster, or more cost-effective by simply using a computer to conduct the process.

Although this Article is the first to articulate the "bound together" standard, the idea of testing the connection between the unpatentable concept and other limitations is already scattered throughout existing subject matter patent eligibility jurisprudence. The machine prong of the machine-or-transformation test examines how strong the nexus is between the unpatentable concept and other more concrete claim limitations (i.e., machines). More recently, *Bancorp*'s requirement that a computer must be "integral to the claimed invention" tests the same connection.[188] The Supreme Court has repeatedly stated that adding insignificant post-solution limitations does not make an abstract idea patentable.[189] But this is simply another way of saying that the nexus between the unpatentable concept and its other limitations is not sufficiently strong.

If these threads existed by themselves, the law would not be so fractured. It would coalesce around a point-of-novelty approach. Unfortunately, the courts have also included many other unrelated factors that confuse subject-matter eligibility determinations. Cases uniformly suggest that claims that sweep too broadly are less likely to be patent eligible.[190] Consequently, they

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

188. *See* Bancorp Servs. v. Sun Life Assurance Co., 687 F.3d 1266, 1278 (Fed. Cir. 2012) (citing SiRF Tech., Inc. v. Int'l Trade Comm'n, 601 F.3d 1319, 1333 (Fed. Cir. 2010)).

189. Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289, 1301 (2012); *Bilski*, 130 S. Ct. at 3231; Diamond v. Diehr, 450 U.S. 175, 191–92 (1981); Parker v. Flook, 437 U.S. 584, 590 (1978).

190. *Mayo*, 132 S. Ct. at 1301; *Bilski*, 130 S. Ct. at 3231 ("Allowing petitioners to patent risk hedging would pre-empt use of this approach in all fields, and would effectively grant a monopoly over an abstract idea."); Gottschalk v. Benson, 409 U.S. 63, 68 (1972) (stating that

ask if a limitation has placed a meaningful limit on claim scope[191] or preempted the entire idea.[192] The Supreme Court has also pointed to *Diehr* and *Flook* as guideposts, suggesting that a claim is patent eligible if it is closer to *Diehr* and is not patent eligible if it is closer to *Flook*.[193] As discussed earlier, these lines of inquiry suffer from both analytical and practical problems.[194] But just as importantly, these inquiries obscure what should be the proper inquiry—assessing the strength of the connection between the unpatentable concept and the other claim limitations. Under the point-of-novelty approach illustrated in this Article, these other lines of inquiry would not be used.[195]

The following examples illustrate how the point-of-novelty approach works. Consider the patent at issue in *Diehr*, which claimed a novel algorithm for curing rubber products.[196] Both the formula and the physical components (the rubber molding press) were necessary to accomplish the invention's goal of making precision molded rubber products.[197] Without the physical device, the formula could not achieve the goal of the invention. Moreover, it makes no sense to discuss the formula apart from the physical devices used to implement it. Thus, the connection between the idea and the device is sufficiently strong such that the subject matter should be patent eligible. The same analysis would also suggest that *Flook* was wrongly decided.[198] The algorithm for updating the alarm limit of a catalytic chemical conversion of hydrocarbons was clearly bound together with industrial equipment and both limitations are necessary to perform the conversion.[199] Accordingly, under

the claims before it were "so abstract and sweeping as to cover both known and unknown uses of the [mathematical formula]").

191. Dealertrack Inc. v. Huber, 674 F.3d 1315, 1333 (Fed. Cir. 2012); *CLS Bank I*, 685 F.3d 1341, 1351 (Fed. Cir. 2012), *reh'g en banc granted, opinion vacated*, 484 F. App'x 559 (Fed. Cir. 2012), *aff'd en banc*, 717 F.3d 1269 (Fed. Cir. 2013).

192. *See, e.g.*, *Bilski*, 130 S. Ct. at 3258 (Breyer, J., concurring); *Bancorp*, 687 F.3d at 1280–81; *Dealertrack*, 674 F.3d at 1331.

193. *See supra* text accompanying notes 126–30.

194. *See supra* text accompanying notes 126–36.

195. Chao, *supra* note 15, at 440.

196. Diamond v. Diehr, 450 U.S. 175, 177 (1981).

197. *See id.* at 189.

198. *See supra* text accompanying note 130 (discussing how the challenged patents in *Flook* and *Diehr* appear to be similar).

199. Perhaps the problem was that claim 1 did not explicitly connect the formula to physical device. But because the claim recited "the catalytic chemical conversion of hydrocarbons," Parker v. Flook, 437 U.S. 584, 596 (1978), the Court could have easily interpreted the claim to require such limitation or noted that adding such a limitation would render Flook's claims patent eligible.

the point-of-novelty approach, *Flook*'s invention would have been patent eligible.

However, in many other cases, the ideas underlying the software patents do not have strong connections to the devices. Consider the patents in *CLS Bank*, where the claims related to a trading system platform for exchanging obligations.[200] The computer limitations were certainly very useful but were not fundamentally necessary for exchanging obligations.[201] Moreover, the idea of exchanging obligations makes perfect sense standing alone; the nexus between the physical components and the underlying concept is weak. In *CLS Bank*, attaching the concept to a machine should not make the concept patentable.

The same analysis would apply to the patents in *Ultramercial*, *Cybersource*, and *Dealertrack*. The ideas underlying all these patents are not sufficiently bound together with their computer limitations to render them patent eligible. Ultramercial's idea of receiving free copyrighted content in exchange for viewing an advertisement does not clearly need the Internet.[202] The idea makes sense without any physical devices. Cybersource's fraud detection patent examined Internet address information and compared it with other transactions utilizing the same credit card.[203] Some type of computer may be necessary to obtain the Internet address information, but the idea of comparing addresses did not need to be performed on a computer, much less any physical device. Thus, this idea is not sufficiently connected to any concrete limitations. Finally, Dealertrack's patents merely automated a method of processing car loans.[204] As a practical matter, computers were undoubtedly necessary to make the system operate efficiently. But that does not satisfy the revised point-of-novelty test. Since the underlying system could operate without computers, albeit inefficiently, the patents do not cover patent-eligible subject matter. Accordingly, none of these patents would survive the proposed point-of-novelty approach described here.[205]

200. *CLS Bank I*, 685 F.3d 1341, 1343 (Fed. Cir. 2012), *reh'g en banc granted, opinion vacated*, 484 F. App'x 559 (Fed. Cir. 2012), *aff'd en banc*, 717 F.3d 1269 (Fed. Cir. 2013).

201. *See id.* at 1358 n.1 (Prost, J., dissenting) (using a table to demonstrate how the steps of one of the claims at issue correspond to ordinary activities).

202. Ultramercial, LLC v. Hulu, LLC, 657 F.3d 1323, 1324 (Fed. Cir. 2011).

203. CyberSource Corp. v. Retail Decisions, Inc., 654 F.3d 1366, 1367–68 (Fed. Cir. 2011).

204. Dealertrack Inc. v. Huber, 674 F.3d 1315, 1317–18 (Fed. Cir. 2012).

205. I do not analyze the *Bancorp Services* patents because, for the most part, the Federal Circuit took the approach I advocate here. *See supra* text accompanying notes 154–57.

As stated earlier, the digital image halftoning patents in *Research Corporation Technologies v. Microsoft* pose a more difficult case.[206] The claims of Research Corporation Technologies ("RCT") revolved around a mathematical algorithm for determining how to display images for different devices like printers and displays. By itself the algorithm was clearly unpatentable.[207] But all the claims added some limitations to this algorithm. At its barest, one method claim merely applied the algorithm to produce "dot profiles."[208] Other claims also required a computer.[209] Still other claims added physical components like "high contrast film," "a film printer," and "printer and display devices."[210]

Under the point-of-novelty approach, the test is whether these limitations are both concrete and strongly connected to the mathematical algorithm that lies at the patents' point of novelty. The last set of claims presents the easiest case. Clearly, "high contrast film," "a film printer," and "printer and display devices" are concrete. Moreover, the nexus between these devices and the mathematical algorithm is strong. Using an algorithm to calculate the proper way to display particular dots is tightly linked with the display devices themselves. Making those calculations without some form of display device achieves nothing. Therefore, under the point-of-novelty approach, these claims would cover patent-eligible subject matter.

The more difficult question arises when these physical components are not recited by a given claim. This leaves only the dot profiles and the computer components to consider. Even though the computer limitations are concrete, they do not have a strong nexus with the underlying mathematical algorithm. There is no special connection between a computer and the algorithm for calculating how to display images. Presumably, the mathematical algorithm can be calculated without the use of a computer. A computer merely makes the calculations faster. So the computer limitations do not "add enough" to make the algorithm patent eligible.[211]

---

206. Research Corp. Techs. Inc., v. Microsoft Corp., 627 F.3d 859 (Fed. Cir. 2010); *see also supra* text accompanying notes 64–66.

207. *See* Diamond v. Diehr, 450 U.S. 175, 187 (1981); *see also* Gottschalk v. Benson, 409 U.S. 63, 67 (1972) (holding that a mathematical expression is simply a "scientific truth" and unpatentable (quoting Mackay Radio & Tel. Co. v. Radio Corp., 306 U.S. 86, 94 (1939))). But the claims were not drawn to the algorithm standing alone. Consequently, the Federal Circuit stated that the claims were not "manifestly abstract" and found that the claims covered patent eligible subject matter. *Research Corp.*, 627 F.3d at 869.

208. *See Research Corp.*, 627 F.3d at 865 (reciting claim 1 of the '310 patent).

209. *See, e.g.*, *id.* at 865–66 (reciting claims 1, 4, and 57 of the '772 patent).

210. *Id.* at 869.

211. *See* Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289, 1297 (2012) (establishing the "add enough" requirement).

But what about the dot profiles? Perhaps these limitations can render the algorithm patent eligible. However, the claimed dot profiles are not physical dots. The limitations actually refer to data that represents how dots are to be printed or displayed. This appears to be a close call. But allowing data that represents concrete objects to satisfy the point of novelty test reopens the door for business method patents.[212] Thus, the "dot profile" limitations should not be considered sufficiently concrete to transform the unpatentable algorithm into patent-eligible subject matter.

This result may appear inconsistent with the goals of providing patent protection to industrial software. But the point is not to give all claims directed at an industrial software protection. The claims still must be drafted to include concrete limitations (e.g., printers or displays). That is why some of RCT's claims survived the point-of-novelty approach and others did not. This has the benefit of limiting claim scope—an outcome that both courts and commentators have sought to make an explicit requirement of § 101 analysis.[213] Although the point-of-novelty approach does not have such an explicit requirement, the RCT example shows how the approach has a claim narrowing effect.

Some critics may argue that the point-of-novelty test proposed here will not make it easier to assess patentable subject matter eligibility. There will, of course, be some close cases, and the "bound together" and "necessary" language is not a magic bullet that will provide sudden clarity. When applying this test, the courts need to appreciate the basis for it. First, the test was designed with the understanding that the Supreme Court wants to do away with *most* business method software patents. Second, the test uses and expands upon the point-of-novelty approach found in *Mayo*. With this understanding, the Federal Circuit can achieve greater clarity as it works through several examples.

In sum, the point-of-novelty approach respects *Bilski*'s hostility to business method patents while operating within the analytical framework required by *Mayo*. The result will be that most business method software patents will be declared ineligible because the connection between the concepts underlying those patents and the computers that they use is not strong. However, most industrial software patents will remain patent-eligible because the concepts underlying these inventions are bound together with specific physical devices.

---

212. *See, e.g.*, Dealertrack, Inc. v. Huber 674 F.3d 1315, 1319–20 (Fed. Cir. 2012) (where the claims contained limitations directed toward credit card applications).

213. *See* notes 133–34 and accompanying text; Lemley et al., *supra* note 44, at 1317 (proposing that the test for patent eligibility under § 101 be based solely on overclaiming).

In February 2013, after this Article was accepted for publication, the Federal Circuit heard oral arguments in the en banc hearing of *CLS Bank*.[214] Interestingly, the U.S. Patent and Trademark Office ("USPTO") took a similar position to the one advocated here. In response to questions from Judge Moore, Deputy Solicitor Nathan Kelly argued that a claim does not become patent-eligible simply because it contains computer limitations.[215] He argued that one must "look deeper into the claim to see if the system and steps are inseparable."[216] The term "inseparable" is another way of assessing the strength of the connection between the point of novelty and any additional concrete claim limitations. Although the USPTO's brief did not use the "inseparable" language, it did characterize the same concept in yet another way. It said that a claim must incorporate "meaningful limitations" (i.e., a limitation that is not a "mere field-of-use limitation, a tangential reference to technology, insignificant, extra-solution activity, an ancillary data-gathering step, or the like.").[217] The patent office's brief also suggested six factors for making this determination.[218] Although some of these factors merely parrot back language from precedent, together they act much like the proposed "bound together and necessary" test. Dennis Crouch has

214. *See* Oral Argument, CLS Bank Int'l v. Alice Corp., No. 2011-1301 (Fed. Cir. Feb. 8, 2013) (en banc), *available at* http://oralarguments.cafc.uscourts.gov/default.aspx?fl=2011-1301_282013.mp3 (for an audio recording of the argument).

215. *See id.* at 28:00–28:41.

216. *See id.* at 28:41.

217. Brief for the United States as Amicus Curiae on Rehearing En Banc in Support of Neither Party at 7, *CLS Bank III*, 717 F.3d 1269 (Fed. Cir. 2013) (en banc).

218. The factors are:

> [1] whether the computer is recited in a manner that is only nominally or tangentially related to the performance of the invention . . . ; [2] whether the computer is generically recited in a manner that would encompass any machine capable of performing the claimed steps, or whether specific, unconventional computer equipment, tools, or processing capabilities are required; [3] whether the invention involves an improvement in the ability of the computer to function as a computer, or whether the invention relates principally to an unrelated, non-technological field . . . ; [4] whether the claim recites a computerized device that manipulates particular data in particular, specific, and useful ways . . . or whether the computer is recited solely for its generic functions of automating tasks or communicating over a distance; [5] whether . . . the abstract idea is bound up in an invention that effects a transformation of matter, or whether . . . the abstract idea is merely described in a particular environment; and [6] whether the computer-related elements of the claim represent conventional steps, described at a high level of generality, that would have to be employed by any person who wished to apply the abstract idea.

Id. at 13–14 (citations omitted).

characterized this as an "odd ball" approach.[219] But, as this Article has argued, this approach sensibly brings together the Supreme Court precedent.

V. POSTSCRIPT: A DEEPENING SCHISM

The Federal Circuit issued its en banc decision in *CLS Bank* on May 10, 2013.[220] As discussed earlier, the patents concerned exchanging obligations using a third party to eliminate risk.[221] The patents contained three types of claims directed (1) towards methods of exchanging obligations between parties, (2) data processing systems, and (3) computer readable media containing programs for exchanging obligations. The parties stipulated that all the claims required a computer.[222] Although the two-paragraph per curiam opinion affirmed the district court's decision finding all the claims patent-ineligible, the judges were badly divided and the decision failed to give the guidance that so many followers of the court sought.[223] The court was split five to five on the eligibility of the system claims. While seven judges did find that the method and computer-readable claims at issue were not patent-eligible, those judges could not agree on the rationale for this conclusion. In fact, no approach was endorsed by a majority of the judges.

The decision contained seven separate opinions reflecting roughly three analytical approaches. These approaches roughly correspond to opinions written by Judge Lourie, Chief Judge Rader, and Judge Newman. For convenience, this Article breaks down the decision into the "strong view" and the "weak view" of § 101's patent-eligibility requirement. The strong view corresponds to Judge Lourie's opinion (joined by Judges Dyk, Prost, Reyna, and Wallach). The weak view corresponds to Chief Judge Rader's first opinion (joined by Judge Moore and joined in part by Judges Linn and O'Malley), Judge Moore's opinion (joined by Chief Judge Rader, Judge Linn, and Judge O'Malley), Judges Linn and O'Malley's opinion, and Chief Judge Rader's second opinion. Although Judge Newman's opinion represents a third distinct analytical approach, the end result would be similar to the weak view. Accordingly, a brief description of her view is included in Section V.B, *infra*, describing the weak view.

---

219. Dennis Crouch, CLS Bank v. Alice Corp*: Oral Arguments Lead to More Questions*, PATENTLY-O (Feb. 9, 2013), http://www.patentlyo.com/patent/2013/02/cls-bank-v-alice-corp-oral-arguments-lead-to-more-questions.html.

220. *CLS Bank III*, 717 F.3d 1269 (Fed. Cir. 2013).

221. *See supra* text accompanying notes 139–41.

222. *CLS Bank III*, 717 F.3d at 1275 (Lourie, J., concurring).

223. In fact, Chief Judge Rader went so far as to say that "nothing" in the en banc opinion "beyond our judgment has the weight of precedent." *Id.* at 1293 n.1 (Rader, C.J., concurring in part and dissenting in part).

A. THE STRONG VIEW

Judge Lourie's concurring opinion argued that none of the claims satisfied § 101 and contains the strongest view of § 101's subject matter patent-eligibility requirement. The opinion was joined by Judges Dyk, Prost, Reyna, and Wallach and thus represents the view of five Federal Circuit judges. Before describing its approach, the opinion took note of some "common themes" found in the Supreme Court's decisions regarding § 101.[224] "First and foremost is an abiding concern that patents should not be allowed to preempt the fundamental tools of discovery."[225] In other words, "claims should not be coextensive with a natural law, natural phenomenon or abstract idea."[226] Second, Judge Lourie characterized the cases as cautioning "against overly formalistic approaches to subject-matter eligibility that invite manipulation by patent applicants."[227] Third, Judge Lourie also said that the cases "urge a flexible, claim-by-claim approach to subject-matter eligibility that avoids rigid line drawing."[228]

Relying on these principles, Judge Lourie then outlined an approach to determine whether a computer-implemented claim is patent-eligible. Assuming that the claim falls within one of the four statutory categories set out in § 101 (i.e., process, machine, manufacture, or composition of matter), the first question to ask is whether the claim poses any risk of preempting an abstract idea.[229] If it does, a court must then identify the "fundamental concept" that is wrapped up in the claim.[230] The analysis then proceeds to preemption analysis and looks to whether the claim covers the entire abstract idea itself.[231]

The opinion's preemption analysis relies on examining the "inventive concept." According to Judge Lourie, an inventive concept must be added to the underlying unpatentable fundamental concept to render the claim patent-eligible.[232] In contrast to a fundamental concept, an inventive concept must be "a product of human ingenuity."[233] Moreover, the inventive concept should not be confused with the novelty or obviousness requirements of

224. *CLS Bank III*, 717 F.3d at 1280–82 (Lourie, J., concurring).
225. *Id.* at 1280.
226. *Id.* at 1281.
227. *Id.*
228. *Id.*
229. *Id.*
230. *Id.*
231. *Id.* at 1282.
232. *Id.* at 1282–84.
233. *Id.* at 1283 (quoting Diamond v. Chakrabarty, 447 U.S. 303, 309 (1980)).

§§ 102 and 103. Interpreting *Parker v. Flook*,[234] Judge Lourie suggests that *Flook* only required that the claim contain an "inventive concept" to be patent-eligible.

For the most part, Judge Lourie defines an inventive concept in the negative. It "must represent more than a trivial appendix to the underlying abstract idea."[235] Thus, limitations that reflect the inventive concept are not "merely tangential, routine, well-understood, or conventional, or in practice fail to narrow the claim relative to the fundamental principle."[236] "Bare field-of-use limitations" do not qualify either.[237] Having said what the inventive concept *is not*, the discussion of how to approach determining patent-eligibility suddenly ends. Instead of explaining the positive characteristics of the inventive concept, the opinion simply acknowledges that it is not offering an "easy bright-line test," but rather one that depends on a "balance of factors."[238]

Judge Lourie's opinion then proceeds to use this approach to analyze the claims. Starting with the method claims, the opinion identifies the underlying abstract idea—"reducing settlement risk by effecting trades through a third-party intermediary . . . empowered to verify that both parties can fulfill their obligations before allowing the exchange—*i.e.*, a form of escrow."[239] Next,

234. Parker v. Flook, 427 U.S. 584 (1978).

235. *CLS Bank III*, 717 F.3d at 1283 (Lourie, J., concurring).

236. *Id.*

237. *Id.* at 1283–84.

238. *Id.*

239. *Id.* at 1286. Judge Lourie's opinion analyzes claim 33 of the '479 patent as a representative method claim. It recites:

> A method of exchanging obligations as between parties, each party holding a credit record and a debit record with an exchange institution, the credit records and debit records for exchange of predetermined obligations, the method comprising the steps of: (a) creating a shadow credit record and a shadow debit record for each stakeholder party to be held independently by a supervisory institution from the exchange institutions; (b) obtaining from each exchange institution a start-of-day balance for each shadow credit record and shadow debit record; (c) for every transaction resulting in an exchange obligation, the supervisory institution adjusting each respective party's shadow credit record or shadow debit record, allowing only these transactions that do not result in the value of the shadow debit record being less than the value of the shadow credit record at any time, each said adjustment taking place in chronological order; and (d) at the end-of-day, the supervisory institution instructing ones of the exchange institutions to exchange credits or debits to the credit record and debit record of the respective parties in accordance with the adjustments of the said permitted transactions, the

the additional limitations—"creating shadow records, using a computer to adjust and maintain those shadow records, and reconciling shadow records and corresponding exchange institution accounts through end-of-day transactions"—were examined.[240] Judge Lourie concludes that they do not add "anything of substance to the claim."[241]

This conclusion is built upon three observations. First, Judge Lourie's opinion says that the claim lacks any express language to define the computer's participation.[242] The computer simply acts as a calculator performing mental steps faster than a human could. Under the strong view, that is not sufficient to show the necessary inventive concept.[243] Second, the opinion views the term "shadow record" as "extravagant language" that merely recites "a basic function required of any financial intermediary in an escrow arrangement."[244] Finally, Judge Lourie characterizes the step of providing end-of-day instructions to reconcile the parties' accounts as a "trivial limitation."[245] Consequently, Judge Lourie concluded that the method claim was not drawn to patent-eligible subject matter.

After having found the method claims ineligible, Judge Lourie's opinion was quickly able to dispense with the computer-readable medium and system claims. The opinion characterizes those ostensibly concrete computer limitations as claim drafting tactics that add nothing of substance. Specifically, the opinion says that claim 39 of the '375 patent, the representative computer-readable medium claim, is not "truly drawn to a specific computer readable medium, rather than to the underlying method."[246] Moreover, the system claims merely "recite a handful of computer components in generic, functional terms that would encompass any device capable of performing the same ubiquitous calculation, storage, and connectivity functions required by the method claims."[247] Accordingly, the opinion concluded that all the asserted claims were invalid under § 101 for failure to recite patent-eligible subject matter.[248]

> credits and debits being irrevocable, time invariant obligations placed on the exchange institutions.

Id. at 1285.

240. *Id.* at 1286.

241. *Id.*

242. *Id.*

243. *Id.*

244. *Id.* at 1287

245. *Id.*

246. *Id.* at 1288 (quoting Cybersource Corp. v. Retail Decisions Inc., 654 F.3d 1366, 1374–75 (Fed. Cir. 2011)).

247. *Id.* at 1290.

248. *Id.* at 1292.

Although I agree with Judge Lourie's ultimate conclusion, I have two critiques of his approach to § 101. First, Judge Lourie overstates the significance of the "inventive concept" in Supreme Court precedent. Judge Rader's first opinion picks up on this flaw and says that Judge Lourie "imbues" the phrase with "a life that is neither consistent with the Patent Act's description of Section 101 nor with the totality of Supreme Court precedent . . . ."[249] I agree. As Chief Judge Rader points out, the *Flook* decision that Judge Lourie relies upon only mentions the "inventive concept" once.[250] The *Mayo* decision does not support Judge Lourie's reliance on the inventive concept either.[251] Although *Flook* and *Mayo* use the phrase to suggest that a patent must claim something more than a natural law, the decisions do not gauge patent-eligibility by using the inventive concept.[252] Judge Lourie just picks out that phrase to describe when additional limitations have added "enough" to render an otherwise unpatentable concept patent-eligible. Thus, his mistake is merely one of nomenclature, not substance. Nonetheless, relying on the term "inventive concept" provides room for critics to argue that Judge Lourie's opinion incorrectly interprets the controlling precedent.

The second difficulty with Judge Lourie's approach is more problematic. The opinion relies on already established principles to say what *is not* an inventive concept. Thus, even if this approach were eventually adopted, it would not provide any practical insights for determining when limitations actually contain an inventive concept. In contrast, the point-of-novelty approach proposed in this Article explains when a claim *is* patent-eligible.

---

249. *Id.* at 1303 n.5 (Rader, C.J., concurring in part and dissenting in part).

250. *Id.* Moreover, Judge Lourie's opinion misinterprets *Flook*. *Flook* suggested that a fundamental concept be treated as if were found in the prior art, and the novelty and non-obviousness requirement had to be satisfied by the other claim limitations. Parker v. Flook 437 U.S. 584, 594 (1978). This interpretation of *Flook* is shared by many commentators, including this Author. *See supra* notes 173–80 and accompanying text; Lemley et al., *supra* note 44, at 1335–36.

251. *See CLS Bank III*, 717 F.3d at 1282 (Lourie, J., concurring) (citing *Mayo* and *Flook* to support the notion of the "inventive concept").

252. Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289, 1294 (2012). In *Mayo*, the Court states:

> [Prior decisions] insist that a process that focuses upon the use of a natural law also contain other elements or a combination of elements, sometimes referred to as an "inventive concept," sufficient to ensure that the patent in practice amounts to significantly more than a patent upon the natural law itself.

Id. However, commentators consistently use this phrase. *See, e.g.*, Kevin Emerson Collins, *Prometheus Laboratories, Mental Steps, And Printed Matter*, 50 HOUS. L. REV. 391, 402 (2012) (discussing the use of the phrase, "inventive concept").

First, the "bound together" standard tests the nexus between the additional limitations and the underlying unpatentable concept to determine when limitations add "enough." Moreover, by prominently relying on *Bilski*, the lower courts can look to business method software patents and industrial software patents as new guideposts for determining patent-eligibility questions. To be fair, the point-of-novelty approach described in this Article is not a bright line test either. But it should provide more clarity than the approach proposed by Judge Lourie and his allies.

B. THE WEAK VIEW

Chief Judge Rader authored two separate opinions. His first opinion, concurring in part and dissenting in part, was joined by Judges Linn (except part VI), Moore, and O'Malley (except part VI), and reflects a comparatively weak view of § 101's subject matter patent-eligibility requirement. The first five parts of the opinion outline a specific approach for determining patent-eligibility and argue that the system claims were patent-eligible under § 101. Part VI distinguishes the method and computer-readable claims from the system claims, and argues that the former claims are not patent-eligible. Judges Linn and O'Malley disagreed with this conclusion and did not join part VI. Instead they wrote their own opinion explaining why the court should have found that the method and computer-readable claims were also patent-eligible.[253] Since the distinction between Chief Judge Rader and Judge Linn's opinions revolves around different claim interpretations, and not how to approach subject matter patent-eligibility, Judge Linn's view does not represent a distinct approach. Therefore, the weak view discussed here still represents the view of four Federal Circuit judges.[254]

Chief Judge Rader's opinion begins in earnest in Part II and emphasizes the breadth of subject matter that is patent-eligible under § 101.[255] After providing a lengthy analysis of the legislative history of § 101, the part concludes by saying: "In sum, any analysis of subject matter eligibility for patenting must begin by acknowledging that any new and useful process,

253. Thus, of those judges who advocated for a weak interpretation of § 101, only Judges Rader and Moore still believed that the method and computer-readable claims were not patent eligible.

254. Judge Linn has recently taken senior status. *2012*, UNITED STATES COURT OF APPEALS FOR THE FEDERAL CIRCUIT, http://www.cafc.uscourts.gov/2012. Therefore, he will no longer participate in any future en banc proceedings.

255. *See, e.g.*, *CLS Bank III*, 717 F.3d at 1294 (Rader, C.J., concurring in part and dissenting in part) ("Underscoring its breadth, Section 101 both uses expansive categories and modifies them with the word 'any.' ").

machine, composition of matter, or manufacture, or an improvement thereof, is eligible for patent protection."[256]

The opinion then goes on to discuss the judicial exceptions to patent-eligibility under § 101. Much like the proponents of the stricter view, Chief Judge Rader notes that a claim cannot merely cover an "abstract idea, law of nature, or natural phenomenon." Of course Chief Judge Rader's opinion strikes an entirely different tone than Judge Lourie's opinion, calling the exceptions "limited"[257] and pointing out that "[a]ny claim can be stripped down, simplified, generalized, or paraphrased to remove all of its concrete limitations, until at its core, something that could be characterized as an abstract idea is revealed."[258]

Under Chief Judge Rader's approach, the primary inquiry "is whether a claim includes *meaningful limitations* restricting it to an application, rather than merely an abstract idea."[259] The opinion discusses different ways for determining whether a limitation is sufficiently meaningful. First, "a claim is not meaningfully limited if it merely describes an abstract idea or simply adds 'apply it.' "[260] Second, a claim "will not be limited meaningfully if it contains only insignificant or token pre- or post-solution activity—such as identifying a relevant audience, a category of use, field of use, or technological environment."[261] Finally, "a claim is not meaningfully limited if its purported limitations provide no real direction, cover all possible ways to achieve the provided result, or are overly-generalized."[262] These three concepts are found in Supreme Court precedent and are also discussed by Judge Lourie.[263] Consequently, they fail to illustrate how the proponents of the weak view of § 101 differ from the proponents of the strong view. It is only when the opinion analyzes the system claims does Chief Judge Rader reveal how he would approach patent-eligibility questions for software patents.

256. *Id.* at 1297 (emphasis added).

257. *Id.*

258. *Id.* at 1298.

259. *Id.* at 1299.

260. *Id.* at 1300 (citing Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289, 1294, 1297 (2012)).

261. *Id.* at 1300–01 (citing *Mayo*, 132 S. Ct. at 1297–98; Bilski v. Kappos, 130 S. Ct. 3218, 3230–31 (2010); Diamond v. Diehr, 450 U.S. 175, 191–92 & n.14 (1981)); Parker v. Flook 437 U.S. 584, 595 n.18 (1978).

262. *CLS Bank III*, 717 F.3d at 1301 (Rader, J., concurring in part and dissenting in part) (citing *Mayo*, 132 S. Ct. at 1300).

263. *See* discussion *supra* Section V.A.

Chief Judge Rader's opinion starts with the system claims, which have the most concrete computer based limitations.[264] The main thrust of Chief Judge Rader's opinion is that each system claim "does not claim anything abstract in its machine embodiments."[265] In support of that conclusion, the opinion points out that the representative claim includes "at least four separate structural components: a *computer*, a first party *device*, a data storage *unit*, and a communication *controller* coupled via machine components to the computer and the first party device."[266] Relying on the specification, the opinion goes on to argue that the structural and functional limitations found in the claims should not be considered post-solution activity but integral to the performance of the claimed system.[267] For example, the specification describes how different computer components operate.[268] It also contains flowcharts that describe specific algorithms that support the recited functions.[269] Finally, the opinion also examines the claims from a more intuitive level. Because the claims contain concrete computer based

264. In contrast, Judge Lourie's opinion begins with the method claims, which appear to recite the least concrete limitations. *See supra* note 239 and accompanying text.

265. *CLS Bank III*, 717 F.3d at 1306 (Rader, C.J., concurring in part and dissenting in part).

266. *Id.* at 1307. Chief Judge Rader's first opinion analyzed claim 26 of the '375 patent as a representative system claim. The claim recites:

> A data processing system to enable the exchange of an obligation between parties, the system comprising: a communications controller, a first party device, coupled to said communications controller, a data storage unit having stored therein (a) information about a first account for a first party, independent from a second account maintained by a first exchange institution, and (b) information about a third account for a second party, independent from a fourth account maintained by a second exchange institution; and a computer, coupled to said data storage unit and said communications controller, that is configured to (a) receive a transaction from said first party device via said communications controller; (b) electronically adjust said first account and said third account in order to effect an exchange obligation arising from said transaction between said first party and said second party after ensuring that said first party and/or said second party have adequate value in said first account and/or said third account, respectively; and (c) generate an instruction to said first exchange institution and/or said second exchange institution to adjust said second account and/or said fourth account in accordance with the adjustment of said first account and/or said third account, wherein said instruction being an irrevocable, time invariant obligation placed on said first exchange institution and/or said second exchange institution.

Id. at 1306.

267. *Id.* at 1306–07.

268. *Id.*

269. *Id.*

limitations, Chief Judge Rader suggests that labeling the system claim an abstract concept "wrenches all meaning from those words, and turns a narrow exception into one which may swallow the expansive rule (and with it much of the investment and innovation in software)."[270]

At this point, Chief Judge Rader's opinion goes on to discuss the method without the concurrence of Judges Linn and O'Malley. In contrast to the system claims, Chief Judge Rader's opinion found that the method claims were not drawn to patent-eligible subject matter.[271] The opinion examined each limitation of the representative claim[272] and found that they were all inherent to the fundamental concept of an escrow. The fact that the parties had stipulated that the method claims required a computer was insufficient to save them.[273] Chief Judge Rader said that "implicit reference to computer 'implementation' is not, by itself, enough."[274] In sum, Chief Judge Rader concluded that the claim as a whole simply encompassed an abstract concept—namely, the entire concept of "using an escrow to avoid the risk of one party's inability to pay."[275]

In essence, Chief Judge Rader and his allies view the vast majority of claims that contain computer limitations as being patent-eligible. The concept of ignoring computer-based limitations because they are not integral to the underlying idea only applies for patents that lie on one end of the spectrum. Under this view, a claim that does no more than simply add a computer to an otherwise unpatentable idea remains unpatentable. But claims that describe how an idea is implemented on particular components should be patent-eligible, even if that description is extremely basic. Of

270. *Id.* at 1309; *see also id.* at 1319 (Moore, J., dissenting in part) ("Looking at these hardware and software elements, it is impossible to conclude that this claim is merely an abstract idea.").

271. Oddly, Chief Judge Rader's opinion does not explicitly analyze the computer-readable medium claims. However, he concludes in the same sentence that the "method and media claims" are not patent-eligible. *Id.* at 1313 (Rader, C.J., concurring in part and dissenting in part). Presumably, Chief Judge Rader is applying the same analysis to both types of claims.

272. *See supra* note 239 for the text of this claim, claim 33 of the '479 patent.

273. This is where Judges Linn and O'Malley disagree. Their opinion interprets the method claims more narrowly and argues that they require "more than the use of computer is some unspecified way." *CLS Bank III*, 717 F.3d at 1329–30 (Linn & O'Malley, JJ., dissenting). Accordingly, Judges Linn and O'Malley would treat the method claims just as Chief Judge Rader treats the system claims. *See id.* at 1330.

274. *Id.* at 1312 (Rader, J., concurring in part and dissenting in part).

275. *Id.*

course almost any patent attorney should be able to draft claims that meet this requirement.[276]

Although the proponents of the weak view ostensibly apply Supreme Court precedent, it is hard to reconcile their analysis with *Bilski* and *Mayo*. First, just one year earlier, *Mayo* had rejected a similar invitation to find "that virtually any step beyond a statement of a law of nature itself should transform an unpatentable law of nature into a potentially patentable application sufficient to satisfy § 101's demands."[277] Second, the "intuitive" understanding of what an abstract idea is that both Judges Rader and Moore apply is clearly inconsistent with *Mayo*. Indeed, a second separate opinion, Chief Judge Rader complains that "equating the personalized medicinal effect of a human-created pharmaceutical in patients of different metabolic rates and genetic makeup with the speed of light is only possible in a netherworld of undefined judicial insights."[278] Clearly, the Supreme Court has a broader view of what an abstract idea is than Chief Judge Rader and his allies. Finally, the proponents of the weak view do not acknowledge the hostility to business method patents found in *Bilski*.[279]

It is hardly surprising that the proponents of the weak view seem at odds with what the Supreme Court has said. The judges who signed on to the opinion appear to be quite dissatisfied with the Supreme Court's recent patent-eligibility jurisprudence. Judge Moore (joined by Chief Judge Rader and Judges Linn and O'Malley) writes that she is "concerned that the current interpretation of § 101, and in particular the abstract idea exception, is causing a free fall in the patent system."[280] Moreover, Chief Judge Rader's second opinion cites to *Bilski* and *Mayo* (as well as several Federal Circuit decisions) as evidence of the failure of § 101 jurisprudence.[281] In short, while the proponents of the weak view give an obligatory salute to the governing Supreme Court precedent, they suggest a very different approach. If anything, the opinions written by Judges Rader, Moore, and Linn/O'Malley appear to be more of a plea to the Supreme Court to lower the bar for patent eligibility and return it to the standard of the pre-*Bilski* and *Mayo* era.

Again, this Article does not explore the possibility of fundamental change to patent-eligibility requirements under § 101. Others may do so. Perhaps the

276. *See supra* note 186.

277. Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289, 1303 (2012).

278. *CLS Bank III*, 717 F.3d at 1335 (Rader, C.J., additional reflections). The opinion does not explicitly mention *Mayo*, but is clearly referring to that decision.

279. *See supra* notes 85–91 and accompanying text.

280. *CLS Bank III*, 717 F.3d at 1313 (Moore, J., dissenting in part).

281. *Id.* at 1333–36 (Rader, C.J., additional reflections).

Supreme Court should throw in the towel and rethink how it approaches basic questions of patent eligibility.[282] One such option is simply lowering the eligibility requirement along the lines suggested by Chief Judge Rader's opinion. But that view is not based on recent precedent. What is more, there are no signs that the Supreme Court will make such a radical shift.[283] In the meantime, the point-of-novelty approach suggested by this Article deals with the here and now. By working within the confines of recent Supreme Court decisions, it sets forth a practical framework for determining patent-eligibility questions raised by software patents.

Finally, Judge Newman's opinion merits some attention. Judge Newman's opinion represents her views alone. Although she would have found the all the claims were patent-eligible, her approach differs from the proponents of the weak view. Like Judge Rader, Judge Newman views the attempts to interpret patent-eligibility under § 101 as a failure.[284] But her solution is quite a bit simpler than the proponents of the weak view. Judge Newman argues that so long as a patent falls within the "useful arts" listed in § 101, the courts should only apply the "laws of novelty, utility, prior art, obviousness, description, enablement, and specificity."[285] Again, this is clearly not the current state of the law. Moreover, because the Supreme Court just rejected this very argument in *Mayo*, Judge Newman's approach is unlikely to provide the realistic solution that this Article seeks.[286]

VI. CONCLUSION

After the fractured decision in *CLS Bank*, the law is still in a state of flux and no one can say with certainty just what kind of software patents, if any, satisfy § 101's patent-eligibility requirement. This Article attempts to identify a realistic path out of the current morass by describing a test for determining when software patents cover patent-eligible subject matter. Relying on bits and pieces from existing precedent, the proposed point-of-novelty approach reins in harmful business method software patents without affecting their more deserving industrial cousins. Moreover, the theory does so without

282. *See* Lemley et al., *supra* note 44.

283. *See supra* text accompanying note 277.

284. *CLS Bank III*, 717 F.3d at 1321 (Newman, J., concurring in part and dissenting in part) ("[A]n all-purpose bright-line rule for threshold portal of section 101 is unavailable as it is unnecessary.").

285. *Id.* at 1322.

286. Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289, 1303 (2012) (rejecting the government's argument that §§ 102, 103, and 112 can perform the proper screening function for patents).

categorically declaring all business method patents ineligible, a step that the Supreme Court refused to take in *Bilski*.

In addition to answering an important doctrinal question, this Article also operates on a more theoretical level. It builds on earlier point-of-novelty works[287] and applies that thinking to one of the most vexing questions facing patent law today—patent eligibility determinations for software patents. This demonstrates that the proposed patent eligibility test is also rooted in a firm theoretical foundation. Moreover, by providing another example of a point-of-novelty solution, this Article hopes to reinforce the case for relying on point-of-novelty thinking more generally in patent law.

287. *See generally* Chao, *supra* note 15; Lemley, *supra* note 130; Chao, *supra* note 170.

REFORMING SURVEILLANCE LAW: THE SWISS MODEL

Susan Freiwald[†] *& Sylvain Métille*[††]

ABSTRACT

As implemented over the past twenty-seven years, the Electronic Communications Privacy Act ("ECPA"), which regulates electronic surveillance by law enforcement agents, has become incomplete, confusing, and ineffective. In contrast, a new Swiss law, CrimPC, regulates law enforcement surveillance in a more comprehensive, uniform, and effective manner. This Article compares the two approaches and argues that recent proposals to reform ECPA in a piecemeal fashion will not suffice. Instead, Swiss CrimPC presents a model for more fundamental reform of U.S. law.

This Article is the first to analyze the Swiss law with international eyes and demonstrate its advantages over the U.S. approach. The comparison sheds light on the inadequacy of U.S. surveillance law, including its recurrent failure to require substantial judicial review, notify targets of surveillance, and provide meaningful remedies to victims of unlawful practices. Notably, through judicial oversight and the requirement that surveillance practices be first approved by the legislature, the Swiss significantly restrict several law enforcement methods that U.S. law leaves to the discretion of the police. This Article explains the differences in approach as stemming from the greater influence of international human rights law in Switzerland and the Swiss people's willingness to engage in a wholesale revision of their procedural law.

In the United States, the courts and Congress have struggled to establish appropriate surveillance rules, as evidenced by recent controversial judgments in the courts and congressional hearings on ECPA reform. In the wake of recent disclosures about massive NSA surveillance programs that have relied on both foreign and domestic surveillance, U.S. citizens have grown increasingly concerned about the excessive use of new surveillance technologies to gather information about their private communications and daily activities. This Article analyzes the Swiss approach to domestic electronic surveillance, which, if adopted here, would significantly improve our laws.

---

† Professor of Law, University of San Francisco School of Law. I thank research librarian John Shafer and research assistants Sydney Archibald, Amy Leifur Halby, Everett Monroe, and David Reichbach for their valuable help. Josh Davis, Jim Dempsey and Judge Stephen Wm. Smith also contributed significantly to my thinking about this paper.

†† Doctor of Law and attorney at the Swiss bar, Lecturer, University of Lausanne, Faculty of Law and Criminal Justice and University of Fribourg International Institute of Management in Technology, Switzerland. This Article was mainly written during my time as a visiting scholar at the Berkeley Center for Law and Technology, University of California, Berkeley, School of Law. I thank research librarian Jean Perrenoud for his valuable help.

We appreciate the comments made by participants at the Privacy Law Scholars' Conference in June 2012, where we presented a draft of this paper: Bryan Cunningham, Danielle Citron, John Grant, Orin Kerr, Greg Nojeim, and Brian Pascal. We particularly thank Stephen Henderson, who moderated the panel devoted to our paper and furnished excellent guidance.

TABLE OF CONTENTS

I. INTRODUCTION

Calls for reform of American laws governing electronic surveillance have multiplied as members of Congress,[1] the judiciary,[2] and the public[3] have recognized that our outdated laws do not adequately protect citizens from law enforcement's abuse of modern surveillance technologies.[4] Congress passed the Electronic Communications Privacy Act ("ECPA")[5] in 1986 to bring government surveillance into the electronic age but has not meaningfully updated it since the advent of the World Wide Web.[6] Bills currently pending in Congress would make small, though significant, changes to ECPA. For example, they would strengthen the protection of location data[7] and stored email.[8] None of the bills proposed, however, would engage in a wholesale overhaul of the electronic surveillance legal regime.

---

1. *See Electronic Communications Privacy Act: Government Perspectives on Protecting Privacy in the Digital Age*: *Hearing Before the S. Comm. on the Judiciary*, 112th Cong. 62 (2011) (statement of Sen. Patrick Leahy, Chairman, S. Comm. on the Judiciary) (describing current electronic surveillance law as out of date and insufficient and in need of legislative update).

2. *See ECPA Reform and the Revolution in Location Based Technologies and Services: Hearing Before the S. Comm. on the Constitution, Civil Rights, and Civil Liberties of the H. Comm. on the Judiciary*, 111th Cong. 76–77, 85–91 (2010) (statement of Stephen Wm. Smith, U.S. Mag. J.) (explaining, for example, that because citizens do not receive notice of surveillance, they do not appeal issuance of warrants and thus the judiciary has insufficient opportunities to interpret and clarify vague aspects of the law); United States v. Jones, 132 S. Ct. 945, 964 (2012) (Alito, J., concurring) ("In circumstances involving dramatic technological change, the best solution to privacy concerns may be legislative.").

3. *See, e.g.*, Editorial, *The End of Privacy?*, N.Y. TIMES, July 14, 2012, at SR10 ("Clearly, federal laws need to be revamped and brought into line with newer forms of surveillance."); *About the Issue*, DIGITAL DUE PROCESS, http://digitaldueprocess.org/index.cfm?objectid=37940370-2551-11DF-8E02000C296BA163 (last visited Mar. 10, 2013).

4. *See It's Time for a Privacy Upgrade*, CENTER FOR DEMOCRACY AND TECHNOLOGY (Oct. 21, 2011), www.cdt.org/blogs/2010ecpas-25th-anniversary-time-change; Paul Ohm, *Probably Probable Cause: The Diminishing Importance of Justification Standards*, 94 MINN. L. REV. 1514, 1551 (2010) ("I agree with essentially everybody who has ever written about ECPA that the law is sorely in need of reform.").

5. 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.). Commentators tend to refer to the Act by its acronym, "ECPA," pronounced "eck-pah," and to drop the definite article when doing so.

6. *See infra* Part V (discussing the evolution of surveillance law in the United States).

7. *See* Online Communications and Geolocation Protection Act, H.R. 93, 113th Cong. (2013) (requiring a warrant for access to both stored email and location data).

8. *See* Electronic Communications Privacy Act Amendments Act of 2013, S. 607, 113th Cong. (2013) (requiring a warrant for access to stored email); Press Release, Patrick Leahy, U.S. Senator for Vt., Leahy Marks 25th Anniversary of ECPA, Announces Plan to Mark Up Reform Bill (Oct. 20, 2011), *available at* www.leahy.senate.gov/press/leahy-marks-25th-anniversary-of-ecpa-announces-plan-to-mark-up-reform-bill.

That overhaul is exactly what Switzerland accomplished when it unified its procedural laws. Switzerland took the opportunity to entirely update its surveillance laws to cover new technologies as well as traditional ones. In January 2011, the Swiss enacted a brand new statute, the Swiss Criminal Procedure Code ("CrimPC"), which covers all provisions for law enforcement surveillance under Swiss law.[9] Extending federal authority to enact CrimPC was complicated because it required an amendment to the Federal Constitution of the Swiss Confederation ("Swiss Constitution" or "Federal Constitution").[10] A series of decisions from the European Court of Human Rights, however, had set forth detailed requirements for law enforcement surveillance by signatories to the European Convention on Human Rights,[11] and the Swiss enacted CrimPC to comply with those decisions.[12]

With surveillance law reform on the agenda in the United States, the Swiss experience offers a unique opportunity to look at a law enforcement surveillance statute started from scratch. Rather than making piecemeal amendments to an entrenched set of rules, as pending bills in the United States currently propose, Swiss legislators started over, writing on a blank slate. Analyzing the resulting statute affords an unusual opportunity to consider what the United States might accomplish if its legislators were also willing to start entirely anew in the field of law enforcement surveillance. A sustained look at CrimPC can open U.S. eyes to new possibilities for surveillance law that reformers have not yet seriously entertained.

A comparison of the two countries' approaches also highlights systematic differences that strongly impact the balance of law enforcement powers and

---

9. CODE DE PROCÉDURE PÉNALE [CRIMPC] [Code of Criminal Procedure] Oct. 5, 2007, RS 312 (Switz.).

10. Before the amendment, the Confederation did not have the power to legislate over criminal law procedure or civil law procedure. The Federal Constitution of the Swiss Federation describes the process by which the people can amend the Swiss Constitution. A partial revision of the Constitution can be decreed by the Federal Assembly or any 100,000 persons eligible to vote. CONSTITUTION FÉDÉRALE [CST] [CONSTITUTION] Apr. 18, 1999, RO 101, art. 139 (Switz.). A revision needs to be adopted only by a majority of the Cantons and a majority of the eligible voters. CST art. 195. It is much easier to amend the Swiss Constitution than to amend the U.S. Constitution. *See generally*, SANFORD LEVINSON, OUR UNDEMOCRATIC CONSTITUTION: WHERE THE CONSTITUTION GOES WRONG (AND HOW WE THE PEOPLE CAN CORRECT IT) 160 (2006) ("no other country . . . makes it so difficult to amend its constitution").

11. Council of Europe Convention for the Protection of Human Rights and Fundamental Freedoms art. 8, Nov. 4, 1950, E.T.S. 5 [hereinafter ECHR], *available at* http://conventions.coe.int/Treaty/en/Treaties/Html/005.htm.

12. Switzerland is a member state of the Council of Europe but not of the European Union. *See infra* Section II.C.

privacy rights in each country. For example, Swiss law precludes the use of surveillance techniques not authorized and regulated by CrimPC; if the law does not explicitly permit and regulate a surveillance technique, such as using a brand new technology to gather data, law enforcement may not use it.[13] In the United States, by contrast, law enforcement considers itself free to use techniques that U.S. law does not yet regulate.[14] Consequently, as new surveillance methods come online, U.S. agents freely use them unless and until Congress tells them not to through regulation,[15] but Swiss agents may not use them unless and until their legislature authorizes them to do so. For example, before CrimPC, law enforcement agents could use GPS surveillance only in those Cantons that authorized it by statute. In the United States, the FBI felt free to use GPS devices to conduct surveillance without warrants, and scrambled to remove them only after the Supreme Court ruled that such surveillance was a search.[16]

This Article describes the passage of CrimPC and its key surveillance provisions, which govern surveillance of mail and telecommunications, the acquisition of user identification data, the use of technical surveillance devices, surveillance of contacts with a bank, the use of undercover agents, and surveillance through physical observation of people and places accessible to the general public.[17] After briefly explaining the structure and history of U.S. surveillance law, this Article contrasts those CrimPC provisions with existing U.S. law.

Before beginning a detailed comparison of the two countries' approaches to law enforcement surveillance, it is important to explain that the two countries, though radically different in size, are worthy subjects of comparison. Switzerland has always been a relatively independent country

13. *See infra* Section VII.H.1.

14. *Compare* Orin Kerr, *Internet Surveillance Law After the USA Patriot Act: The Big Brother That Isn't*, 97 NW. U. L. REV. 607, 645–47 (2003) (arguing that prior to their inclusion in a 2001 law, surveillance devices that recorded electronic addressing information were entirely unregulated and hence permitted without restriction), *with* Susan Freiwald, *Online Surveillance: Remembering the Lessons of the Wiretap Act*, 56 ALA. L. REV. 9, 72–73 (2004) (describing how courts have sometimes viewed practices not subject to statutory regulation as nonetheless subject to Fourth Amendment restrictions). The views of Professor Kerr, a principle author of an early version of the federal prosecutor's training manual, have generally prevailed. *See* COMPUTER CRIME & INTELLECTUAL PROP. SECTION, DEP'T OF JUSTICE, SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS vii (3d ed. 2009), *available at* www.justice.gov/criminal/cybercrime/docs/ssmanual2009.pdf.

15. State legislators may also constrain law enforcement use of new technologies, as may courts through application of constitutional constraints.

16. *See infra* note 83.

17. *See infra* Part VII.

that currently operates outside the strictures of the European Union,[18] although it shares many cultural values with other European countries.[19] As a Western European country, Switzerland is also a close cultural relative of the United States. While it has a lower homicide rate than the United States, it has a comparable number of burglaries and thefts per capita, a comparable number of professional judges and magistrates per capita, and a comparable number of police officers per capita.[20] Other comparative law articles have considered the United States and Germany, a country with geographic and language ties to Switzerland, but which is a member of the European Union and therefore less independent than Switzerland.[21]

Through a detailed, section-by-section comparison of each major surveillance provision of CrimPC to its U.S. counterpart, clear patterns

---

18. The European Council, sometimes called the Council of the European Union, is a body of the European Union; it consists of state or executive leaders from the member states who meet for the purpose of planning E.U. policy. *See* COUNCIL OF THE EUROPEAN UNION, www.consilium.europa.eu (last visited Mar. 13, 2013). Twenty-eight States are members of the European Union, but Switzerland is not among them. The European Council is sometimes confused with the Council of Europe, of which Switzerland is a member. *See infra* note 46.

19. The decisions of the European Court of Human Rights have significantly influenced Swiss law. *See infra* Section II.C.

20. *See* U.N. Office on Drugs & Crime, Theft at the National Level, Number of Police-Recorded Offences, www.unodc.org/documents/data-and-analysis/statistics/crime/CTS12_Theft.xls (last visited Mar. 14, 2013) (reporting theft rate per 100,000 population for the year 2010 as 1993.0 in the United States and 1560.3 in Switzerland); *Statistics on Burglary*, United Nations Office on Drugs and Crime, Burglary Breaking and Entering at the National Level: Number of Price-Recorded Offenses, www.unodc.org/documents/data-and-analysis/statistics/crime/CTS12_Burglary.xls (last visited Mar. 14, 2013) (reporting burglary rate per 100,000 population for the year 2010 as 695.9 in the United States and 812.1 in Switzerland); U.N. Office on Drugs & Crime, Statistics on Criminal Justice Resources: Total Police Personnel at the National Level, www.unodc.org/documents/data-and-analysis/statistics/crime/CTS12_Criminal_justice_resources.xls (last visited Mar. 14, 2013) (reporting police force per 100,000 population in the year 2008 as 232.3 in the United States and 215.6 in Switzerland); European Institute for Crime Prevention and Control, *International Statistics on Crime and Justice*, at 139, HEUNI Publication Series No. 64 (2010) (reporting the rate of professional judges per 100,000 population as 10.8 in the United States in the year 2001 and 10.6 in Switzerland in the year 2002). *But see* U.N. Office on Drugs & Crime, Intentional Homicide: Count and Rate per 100,000 Population, www.unodc.org/documents/data-and-analysis/statistics/crime/Homicide_statistics2012.xls (last visited Mar. 14, 2013) (reporting homicide rate per 100,000 population for the year 2010 as 4.2 in the United States and 0.7 in Switzerland.).

21. *See, e.g.*, Paul M. Schwartz, *German and U.S. Telecommunications Privacy Law: Legal Regulation of Domestic Law Enforcement Surveillance*, 54 HASTINGS L.J. 751 (2002); Paul M. Schwartz, *Evaluating Telecommunications Surveillance in Germany: The Lessons of the Max Planck Institute's Study*, 72 GEO. WASH. L. REV. 1244 (2003); Jacqueline E. Ross, *The Place of Covert Surveillance in Democratic Societies: A Comparative Study of the United States and Germany*, 55 AM. J. COMP. L. 493 (2007).

emerge, which illustrate the superior attributes of the Swiss approach. CrimPC provides greater coverage, less complexity, and more comprehensive protections for the Swiss people. First, CrimPC regulates more surveillance techniques than ECPA, the closest U.S. analog. For example, CrimPC restricts the use of undercover agents in law enforcement, but neither ECPA nor any other U.S. statute or constitutional provision regulates undercover operatives.[22] Also, as mentioned above, Swiss law precludes the use of unregulated techniques, whereas, subject to the Fourth Amendment, U.S. law enforcement agents make unlimited use of techniques not covered by ECPA.[23] Second, CrimPC is fundamentally easier to understand, which will surely make it easier for judges to apply. While commentators have criticized the complexity of ECPA rules that govern electronic communications surveillance, CrimPC's nearly uniform and technology-neutral approach contrasts strikingly with ECPA's thicket of categories and distinctions.[24] Finally, for those techniques that are covered by both CrimPC and ECPA, CrimPC almost always provides substantially greater protections against law enforcement abuse. In particular, CrimPC offers significantly greater judicial oversight, including by providing notice to targets that they have been the subjects of surveillance and real remedies for those who have been surveilled in violation of the law.

U.S. reformers should keep the Swiss approach in mind as they turn to ECPA reform in the coming months and years. In particular, Switzerland's requirement that statutory law must first authorize new surveillance techniques with appropriate restrictions before law enforcement may use them should encourage U.S. legislators to act quickly when faced with reports that U.S. agents are using new surveillance techniques to violate privacy. In addition, legislators should take critiques of the U.S. system more seriously, especially those founded on claims that current laws provide inadequate due process and call for better notice to targets, adequate remedies for improper investigations, and meaningful judicial oversight of

22. The Supreme Court has held that the Fourth Amendment does not apply to undercover surveillance. *See infra* Section VII.F.2. If undercover agents use wiretaps or other techniques regulated by ECPA, then those techniques are regulated, but the use of agents per se is not. *See id.*

23. *See infra* Section VII.B. Some states provide greater restrictions than ECPA for agents acting under the jurisdictions of those state statutes. *See generally* Stephen E. Henderson, *Learning from All Fifty States: How to Apply the Fourth Amendment and Its State Analogs to Protect Third Party Information from Unreasonable Search*, 55 CATH. U. L. REV. 373 (2006) (providing a comprehensive overview of state statutes that provide greater protection to targets of some surveillance practices than federal law).

24. *See infra* Sections VII.B and VII.C.

surveillance practices. Finally, legislators should seriously consider starting over with a regime that scraps ECPA's outdated and confusing categories and starts anew with a scheme that, like CrimPC, is clear, comprehensive, and, at least on its face, adequately protective of privacy rights.

II. THE SWISS LEGAL FRAMEWORK FOR SURVEILLANCE

A. SWISS LEGAL STRUCTURE

As in the United States, the Swiss legal system operates at both a federal and state level, with the states in Switzerland known as "Cantons."[25] The Swiss Confederation (also known as "Switzerland" or "Confederatio Helvetica") has 7.9 million inhabitants.[26] Each Canton may exercise the power over its own institutions given by the terms of the Federal Constitution.[27] Until the Federal Constitution was amended to provide federal power over all aspects of criminal and civil procedure, criminal law procedures, including surveillance for criminal law enforcement, were solely within the legislative competence of the Cantons.[28]

As in most European countries, the Constitution limits public activities.[29] The constitutional principle of legality requires that all activities of the State, including surveillance by state authorities, shall be based on and limited by enacted law.[30] CrimPC provides the specific legislative enactment required for law enforcement surveillance. Because everyone must abide by public regulations, whether or not they have individually consented to them, rights

25. CST art. 1.

26. 5.1 million people are eligible to vote in Switzerland. Arrêté du Conseil fédéral, constatant le résultat de la votation populaire du 23 septembre 2012 [Decree ascertaining the result of the vote of September 23, 2012] FF 1053, 1055 (2013), www.bfs.admin.ch/bfs/portal/fr/index/themen/01/02/blank/key/bevoelkerungsstand.html.

27. Jean-François Aubert & Etienne Grisel, *The Swiss Federal Constitution*, *in* INTRODUCTION TO SWISS LAW 15–25 (François Dessemontet & Tuğrul Ansay eds. 2004); THOMAS FLEINER, ALEXANDER MISIC & NICOLE TÖPPERWIEN, SWISS CONSTITUTIONAL LAW 122 (2005).

28. The Federal Constitution provides that the Cantons shall exercise all rights that are not vested in the Confederation. CST art. 3; JEAN-FRANÇOIS AUBERT & PASCAL MAHON, PETIT COMMENTAIRE DE LA CONSTITUTION FÉDÉRALE DE LA CONFÉDÉRATION SUISSE DU 18 AVRIL 1999 [SHORT COMMENTARY ON THE SWISS CONSTITUTION OF APRIL 18, 1999] 30–31 (2003); FLEINER, MISIC & TÖPPERWIEN, *supra* note 27, at 122–26; RENÉ A. RHINOW & MARKUS SCHEFER, SCHWEIZERISCHES VERFASSUNGSRECHT [SWISS CONSTITUTIONAL LAW] 147 (2009).

29. CST art. 5.

30. *See* CST art. 5; AUBERT & MAHON, *supra* note 28, at 39–50; Thomas Fleiner, *Cantonal and Federal Administrative Law of Switzerland*, *in* INTRODUCTION TO SWISS LAW, *supra* note 27, at 35–37.

and obligations can be imposed only if they arise from a statute, such as CrimPC.[31]

Written law, enacted by the legislature, is by far the most important source of law in Switzerland.[32] Different forms of written law have different hierarchical values that operate similarly to the hierarchical values of American laws. Constitutional rules prevail over ordinary acts, federal law takes precedence over cantonal law, and legislative statutes take priority over regulations promulgated by the Federal Council[33] or administrative authorities.[34] Both the Swiss Constitution and the European Convention on Human Rights ("ECHR") provide significant privacy rights that the legislature had to respect when enacting CrimPC.[35] The next two Sections discuss those privacy rights.

B. RIGHTS TO PRIVACY UNDER THE SWISS CONSTITUTION

At the constitutional level, the right to privacy derives primarily from Article 13 of the Swiss Constitution, which states that "everyone has the right to privacy in their private and family life and in their home, and in relation to their mail and telecommunications," and "everyone has the right to be protected against the misuse of their personal data."[36] The first sentence protects privacy in general and emphasizes the protection of the person and of his or her living quarters and workspace and his or her communications with others. The second sentence establishes the traditional protection of personal data, or what U.S. commentators refer to as "information privacy."[37] This informational self-determination right gives every person the power to decide whether and for which purpose personal information shall be processed.[38] As a fundamental right, the right to privacy limits the power of the State but cannot be invoked against other private persons.

---

31. A statute's legitimacy derives from the consent of the people expressed through the democratic adoption of the law.

32. In fact, the Swiss do not have judge-made common law as we do in the United States.

33. In Switzerland, the term "government" describes the executive branch, which is the Federal Council, composed of seven members. Each member is the head of one of seven departments that together form the federal administration. CST arts. 175, 178.

34. ANDREAS AUER, GIORGIO MALINVERNI & MICHEL HOTTELIER, DROIT CONSTITUTIONNEL SUISSE I [SWISS CONSTITUTIONAL LAW] 491–517 (2d ed. 2006).

35. Courts must also consider these rights when evaluating the application of a surveillance law to a particular person.

36. CST art. 13.

37. *See generally* DANIEL J. SOLOVE & PAUL M. SCHWARTZ, INFORMATION PRIVACY LAW (4th ed. 2011) (assembling cases and readings for law school courses on the protection of personal data).

38. Tribunal Fédéral [TF] [Federal Supreme Court] July 9, 2003, 129 ARRÊTS DU TRIBUNAL FÉDÉRAL SUISSE [ATF] I 232, 245–45; TF, May 29, 2002, 128 ATF II 259, 268.

The Swiss Supreme Court has refused to define the right to privacy, but it has made clear that the right covers every piece of personal data that is not publicly accessible.[39] Europeans generally view privacy as relating to the dignity and autonomy of the person.[40] Article 7 of the Swiss Constitution provides that human dignity must be respected and protected.[41] The right to personal freedom under Article 10 also protects human dignity.[42]

Although the right to privacy is considered a fundamental right, it is not absolute and can be subject to limitation. According to Article 36 of the Swiss Constitution, a restriction on the right of privacy, such as a statute that permits law enforcement surveillance, must satisfy four conditions: (1) it must have a legal basis, (2) it must be justified in the public interest or for the protection of the fundamental rights of others, (3) it must meet the standard of proportionality of means and ends,[43] and (4) it may not violate the essence of the fundamental right at stake.[44] When possible, courts interpret laws consistently with the Constitution.[45]

---

39. Some examples of personal data are: identification data, TF, Apr. 23, 1998, 124 ATF I 85, 87; medical data, TF, June 19, 1996, 122 ATF I 153, 155; data about sexual identity and orientation, TF, Mar. 3, 1993, 119 ATF II 264, 268; data about relationships with other human beings; and files of judicial proceedings, TF, Mar. 17, 1993, 199 ATF Ia 99, 101.

40. For further comparisons of American and European notions of privacy, see Paul M. Schwartz & Karl-Nikolaus Peifer, *Prosser's Privacy and the German Right of Personality: Are Four Privacy Torts Better than One Unitary Concept?*, 98 CALIF. L. REV. 1925 (2010); James Q. Whitman, *The Two Western Cultures of Privacy: Dignity Versus Liberty*, 113 YALE L.J. 1151 (2004); Francesca E. Bignami, *European Versus American Liberty: A Comparative Privacy Analysis of Antiterrorism Data Mining*, 48 B.C. L. REV. 609 (2007).

41. AUBERT & MAHON, *supra* note 28, at 67; JÖRG PAUL MÜLLER & MARKUS SCHEFER, GRUNDRECHTE IN DER SCHWEIZ IM RAHMEN DER BUNDESVERFASSUNG, DER EMRK UND DER UNO-PAKTE [BASIC RIGHTS IN SWITZERLAND ACCORDING TO THE FEDERAL CONSTITUTION, THE ECHR AND THE U.N. COVENANTS] 1–4 (2008).

42. CST art. 10 ("Everyone has the right to life. The death penalty is prohibited. Everyone has the right to personal liberty and in particular to physical and mental integrity and to freedom of movement. Torture and any other form of cruel, inhuman or degrading treatment or punishment are prohibited.").

43. Article 5 of the Swiss Constitution also mentions the principle of proportionality, which governs all activity of the State. CST art. 5.

44. According to the Swiss Constitution, the essence of fundamental rights is sacrosanct. CST art. 36; *see also* ANDREAS AUER, GIORGIO MALINVERNI & MICHEL HOTTELIER, DROIT CONSTITUTIONNEL SUISSE II 79–119 (2d ed. 2006); ULRICH HÄFELIN, WALTER HALLER & HELEN KELLER, SCHWEIZERISCHES BUNDESSTAATSRECHT 90–101 (7th rev. ed. 2008); WALTER HALLER, THE SWISS CONSTITUTION IN A COMPARATIVE CONTEXT 157–62 (2009).

45. Courts in the United States use the same interpretative approach, which is known as constitutional avoidance. *See, e.g.*, Edward J. DeBartolo Corp. v. Fla. Gulf Coast Building & Constr. Trades Council, 485 U.S. 568, 575 (1988) ("[E]very reasonable construction must

In summary, because CrimPC authorizes the restriction of fundamental rights during an investigation, the Swiss Constitution required that it be enacted as a federal law, that it be justified in the public interest to protect other fundamental rights, and that it respect the principle of proportionality and the essence of the right to privacy. These constraints no doubt contributed to CrimPC's comprehensive protections, which distinguish it from its significantly less protective U.S. counterparts.

C. RIGHTS TO PRIVACY UNDER THE EUROPEAN CONVENTION ON HUMAN RIGHTS

As a member of the Council of Europe,[46] Switzerland enacted the European Convention on Human Rights ("ECHR") in 1974, at which time it became directly binding in the Swiss legal system.[47] ECHR is an international treaty under which the member States of the Council of Europe promise to secure fundamental civil and political rights, both to their own citizens and to everyone within their jurisdictions. The European Court of Human Rights ("ECtHR"), a permanent international court based in Strasbourg and known for its progressive and dynamic interpretation of the Convention, enforces the ECHR. Judgments from the ECtHR are binding on the defendant country and persuasive in other signatory countries. The Court's case law spans more than fifty years.

The ECHR has played and continues to play an important role in shaping surveillance law in Switzerland and many other countries. The ECtHR develops its own case law and interprets the Convention so as to keep it current.[48] As a superior international body, the ECtHR governs how national courts apply the ECHR. Swiss courts are required to apply international law, and when domestic law conflicts with international law, international law

be resorted to, in order to save a statute from unconstitutionality." (quoting Hooper v. California, 155 U.S. 648, 657 (1895))).

46. The Council of Europe is an international organization located in Strasbourg, comprised of forty-seven European countries and established to promote democracy, protect human rights, and enforce the rule of law in Europe. *Who We Are*, COUNCIL OF EUROPE, www.coe.int/aboutcoe/index.asp (last visited Mar. 10, 2013).

47. In Switzerland, ratification of an international treaty like ECHR immediately incorporates the terms of that treaty into federal law. *See* FLEINER, MISIC & TÖPPERWIEN, *supra* note 27, at 43–45.

48. The European Court of Human Rights considers the ECHR to be a living instrument, which must (1) be interpreted in a dynamic and evolutionary way, (2) meet present day conditions, (3) be interpreted according to the purpose of the Convention, and (4) be interpreted such that the rights it grants are practical and effective. In addition, the Court must elucidate, safeguard, and develop the rules instituted by the Convention. *See* Golder v. United Kingdom, App. No. 4451/70, Eur. Ct. H.R. (1975) (hudoc.echr.coe.int).

prevails.[49] Swiss courts may not invalidate Swiss statutes on the grounds that they violate the Swiss Constitution. However, if a statute violates a provision contained in the Constitution and in the ECHR, the ECHR prevails on statutes and the provision of the statute that cannot be interpreted in accordance with the ECHR will not be applied to the case reviewed by the court.[50]

Like the Swiss Constitution, the ECHR establishes a right to privacy and provides similar protections. Article 8 of the ECHR states that "[e]veryone has the right to respect for his private and family life, his home and his correspondence."[51] The ECtHR views any State that chooses to employ new surveillance technologies as bearing a special responsibility to strike the right balance between the potential benefits of such surveillance techniques and the private lives with which they interfere.[52]

Like the Swiss Supreme Court, the ECtHR has not precisely defined "private life." It certainly covers the physical and psychological integrity of a person and incorporates the notion of personal autonomy.[53] It also protects a right to one's own identity and personal development, such as the right to establish relationships with other human beings and the outside world.[54] This right may also include protection for activities of a professional or business nature.[55] There is, therefore, a category of interaction people have with others that falls within the scope of one's "private life," even if conducted in the public sphere. A person's reasonable expectations of privacy may be a significant, although not necessarily conclusive, factor in determining whether he has a right to privacy.[56]

---

49. CST art. 190.

50. AUBERT & MAHON, *supra* note 28, at 1453–62.

51. ECHR art. 8.

52. S. & Marper v. United Kingdom, App. Nos. 30562/04, 30566/04, § 112, Eur. Ct. H.R. (2008) (hudoc.echr.coe.int) (finding that the retention of DNA profiles, samples, and fingerprints of persons not convicted of a crime violates Article 8 of the ECHR).

53. *Id.* § 66 (finding that the retention of DNA profiles, samples, and fingerprints of persons not convicted of a crime violates Article 8 of the ECHR).

54. Amann v. Switzerland, App. No. 27798/95, § 65, Eur. Ct. H.R. (2000) (hudoc.echr.coe.int).

55. *Id.*

56. *See, e.g.*, *Marper*, § 66 (hudoc.echr.coe.int) (finding that retention of DNA profiles, samples, and fingerprints of persons not convicted of a crime violates Article 8); Gillan & Quinton v. United Kingdom, App. No. 4158/05, § 61, Eur. Ct. H.R. (2010) (hudoc.echr.coe.int) (finding that U.K. law authorizing mandatory searches of persons at the discretion of police within a predetermined geographic area violates Article 8 of the European Convention on Human Rights).

A number of elements determine whether surveillance conducted outside a person's home or private property infringes on that person's private life. The Court has not enumerated those elements explicitly; rather, it considers each case as a whole and engages in fact-specific inquiries based on common norms. For example, in *Niemietz v. Germany* the ECtHR held that the notion of a "private life" is not restricted to an inner circle that entirely excludes the outside world; it also comprises the right to establish and develop relationships with other human beings.[57] The court held that a warrant for the search and seizure of any documents found in the applicant's office impinged on professional secrecy to an extent that was not proportional to the ends achieved under the circumstances.[58]

Like the Swiss Constitution, the ECHR permits some restrictions on the right to a private life. Article 8.2 provides:

> There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.[59]

Accordingly, any governmental interference in private lives must, among other things, (1) have some basis in domestic law, (2) have a legitimate aim, and (3) be necessary in a democratic society. The last requirement incorporates the notion that the means (e.g., surveillance) must be proportional to the ends achieved (e.g., law enforcement benefits).

Under the ECtHR's jurisprudence, surveillance generally constitutes an intrusion into private life.[60] In cases involving surveillance laws, the Court emphasizes seven requirements for any law authorizing government surveillance,[61] which explain why CrimPC provides much more

57. Niemietz v. Germany, App. No. 13710/88, § 29, Eur. Ct. H.R. (1992) (hudoc.echr.coe.int).

58. *Id.* (interpreting the words "private life" and "home" in Article 8 to include certain professional or business activities or premises).

59. ECHR art. 8.2.

60. Malone v. United Kingdom, App. No. 8691/79, § 64, Eur. Ct. H.R. (1984) (hudoc.echr.coe.int).

61. The recent cases of Kvasnica v. Slovakia, App. No. 72094/01, Eur. Ct. H.R. (2009) (hudoc.echr.coe.int), Calmanovici v. Romania, App. No. 42250/02, Eur. Ct. H.R. (2008) (hudoc.echr.coe.int), and Popescu v. Romania (No. 2), App. No. 71525/01, Eur. Ct. H.R. (2007) (hudoc.echr.coe.int), have confirmed the previous jurisprudence in cases such as Klass v. Germany, App. No. 5029/71, Eur. Ct. H.R. (1978) (hudoc.echr.coe.int), Malone, *supra* note 60, Kruslin v. France, App. No. 11801/85, Eur. Ct. H.R. (1990)

comprehensive privacy protection than comparable U.S. law. First, exploratory surveillance for preventive monitoring is prohibited.[62] Second, any surveillance should have a basis in domestic law and this law should be compatible with the rule of law and accessible to the person concerned who must, moreover, be able to foresee its consequences for him or her.[63] Third, data may only be used for the specific purposes for which it was collected.[64] Fourth, surveillance should be authorized by an independent body, preferably a judicial body, which is not in any way associated with the executive power.[65] In a later decision, the ECtHR elaborated that an independent judicial authority should authorize surveillance either before or after it takes place.[66] As the comparison between the two systems will show, CrimPC provides for significantly more judicial review than do the U.S. legal rules.

Fifth, the ECtHR requires such effective remedies as notification to the surveillance target within a reasonable time after the grounds necessitating the surveillance have ceased,[67] an opportunity to contest the surveillance or its effects on protected rights before an independent judicial authority,[68] and standing to bring a civil claim for any damage suffered as a result of the surveillance. Accordingly, CrimPC provides more extensive notice and more significant remedies than are available to the targets of surveillance in the United States. The sixth and seventh requirements provide data privacy rights that U.S. law generally does not afford.[69]

(hudoc.echr.coe.int), and Huvig v. France, App. No. 11105/84, Eur. Ct. H.R. (1990) (hudoc.echr.coe.int).

62. *See Klass*, § 51 (hudoc.echr.coe.int).

63. *See Kvasnica*, §§ 78–79 (hudoc.echr.coe.int); *Kruslin*, § 27 (hudoc.echr.coe.int); *Huvig*, § 26 (hudoc.echr.coe.int), *Popescu*, § 61 (hudoc.echr.coe.int), *Calmanovici*, §§ 118, 121 (hudoc. echr.coe.int).

64. *Calmanovici*, §§ 118, 121 (hudoc.echr.coe.int).

65. *See Klass*, § 56 (hudoc.echr.coe.int).

66. *See Popescu*, §§ 69–75 (hudoc.echr.coe.int). The Swiss Federal Supreme Court requires a judicial body to authorize surveillance beforehand and to consider objections to it afterwards when the surveillance pertains to communications. TF, Dec. 27, 1994, 120 ATF Ia 314, 318.

67. *See Popescu*, § 73 (hudoc.echr.coe.int).

68. *See Kruslin*, § 33–34 (hudoc.echr.coe.int); *Popescu*, §§ 73, 77 (hudoc.echr.coe.int).

69. Under the sixth requirement, the defendant should have access to data that could be used against him or her in a trial, at least by end of the investigation, and the defendant should have access to the original recordings until the end of the trial. *Popescu*, §§ 80–109 (hudoc.echr.coe.int). The surveillance target should also have the right to obtain review by a public or private expert of the authenticity or accuracy of the recording or associated transcript. *See Kruslin*, § 20(m) (hudoc.echr.coe.int); *Popescu*, §§ 80–81 (hudoc.echr.coe.int). The seventh requirement is that the law should indicate when and how data collected by surveillance shall be destroyed. *See Kruslin*, §§ 35, 52 (hudoc.echr.coe.int); *Popescu*, §§ 78–79

To summarize, to the extent it imposes a restriction on private life, surveillance law in Switzerland must have a legitimate aim and be necessary in a democratic society. It must be conducted only in accordance with enacted law, and the law must require that any surveillance be authorized by an independent body not associated with the executive branch. During that review, the independent body will also determine if the means of surveillance is proportional to the ends to be achieved. The target of surveillance must (1) be notified of the surveillance, (2) be provided access to the results of the surveillance, (3) have the opportunity to bring those results to an expert who can evaluate their authenticity, (4) have the opportunity to challenge the surveillance in court,[70] if so desired, and (5) be awarded damages if that challenge is successful. As we shall see, no comparable restrictions or rights underlie much of the surveillance that occurs in the United States.

Surveillance conducted according to CrimPC, therefore, is subject to challenge on the grounds that the statute conflicts with the ECHR.[71] Such a challenge, however, would likely fail because the Swiss legislature drafted CrimPC specifically to conform to ECtHR decisions and other national precedents involving the ECHR.[72] For example, to erase any uncertainty regarding the sufficient legal basis to use government monitoring software and IMSI-Catchers, the Federal Council proposed an amendment to the Parliament in 2013, which would add two new articles permitting the use of government monitoring software and IMSI-Catchers.[73]

In theory, the ECHR plays a similar role in Swiss law as the Fourth Amendment plays in U.S. law.[74] In practice, however, the ECHR has arguably shaped current Swiss law much more than the Fourth Amendment has influenced U.S. law because Swiss lawmakers have drafted legislation to comply with its mandates and because all law enforcement surveillance in Switzerland may proceed only according to that law.

(hudoc.echr.coe.int). Under U.S. law, the only comparable right is the wiretap target's right to request a copy of the recording. *See* 18 U.S.C. § 2518(8)(d) (2012).

70. CRIMPC art. 393.

71. If a court finds that a particular surveillance technique exceeds the mandates of CrimPC, it could render the results of the surveillance unusable. Typically, the legislature amends the law to address the technique.

72. Conseil Fédéral, Message relatif à l'unification du droit de la procédure pénale [Message about Unification of Criminal Procedure Law], FF 1057, 1075 (2006).

73. Conseil Fédéral, Message concernant la loi fédérale sur la surveillance de la correspondance par poste et télécommunication [LSCPT] [Message About the Modification of the Surveillance of Post and Telecommunications Act], FF 2379 (2013).

74. For further discussion of the Fourth Amendment, see *infra* Section III.A.

In the United States, by contrast, the Fourth Amendment protects against excessive surveillance more in theory than in practice. As Part III discusses, the U.S. Supreme Court has interpreted the Fourth Amendment to apply to a small subset of surveillance practices. Litigators for the Department of Justice ("DOJ") have endeavored to limit the scope of the surveillance practices subject to the Fourth Amendment and have generally achieved success in the courts. As a result, unlike the meaningful limits that the Swiss Constitution and the ECHR impose on surveillance practices in Switzerland, the Fourth Amendment constrains a limited subset of surveillance methods in the United States.

III. THE U.S. FRAMEWORK FOR SURVEILLANCE—COMPARED

A. U.S. LEGAL STRUCTURE

The structure of U.S. law is, at least superficially, similar to the structure of Swiss law. Both federal and state laws in the United States regulate law enforcement surveillance practices, with the U.S. Constitution providing a means to strike down laws that do not satisfy its mandates. In the United States, however, determining the applicable legal rule to govern a given act of law enforcement surveillance may not be easy. Government agents may conduct surveillance activities for law enforcement purposes and to gather foreign intelligence; different rules apply depending on the purpose of the surveillance.[75] Although federal legislation trumps inconsistent state legislation and provides a single law for federal actors all over the United States,[76] federal appellate courts differ as to how they interpret the federal surveillance provisions; consequently, the applicable rules vary by jurisdiction.[77] Finally, states have passed their own laws to regulate the surveillance practices of state and local law enforcement agents as well as private actors.[78] Those laws, which must respect the floor set by federal law,[79]

75. Other than a short discussion, *infra* Section V.C, this Article will not cover surveillance for foreign intelligence gathering.

76. Under federal statutory law, applications for wiretapping are made by federal law enforcement officials to federal magistrate judges for violations of federal law, and to state judges for investigation by state law enforcement agents of violations of state laws. *See* 18 U.S.C. § 2516(2)–(3) (2012).

77. *See, e.g.*, Ohm, *supra* note 4, at 1538–42 (describing how the Ninth Circuit interprets an ECPA provision pertaining to email surveillance differently from the Department of Justice).

78. *See, e.g.*, Charles H. Kennedy & Petper P. Swire, *State Wiretaps and Electronic Surveillance After September 11*, 54 HASTINGS L.J. 971, 977 (2003) (surveying state wiretap laws enacted since September 11, 2011). State statutes are subject to judicial review in either state

may be more restrictive of law enforcement practices and therefore more protective of privacy interests.[80] To avoid undue complexity, this Article will focus on federal statutes and federal constitutional law.

The most important difference between the Swiss and American legal systems lies not in the hierarchy of laws, but in the defaults that operate in the absence of legislation. Laws, both statutory and constitutional, *restrict* government action in the United States. That means that ECPA and the Fourth Amendment restrict government surveillance practices, but if they do not preclude a particular surveillance technique, government actors feel free to engage in it.[81] An example is the use of undercover agents, which neither a statute nor the Fourth Amendment regulate in the United States.[82] As previously discussed, the Swiss Constitution and the ECHR require enacted law to *authorize* their surveillance practices before they may be used. Once one understands what CrimPC covers, one knows the scope of law enforcement surveillance in Switzerland. Because law enforcement agents in the United States conduct surveillance until a statute or a court decision restricts them from doing so,[83] however, it is just as important to understand what statutory law (usually ECPA) and the Fourth Amendment do not cover as what they do. The comparison to CrimPC helps to bring that to light.

B. RIGHTS TO PRIVACY UNDER THE U.S. CONSTITUTION

Historically, judges have used the Fourth Amendment[84] to set standards when evaluating law enforcement surveillance practices.[85] Concerns about

---

or federal courts to ensure their compliance with both the federal and applicable state constitutions).

79. *See* Lane v. CBS Broad. Inc., 612 F. Supp. 2d 623, 637 (E.D. Pa. 2009) (reviewing legislative history to find that Congress intended for the federal law to set a baseline of protection above which states could legislate).

80. *See supra* note 23.

81. *See supra* note 14 and accompanying text.

82. *See infra* Section VII.F.2. CrimPC regulates the practice. *See id.*

83. *See, e.g.*, Kevin Johnson, *FBI Cuts Back on GPS Surveillance After Supreme Court Ruling*, USA TODAY, Feb. 7, 2012, www.usatoday.com/news/washington/story/2012-02-03/fbi-gps-surveillance-supreme-court-ruling/52992842/1 (reporting that the FBI had been operating under the assumption that use of GPS trackers did not require a court order or warrant prior to the Supreme Court's decision that it constituted a Fourth Amendment search); Julia Anguin, *FBI Turns Off Thousands of GPS Devices After Supreme Court Ruling,* WSJ.COM (Feb. 25, 2012), http://blogs.wsj.com/digits/2012/02/25/fbi-turns-off-thousands-of-gps-devices-after-supreme-court-ruling.

84. U.S. CONST. amend. IV. The Fourth Amendment requires that:

> [T]he right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no warrant shall issue, but upon probable cause, supported by Oath

First Amendment rights of free speech have also animated courts' reasoning in some surveillance cases,[86] but they have not yet provided an independent basis for review.[87]

The Fourth Amendment governs electronic surveillance practices more in theory than in practice. Courts have required challengers to overcome such hurdles as the requirement that they have standing to sue,[88] that the controversy be ripe for review,[89] and that the court cannot avoid the constitutional issue by statutory construction.[90] In addition, because many people targeted for law enforcement surveillance never learn about that surveillance, they cannot bring challenges to those practices of which they are unaware.[91] Finally, the federal appellate courts have taken few cases that

---

> or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.

Id.

85. *See, e.g.*, Berger v. New York, 388 U.S. 41, 51–53 (1967) (reviewing the history of the U.S. Supreme Court's surveillance decisions); United States v. Warshak, 631 F.3d 266, 283–88 (6th Cir. 2010) (finding federal surveillance statute unconstitutional to the extent it permits law enforcement access to stored email without a warrant).

86. *See, e.g.*, United States v. U.S. Dist. Court, 407 U.S. 297, 314 (1972) ("The price of lawful public dissent must not be a dread of subjection to an unchecked surveillance power. Nor must the fear of unauthorized official eavesdropping deter vigorous citizen dissent and discussion of Government action in private conversation.").

87. *See generally* Daniel J. Solove, *The First Amendment as Criminal Procedure*, 82 N.Y.U. L. REV. 112, 165–76 (2007) (identifying implications of electronic surveillance for First Amendment interests).

88. *See, e.g.*, Jewel v. NSA, 673 F.3d 902, 912 (9th Cir. 2011) (reversing lower court decision that plaintiffs lacked standing to challenge widespread warrantless surveillance of their communications phone calls and emails as part of terrorist surveillance program); ACLU v. NSA, 493 F.3d 644, 657 (6th Cir. 2007) (finding that plaintiffs lacked standing under Fourth Amendment to challenge the same practices).

89. *See, e.g.*, Warshak v. United States 532 F.3d 521, 525–34 (6th Cir. 2008) (en banc) (denying claim for injunctive relief from law enforcement surveillance on the grounds that claim was not ripe).

90. *See supra* note 45 and accompanying text; *see also* Susan Freiwald, *Cell Phone Location Data and the Fourth Amendment: A Question of Law, Not Fact*, 70 MD. L. REV. 681, 695 (2011) [hereinafter Freiwald, *Cell Phone Location Data*] (discussing successful arguments in recent case that courts should avoid constitutional ruling); Susan Freiwald, *The* Davis *Good Faith Rule and Getting Answers to the Questions That* Jones *Left Open*, 14 N.C. J. L. & TECH. 341 (2013) (discussing how courts are avoiding constitutional analysis by relying on a recent expansion in the exceptions to the exclusionary rule).

91. *See infra* Section VII.C. (discussing how some statutes require notice to targets of surveillance); *see also* Stephen Wm. Smith, *Gagged, Sealed & Delivered: Reforming ECPA's Secret Docket*, 6 HARV. J. L. & POL'Y REV., 313, 328 n.83 (2012) (discussing huge number of electronic surveillance orders that do not lead to prosecutions and of which the targets never obtain notice).

pertain to surveillance.[92] Among those few instances when higher-level courts do take on cases involving modern day surveillance questions, those courts often avoid the constitutional analysis altogether.[93]

The Supreme Court did issue a constitutional decision in 2012 in *United States v. Jones*, a case that addressed law enforcement's use of a GPS tracker attached to a car for an extended period.[94] Although all nine Justices agreed that the practice implicated the Fourth Amendment, the fractured opinion yielded no clear constitutional test beyond the facts of the case.[95] Importantly, the Court provided little guidance on how the Fourth Amendment applies, if at all, to location data surveillance accomplished by remote GPS tracking surveillance such as when officers monitor devices installed in cars or smartphones or when they acquire location data records from cell phone providers.[96] A broadly written decision might have motivated Congress to dramatically revamp ECPA, but the narrow decision in *Jones* certainly did not.[97] Even after *Jones*, litigants continue to debate how to apply decades-old precedents to modern surveillance methods.[98]

The older cases do make some things clear. In *Berger v. New York*, the Supreme Court found unconstitutional a New York statute that regulated electronic surveillance because the state law did not impose sufficient

92. *See* Smith, *supra* note 91, at 326–31 (discussing lack of appellate oversight of electronic surveillance cases).

93. *See* City of Ontario v. Quon, 560 U.S. 746, 130 S. Ct. 2619, 2629 (2010) ("The judiciary risks error by elaborating too fully on the Fourth Amendment implications of emerging technology before its role in society has become clear.").

94. United States v. Jones, 132 S. Ct. 945 (2012).

95. *See id.* at 954 (noting that a later case may require the Court to resort to a reasonable expectation of privacy but that the present case could be resolved on the basis of trespass); *see also* Paul Ohm, *United States v. Jones Is a Near-Optimal Result*, FREEDOM TO TINKER (Jan. 23, 2012), https://freedom-to-tinker.com/blog/paul/united-states-v-jones-near-optimal-result (describing it as positive that Court issued a narrow decision and avoided the debate over "reinventing Katz"). For further discussion, see *infra* Section VII.C.2.e.

96. *See* sources cited *supra* note 90 (discussing cases addressing surveillance through acquisition of location data from cell phone service providers and the questions *Jones* left unanswered).

97. For example, had Justice Sotomayor's concurrence been the majority decision, it would presumably have made any use of GPS tracking a search and dramatically undermined ECPA's lesser protection for electronic communications held by third parties. *See Jones*, 132 S. Ct. at 955–57 (Sotomayor, J., concurring).

98. *See, e.g.*, Brief for the United States at 16–26, *In re* Application of the U.S. for Historical Cell-Site Data, No. 11-20884 (5th Cir. Feb. 15, 2012), 2012 WL 1197699 [hereinafter Government Brief 5th Circuit] (arguing that Supreme Court cases from the 1970s and 1980s determine the outcome of the case).

procedural hurdles on law enforcement agents.[99] In *Katz v. United States*, concurring Justice Harlan formulated the reasonable expectation of privacy test[100] and the majority opinion announced that surveillance practices that intrude upon such expectations must comply with the restrictions set out in *Berger*.[101] In a series of cases in the late 1980s and early 1990s, seven federal courts of appeal extended the core Fourth Amendment protections established in *Berger* to government use of video surveillance cameras that record activities subject to a reasonable expectation of privacy.[102] The appellate courts found video surveillance to share the features of wiretapping that make it particularly prone to abuse in that such surveillance is hidden, indiscriminate, intrusive, and continuous and therefore it must be subject to the same restrictions as wiretapping.[103]

The crucial question in the United States is whether the law enforcement practice at issue constitutes a "search" under the Fourth Amendment like wiretapping, bugging, and some types of silent video surveillance. Unlike in Switzerland, constitutional privacy principles apply only to that subset of practices that are considered to be such searches. Practices that are not searches under the Fourth Amendment are subject to no constitutional regulation, and are regulated, if at all, by Congress, subject to no constitutional constraints.

In two important cases, the Supreme Court significantly limited what surveillance-type practices count as constitutional searches. In *United States v. Miller*, the Court found no Fourth Amendment search when law enforcement agents compelled a bank to produce records of the defendant's transactions

99. Berger v. New York, 388 U.S. 41, 60 (1967) (emphasizing the need for "adequate judicial supervision or protective procedures").

100. Katz v. United States, 389 U.S. 347, 360–62 (1967) (Harlan, J., concurring).

101. *See Katz*, 389 U.S. at 354–56 (noting that a judicially-authorized warrant that had "carefully limited use of electronic surveillance" could have been acceptable).

102. *See* Susan Freiwald, *First Principles of Communications Privacy*, 2007 STAN. TECH. L. REV. 3, ¶¶ 53–56.

103. *See id.*; *see also* Freiwald*, Cell Phone Location Data*, *supra* note 90, at 746–49 (arguing that these four factors—"hidden, indiscriminate, intrusive, and continuous"—should be used to find cell site location data protected by the Fourth Amendment); Brief for Yale Law Sch. Info. Soc'y Project Scholars et al. as Amici Curiae Supporting Respondent at 34–35, United States v. Jones, 132 S. Ct. 945 (2012) (arguing that the four factors should be used to find GPS tracking data protected by the Fourth Amendment). Arguments to extend the category of searches subject to the *Berger* standard beyond wiretapping, bugging and silent video surveillance to their modern analogues, such as that made in the *Jones* case, have not been successful. *But see In re* Application of the U.S. for an Order Directing a Provider of Elec. Commc'n Serv. to Disclose Records to the Gov't, 534 F. Supp. 2d. 585, 586 n.7 (W.D. Pa. 2008) (discussing four factors in reference to cell site location information).

with the bank such as his deposit slips and account statements.[104] The Court stated:

> [T]he Fourth Amendment does not prohibit the obtaining of information revealed to a third party and conveyed by him to Government authorities, even if the information is revealed on the assumption that it will be used only for a limited purpose and the confidence placed in the third party will not be betrayed.[105]

Government litigators and academics have disagreed over the implications of *Miller*. Some have argued that it establishes that the Fourth Amendment does not protect information obtained from a third party, which would include records of electronic communications stored with service providers.[106] Others have promoted a narrow construction of *Miller*,[107] under which, for example, customers would not forfeit their Fourth Amendment interests by sharing information with intermediaries such as electronic communication providers.[108] Whatever the proper application of *Miller* to new technologies, it clearly inspired Congress to provide only limited restrictions on law enforcement access to stored electronic records in ECPA.[109]

The Supreme Court extended *Miller* to the communications context in 1979 when it found law enforcement acquisition of dialed telephone numbers not to be an unconstitutional search in *Smith v. Maryland*.[110] Law enforcement agents used a device known as a "pen register" to obtain the

104. United States v. Miller, 425 U.S. 435, 442–45 (1976).

105. *Id.* at 443.

106. *See, e.g.*, Final Reply Brief for Defendant-Appellant United States of America at 17, Warshak v. United States, 532 F.3d 521 (6th Cir. 2008), 2007 WL 2085416 (proposing the rule that "the government may compel an entity to disclose any item that is within its control and that it may access").

107. *See, e.g.*, Patricia L. Bellia, *Surveillance Law through Cyberlaw's Lens*, 72 GEO. WASH. L. REV. 1375, 1403–09 (2004) (arguing that a broad reading of *Miller* is inconsistent with *Katz*); Deirdre K. Mulligan, *Reasonable Expectations in Electronic Communications: A Critical Perspective on the Electronic Communications Privacy Act*, 72 GEO. WASH. L. REV. 1557 (2004). Under a narrow construction, the *Miller* case would apply only when the target has knowingly and voluntarily shared his information with a service provider and the provider has stored the records in the ordinary course of its business. *See, e.g.*, *In re* Application of the U.S. for an Order Directing a Provider of Elec. Commc'ns Serv. to Disclose Records to the Gov't, 620 F.3d 304, 317–18 (3d Cir. 2010) (rejecting applicability of *Miller* to the acquisition of cell site location data).

108. *See, e.g.*, Patricia L. Bellia & Susan Freiwald, *Fourth Amendment Protection for Stored Email*, 2008 U. CHI. LEGAL F. 121, 158–69 (2008). In *Miller*, government agents acquired Miller's records from his bank, which was considered a party to his bank records. *Miller*, 425 U.S. at 438, 440–41.

109. *See* H.R. Rep. No. 99-647, at 23, 73 (1986) (referring to the *Miller* case when explaining lesser protections for electronic communications in storage); *see also infra* Section VII.B.2e.

110. 442 U.S. 735, 745–46 (1979).

telephone numbers dialed on a telephone.[111] The Supreme Court considered the limited intrusiveness of the pen register investigation as well as the target's voluntary and knowing disclosure of his telephone numbers to telephone company employees when it found the technique to intrude on no reasonable expectation of privacy.[112] As with the *Miller* case, the *Smith* decision does not have to be read to imply a lack of constitutional protection for modern electronic communications information.[113] Justice Department litigators, however, have maintained that *Smith* establishes that all "non-content" information lacks Fourth Amendment protection.[114] Whatever the appropriate reading of the case, it inspired Congress to provide for relatively little restriction in ECPA on law enforcement access to communication attributes, which include all non-content features of communications.[115]

Miller and *Smith* established that the practices they considered—compelled disclosure of stored bank records and acquisition of telephone numbers dialed—fell entirely outside the protection of the Fourth Amendment because they were not "searches" that intruded upon the targets' reasonable expectations of privacy. Some U.S. courts have read *Miller* and *Smith* more expansively and have found modern surveillance practices, such as IP address and cell site location acquisition, to be similarly outside the protection of the Fourth Amendment.[116] Some courts have recently rejected such broad readings, and found new practices, such as stored email acquisition, to be constitutionally protected because they differ significantly

111. Pen registers were mechanical surveillance devices that originally recorded only the numbers dialed, and did not determine whether a call had succeeded, its duration or the identity of the parties to it. *See generally* Susan Freiwald, *Uncertain Privacy: Communication Attributes After the Digital Telephony Act*, 69 S. CAL. L. REV. 949, 982–89 (1996) (describing the mechanics of early pen registers and reviewing their evolution over time).

112. *Id.* at 741–44.

113. *See, e.g.*, United States v. Jones, 132 S. Ct. 945, 957 (2012) (Sotomayor, J., concurring) ("More fundamentally, it may be necessary to reconsider the premise that an individual has no reasonable expectation of privacy in information voluntarily disclosed to third parties . . . ." (citing *Smith*, 442 U.S. at 742, and *Miller*, 425 U.S. at 443)).

114. *See, e.g.*, Gov't Reply Brief at 2–3, *In re* Application of the U.S. for an Order Directing a Provider of Elec. Commc'n Serv. to Disclose Records to the Gov't, 620 F.3d 304 (3d Cir. 2010) (arguing that "non-content" cell-site location records are not subject to Fourth Amendment protection).

115. *See infra* Section VII.C.2; Freiwald, *supra* note 111, at 969–75, 993–1007 (describing how Congress accorded weak protections to communications attributes in the federal surveillance statutes).

116. *See, e.g.*, United States v. Forrester, 512 F.3d 500, 511 (9th Cir. 2008) (finding real-time collection of IP addresses by law enforcement agents to be unprotected by the Fourth Amendment); Government Brief 5th Circuit, *supra* note 98, at 25–26 (listing five federal "district court cases [that] have relied on Smith and Miller and rejected Fourth Amendment challenges to acquisition of historical cell-site records without a warrant.").

from the practices considered in *Miller* and *Smith* and instead more analogous to wiretapping and acquisition of postal mail.[117]

Congress retains complete discretion over how to regulate those practices that do not implicate the Fourth Amendment. Unlike Swiss legislators, Congress has not produced a comprehensive surveillance law that covers all types of surveillance used during law enforcement investigations. Instead, restrictions derive from piecemeal legislation such as ECPA, which has fallen out-of-date in the more than twenty-five years since its passage. As the next section shows, in the United States, there is nothing comparable to the restrictions imposed by the ECtHR to inspire or require Congress to bring U.S. laws up to date.

C. RIGHTS TO PRIVACY UNDER INTERNATIONAL LAW

The United States is not a signatory to the European Convention on Human Rights and is not a member of the Council of Europe. Nor is the United States a party to an international treaty that would regulate its national law enforcement practices directly, with the exception of the Convention on Cybercrime. Article 15 of the Convention on Cybercrime requires that parties to the treaty include safeguards which "provide for the adequate protection of human rights and liberties."[118] Individual state parties may determine which specific safeguards to impose, however, and the treaty imposes no specific due process requirements on the United States, nor does it empower an international enforcement body.[119]

The United States does not fully submit to treaty obligations that could impose restrictions like those imposed by the ECHR. For example, the United States is a party to the International Covenant on Civil and Political Rights, but during ratification the Senate declared non-self-executing[120] that part of the treaty that protected against unlawful interference with a person's

117. *See* United States v. Warshak, 631 F.3d 266, 288 (6th Cir. 2010) (holding that acquisition of stored email without a warrant is unconstitutional); *see also In re* Application of the U.S. for Historical Cell Site Data, 747 F. Supp. 2d 827, 846 (S.D. Tex. 2010) (finding warrantless acquisition of historical cell site location information to violate the Fourth Amendment), *vacated*, 724 F.3d 600 (2013).

118. Council of Europe, Convention on Cybercrime art. 15, Nov. 23, 2001, T.I.A.S. No. 13174.

119. Miriam Miquelon-Weisman, *The Convention on Cybercrime: A Harmonized Implementation of International Penal Law: What Prospects for Procedural Due Process?*, 23 J. MARSHALL J. COMPUTER & INFO. L., 329, 340–41 (2005).

120. *See* S. Treaty Doc. No. 95-20 (1992) (providing resolution that sections of the International Covenant on Civil and Political Rights listing the rights of individuals are not self-executing).

"privacy, family, home, or correspondence."[121] In the absence of additional legislation, a U.S. citizen cannot challenge surveillance on the basis of that treaty language. While the United States is a party to the International Court of Justice, only other state parties, not individuals or non-state organizations, can bring matters before it.[122] Therefore, no United States citizen can use its dispute resolution mechanisms to challenge domestic law enforcement surveillance.

The absence of a higher order treaty like the ECHR has left law enforcement surveillance in the United States to the discretion of Congress, constrained to a limited degree by the Fourth Amendment. As later sections of this paper will show, Congress has used its discretion to produce an electronic surveillance regime with less expansive coverage, more complexity, and less comprehensive privacy rights than the Swiss statutory regime of CrimPC, to which we now turn.

IV. SWITZERLAND: APPLICABLE LAW ENFORCEMENT SURVEILLANCE ACTS

A. THE LAWS PRIOR TO THE SWISS CRIMINAL PROCEDURE CODE ("CRIMPC")

Current regulations for the various types of surveillance practices stem from the historical regulation of the mail and telephone networks. In 1889, the federal Act on Telephones made the content of telephone calls secret.[123] This first law protected all users by treating all phone calls as private matters. Thirty years later, however, two laws gave significant surveillance power to the State by providing law enforcement authorities the right to access the content of telephone calls, telegraph messages, and mail.[124] Decades later,

121. International Covenant on Civil and Political Rights art. 17, Dec. 19, 1966, 999 U.N.T.S. 171.

122. Charter of the United Nations and Statute of the International Court of Justice art. 34 para. 1, June 26, 1945, 59 Stat. 1055.

123. Loi Fédérale Sur Les Téléphones (du 27 Juin 1889) Avec Les Changements Y Apportés Par La Loi Fédérale Du 7 Décembre 1894, Et Ordonnance Sur Les Téléphones (du 24 Septembre 1895), FF III 902 (1889), RO 11 256, *available at* www.amtsdruckschriften.bar.admin.ch/viewOrigDoc.do?id=10069429.

124. Loi fédérale du 14 octobre 1922 réglant la correspondance télégraphique et téléphonique [Federal Act Regulating Telegraph and Telephone Communications] RS 7872 (1922); Loi fédérale du 2 octobre 1924 sur le Service des postes [Federal Act on the Postal Service] (1924).

both acts were modified again to restrict surveillance so that it could no longer be used to investigate civil matters or minor crimes (non-felonies).[125]

Viewing private life as insufficiently protected by the law, the federal Parliament amended the Criminal Code to add offenses for breach of privacy or secrecy in 1968.[126] The new Criminal Code provisions should have protected citizens' privacy from individual and state surveillance, but the Swiss Supreme Court held that an official who conducted surveillance in violation of the Criminal Code was not guilty on the grounds that he was doing his official duty.[127] This case spurred reform proposals in the Swiss Parliament.

A few years after Switzerland enacted the ECHR in 1974, the ECtHR held in a case brought against Germany that any interference with an Article 8 privacy right needed some basis in domestic law.[128] Even with those changes to its Criminal Code, Switzerland had no clear rule of law for surveillance that satisfied the requirement of proportionality of means and end. Switzerland needed to update its surveillance law to conform to the requirements of ECHR as recently interpreted by the Court.

As a result, Parliament enacted the federal Act on Privacy Protection in 1979,[129] which endeavored to regulate secret surveillance using the same principles that regulated the search of a house or the conduct of an arrest. It enumerated the conditions for surveillance and provided legal protection for individual subjects. The Act's provisions covered surveillance of post, telephone, and telegraph traffic. CrimPC retains several of the Act's basic principles such as the conditions imposed on surveillance, the requirement of proportionality, and the subject's right to go to court to contest surveillance. Parliament also amended the Criminal Code to preclude courts from excusing official surveillance merely on the grounds that the breach was conducted as part of official duties.[130]

---

125. In the Swiss Criminal Code, felonies are distinguished from misdemeanors according to the severity of the penalties that the offense carries. CODE PÉNAL SUISSE [CP] [Criminal Code] Dec. 21, 1937, RS 311, art. 10. Felonies carry a custodial sentence of more than three years and misdemeanors carry a monetary penalty or a custodial sentence not exceeding three years. Contraventions are punishable by a fine. CP art. 103.

126. CP art. 179bis–179septies.

127. Tribunal Fédéral [TF] [Federal Supreme Court] Mar. 8, 1974, 100 ARRÊTS DU TRIBUNAL FÉDÉRAL SUISSE [ATF] Ib 13, para. 5.

128. Klass v. Germany, App. No. 5029/71, Eur. Ct. H.R. (1978) (hudoc.echr.coe.int).

129. Loi fédérale sur la protection de la vie privée du 23 mars 1979 (modifications de lois fédérales) RO 1170 (1979). The Act amended the Federal Act on Telegraph and Telephone Traffic and the Federal Postal Service Act.

130. CP art. 179octies.

The Swiss Parliament enacted the law that inspired the new CrimPC in 2002.[131] That law, known as the Surveillance of Post and Telecommunications Act ("SPTA"), brought all provisions pertaining to the surveillance of post and telecommunications together in the same act.[132] Parliament designed SPTA to be as uniform as possible and to protect every kind of letter, parcel, and telecommunication from surveillance.[133] It covered the content and attributes of letters and parcels,[134] phone calls (including Voice over IP), email, text messages, faxes, and pager transmissions.[135] The next section describes the passage of CrimPC.

B. CRIMPC

After seven years of work, a committee of experts charged with unifying criminal procedure developed a draft of CrimPC.[136] The experts designed

131. Loi fédérale sur la surveillance de la correspondance par poste et télécommunication ("LSCPT") [The Federal Act of October 6, 2000 on the Surveillance of Post and Telecommunications ("SPTA")], RS 780.1. Parliament passed the Federal Law on Undercover Investigation on June 20, 2003 and CrimPC now includes important rules from that law as well.

132. SPTA did not cover the use of tracking devices and video surveillance equipment because such surveillance was not yet within the federal power and was therefore allowed only pursuant to cantonal law, if at all. For more on the situation prior to the SPTA and SPTA in general, see THOMAS HANSJAKOB, *BÜPF/VÜPF*: KOMMENTAR ZUM BUNDESGESETZ UND ZUR VERORDNUNG ÜBER DIE ÜBERWACHUNG DES POST- UND FERNMELDEVERKEHRS [COMMENTARY TO THE SURVEILLANCE OF POST AND TELECOMMUNICATIONS ACT AND ORDINANCE] 1–18 (2006).

133. Conseil Fédéral, Message concernant les lois fédérales sur la surveillance de la correspondance postale et des télécommunications et sur l'investigation secrète du 1er juillet 1998 [Message concerning the Federal Acts on the Surveillance of Post and Telecommunications and Undercover Investigation of July 1, 1998], FF IV 3689, 3703 (1998).

134. GÉRARD PIQUEREZ, TRAITÉ DE PROCÉDURE PÉNALE SUISSE [TREATY OF SWISS CRIMINAL PROCEDURE] 615 (2006); Bernhard Sträuli, La surveillance de la correspondance par poste et télécommunication: aperçu du nouveau droit [Surveillance of Post and Telecommunications: an Overview of the New Law], in PLUS DE SÉCURITÉ—MOINS DE LIBERTÉ? LES TECHNIQUES D'INVESTIGATION ET DE PREUVE EN QUESTION [MORE SECURITY—LESS FREEDOM? INVESTIGATION TECHNIQUES AND EVIDENCE IN QUESTION] 95–99 (2003).

135. SPTA did not cover communications made in Internet public forums or chat rooms. Police officer interventions in such conversations would be covered under the CrimPC rules pertaining to undercover agents. Beat Rhyner & Dieter Stüssi, *Kommentar zu Art. 269–279 StPO (Commentary to articles 269–279 CrimPC)*, in POLIZEILICHE ERMITTLUNG 443 (Gianfranco Albertini, et al. eds., 2008); Beat Rhyner & Dieter Stüssi, *Kommentar zu Art. 286-298 StPO (Commentary to articles 286–298 CrimPC)*, in POLIZEILICHE ERMITTLUNG, *supra*, at 498–99.

136. The Federal Council submitted the draft to the legislative process along with the committee of experts in 2001; the committee had begun their work in 1994.

CrimPC to treat every method of surveillance consistently with the treatment of surveillance of post and telecommunications under SPTA.[137]

Although CrimPC passed with great support from the Swiss people in 2007, it required a constitutional amendment to pass into law.[138] CrimPC represented a significant change in that it replaced twenty-seven different codes of criminal procedure (twenty-six cantonal and one federal).[139] Because some Cantons had to make extensive administrative or organizational changes to conform to the new federal CrimPC, the legislature decided to delay the new law's introduction until January 1, 2011.[140]

CrimPC provides for a public prosecutor, among other duties, to lead preliminary proceedings, conduct the examination of witnesses and others, bring charges, and represent cases before the courts.[141] Newly created Compulsory Measures Courts offset the public prosecutor's power.[142] In addition to overseeing surveillance activities, the new courts approve pretrial and security detentions and authorize the deployment of undercover investigators.[143]

Swiss law significantly deters violations of CrimPC. Only government officials may use one of the surveillance measures listed under CrimPC, and only after satisfying its statutory requirements.[144] The Criminal Code prohibits the use of surveillance without authorization and treats any information gathered by such surveillance as illegally obtained and subject to

137. Conseil Fédéral, Message relatif à l'unification du droit de la procédure pénale (Message about Unification of Criminal Procedure Law), FF 1057, 1099–1100, 1230 (2006).

138. All Cantons and 86.4% of the people eligible to vote approved the constitutional amendment needed. Arrêté du Conseil fédéral du 17 mai 2000 constatant le résultat de la votation populaire du 12 mars 2000, FF 2814–2820 (2000). According to the Swiss Constitution, the Confederation had the power to legislate over criminal and civil law but not over criminal law procedure or civil law procedure.

139. Under CrimPC, cantonal bodies continue to enforce substantive federal criminal law but comply in addition with the federal CrimPC.

140. CrimPC required many practical changes for some Cantons, especially those in the French part of Switzerland. Such Cantons, which used to have an independent and impartial investigating magistrate responsible for gathering the necessary evidence and conducting other pretrial steps, had to adopt the more adversarial prosecutorial role established in CrimPC.

141. CRIMPC art 16.

142. CRIMPC art 18.

143. *Id.* The Compulsory Measures Court is a regular court. *Id.* For more about the Compulsory Measures Courts, *see* André Kuhn, Procédure pénale unifiée: reformatio in pejus aut in melius? [Unified Criminal Procedure: Reformation in Pejus aut in Melius?] 45–49 (2008); Mark Pieth, Schweizerisches Strafprozessrecht: Grundriss für Studium und Praxis [Swiss Criminal Procedure Law: Basics for Academia and Practice] 63–64 (2009).

144. CP art. 179octies.

the exclusionary rule when challenged by the subject.[145] In addition, officials who conduct surveillance in violation of CrimPC risk disciplinary measures and prosecution.[146]

C. OTHER ACTS PERTINENT TO LAW ENFORCEMENT SURVEILLANCE

Swiss intelligence agencies do not conduct surveillance pursuant to CrimPC,[147] but instead operate according to the Internal Security Act ("ISA"), which addresses dangers relating to terrorism, illegal intelligence, violent extremism, and illegal arms and radioactive materials trade.[148] ISA permits preventative surveillance of those not suspected of criminal activity but limits surveillance under its auspices to publicly available information.[149] The Swiss Constitution does not require the limited intelligence surveillance under ISA to proceed with prior judicial authorization, unlike law enforcement surveillance under CrimPC.[150]

Since the enactment of CrimPC, the Swiss Criminal Code,[151] the Swiss Civil Code,[152] and the Federal Act on Data Protection do not generally

145. For more on the remedies for unlawful surveillance, *see infra* Section VI.D.

146. The provisions contained in the Criminal Code aim to avoid private surveillance and official surveillance without authorization, or "wild surveillance."

147. CrimPC does not apply to intelligence activities. Conseil Fédéral, Message relatif à l'unification du droit de la procédure pénale (Message about Unification of Criminal Procedure Law), FF 1057, 1112 (2006).

148. Loi fédérale du 21 mars 1997 instituant des mesures visant au maintien de la sûreté intérieure [The Federal Act on Measures to Safeguard Internal Security of March 21, 1997 ("LMSI")] RS 120 (1997). The ISA is used for all civil (non-military) surveillance conducted inside the country, whether or not the target is a Swiss citizen.

149. Intelligence agents may gather information through sources open to the public, and cantonal and federal authorities may transmit information to intelligence agencies. ISA art. 14. They may also conduct physical observation, video, and audio recording of public and freely accessible places.

150. *See supra* text at notes 65–66. The Government is currently drafting a bill that may allow for preventive surveillance. This surveillance would be subject to similar requirements to the ones in CrimPC (judicial plus political oversight, notice, and exclusionary rules). *See* Avant-projet de Loi fédérale sur le Service de renseignement civil (First Draft of Civil Intelligence Service Act), *available at* www.admin.ch/ch/f/gg/pc/ind2013.html.

151. The Swiss Criminal Code penalizes as misdemeanors unlawful entry (CP art. 186) and breach of postal or telecommunications secrecy (CP art. 321ter). It treats as felonies: breach of the privacy of a sealed document (CP art. 179), listening in on and recording the conversations of others (CP art. 179bis), unauthorized recording of conversations (CP art. 179ter), breach of secrecy or privacy through the use of an image-carrying device (CP art. 179quater), marketing and promotion of devices for unlawful listening or sound or image recording (CP art. 179sexies), misuse of a telecommunications installation (CP art. 179septies), and obtaining personal data without authorization (CP art. 179novies). *See* Sylvain Métille, *L'utilisation privée de moyens techniques de surveillance et la procédure pénale* (*Private Use of Surveillance and Criminal Procedure*), in "LE DROIT DÉCLOISONNÉ", INTERFÉRENCES ET INTERDÉPENDANCES ENTRE DROIT PRIVÉ ET DROIT PUBLIC ("DECOMPARTMENTALIZED

govern surveillance by law enforcement, but they do contain rules relevant to surveillance by private parties.[153] Law enforcement agents who conduct surveillance in accordance with CrimPC commit no offenses under these laws.[154]

V. UNITED STATES: APPLICABLE SURVEILLANCE ACTS

A. THE WIRETAP ACT

In 1968, Congress passed the Wiretap Act,[155] the precursor to ECPA, to codify the Fourth Amendment protections the Supreme Court had established in *Berger* the year before.[156] The Wiretap Act's procedural safeguards are closest to those provided by CrimPC, offering the highest level of judicial oversight of any of the surveillance laws in the United States. Under the Wiretap Act, for example, law enforcement agents must show that other less intrusive methods will not work before they may wiretap, and they must establish a tight nexus between the communications they seek to obtain and the criminal activity they are investigating.[157] Like CrimPC, the Wiretap Act requires that targets receive notice of the surveillance and provides real remedies for victims of improper investigations.[158]

But while the Wiretap Act has comprehensive protections like CrimPC, its coverage is dramatically more limited. The Wiretap Act applies to the use of traditional wiretaps (for telephone calls), bugs (to record oral conversations), and silent video surveillance conducted where targets have a reasonable expectation of privacy.[159] All other types of law enforcement

LAW," INTERFERENCES AND INTERDEPENDENCES BETWEEN PRIVATE LAW AND PUBLIC LAW) (Jean-Philippe Dunand & Pascal Mahon eds., 2009).

152. Art. 28 provides a general protection of legal personality: any person whose personality rights are unlawfully infringed may apply to the court for protection against any infringers. An infringement is unlawful unless it is justified by the consent of the person whose rights are infringed or by an overriding private or public interest or by law. STÉPHANE BONDALLAZ, LA PROTECTION DES PERSONNES ET DE LEURS DONNÉES DANS LES TÉLÉCOMMUNICATIONS (PROTECTION OF PERSONS AND THEIR DATA IN TELECOMMUNICATIONS) 146–56 (2007).

153. They apply, for example, to monitoring at the workplace or on private property.

154. CP art 179octies.

155. Omnibus Crime Control and Safe Streets Act of 1968, Pub. L. No. 90-351, Title III, 82 Stat. 212 (codified as amended at 18 U.S.C. §§ 2510–2522 (2012)). Commentators refer to the law as either "Title III" or the more intuitive "Wiretap Act."

156. Berger v. New York, 388 U.S. 41, 56–59 (1967); *see supra* Section III.B.

157. 18 U.S.C. § 2518 (2012); *see also* James G. Carr & Patricia L. Bellia, The Law of Electronic Surveillance § 4.17–4.48 (2011 ed.) (describing the requirements of the Wiretap Act).

158. *See infra* Section VII.B.2.

159. 18 U.S.C. § 2511 (2012); *see infra* Section VII.D.2. (describing how most federal appellate courts applied the substantive provisions of the Wiretap Act to silent video

surveillance must satisfy other statutes, such as ECPA, or are unregulated by federal statutory law.[160]

B. THE ELECTRONIC COMMUNICATIONS PRIVACY ACT ("ECPA")

While Congress endeavored to regulate the surveillance of modern communications technologies by passing ECPA in 1986 to amend the Wiretap Act,[161] ECPA's complexity has created considerable controversy about exactly what it covers.[162] ECPA extended some but not all of the Wiretap Act's protections to electronic communications' content and also includes entirely new provisions to govern some new surveillance practices Congress viewed as less intrusive than traditional wiretapping.[163]

ECPA contains three titles. The first extends the Wiretap Act provisions to the acquisition in real time of electronic communications such as email.[164] As this Article will discuss in more detail, it is easier for agents to obtain approval for such surveillance than for a traditional wiretap.[165] Significantly, and unlike under CrimPC, no information obtained in violation of ECPA is subject to a statutory exclusionary remedy, which significantly reduces ECPA's deterrent effect.[166] ECPA's second title, the "Stored Communications Act," addresses the acquisition of stored electronic information.[167] It has significantly fewer protections for such information than the first title and accords different protections to the contents of electronic communications and the non-content information associated with

surveillance by analogy despite the absence of explicit language in the Act); *see also supra* text accompanying notes 101–02 (describing federal appellate courts' finding that silent video surveillance is protected by the Fourth Amendment).

160. State law may provide greater regulation than federal law, both by providing greater coverage and by providing more comprehensive rights. But a discussion of state law is beyond the scope of this article. *See supra* note 23.

161. Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848 (1986) (codified as amended in scattered sections of 18 U.S.C.).

162. *See, e.g.*, Mink v. Salazar, 344 F. Supp. 2d 1231, 1239 (D. Colo. 2004) ("As several courts have noted, the [ECPA] is 'famous (if not infamous) for its lack of clarity.' " (citations omitted)).

163. *See supra* text accompanying notes 103–14.

164. Title I, Pub. L. No. 99-508, § 101, 100 Stat. 1848, 1848 (1986) (codified in scattered sections of 18 U.S.C.). There is no short form name given to the first title of ECPA.

165. *See infra* Section VII.B.2d).

166. It also reduces the number of cases brought to contest surveillance conducted according to its authority. *See* Orin S. Kerr, *Lifting the "Fog" of Internet Surveillance: How a Suppression Remedy Would Change Computer Crime Law*, 54 HASTINGS L.J. 805, 817 (2003); *see also* Freiwald, *supra* note 102, ¶¶ 19–35 (arguing that difficulties in determining constitutional questions have also inhibited their resolution).

167. Title II, Pub. L. No. 99-508, § 201, 100 Stat. 1848, 1860 (1986) (codified as amended at 18 U.S.C. §§ 2701–11 (2012)).

such communications—"communication attributes."[168] The third title, known as the "Pen Register Act,"[169] covers law enforcement use of pen registers and "trap and trace devices" to obtain dialing and addressing information for both wire and electronic communications.[170] Provisions in both the Stored Communication Act and the Pen Register Act restrict law enforcement surveillance significantly less than do comparable provisions in CrimPC.

C. THE USA PATRIOT ACT AND OTHER AMENDMENTS

Congress passed the USA PATRIOT Act in 2001 ("Patriot Act"),[171] just six weeks after the terrorist attacks of September 11.[172] Most of the Patriot Act's many provisions have nothing to do with surveillance, but a few of them further eased the restrictions on law enforcement surveillance.[173] For example, the Patriot Act amended ECPA so that acquisition of voicemail would receive the same reduced protection as stored electronic messages instead of the stronger protections that the Wiretap Act accorded telephone calls.[174] The Patriot Act also clarified that the weak provisions of the Pen Register Act would apply to the acquisition of electronic communication

168. *See* Freiwald, *supra* note 111, at 951 (introducing and explaining use of the term "communication attributes"). The statute treats different subcategories of communication attributes differently. *See infra* Section VII.C.2.

169. Title III, Pub. L. No. 99-508, § 201, 100 Stat. 1848, 1873 (1986) (codified as amended at 18 U.S.C. §§ 3121–27 (2012)).

170. Traditional pen registers acquired the telephone numbers dialed by the target's phone while trap and trace devices acquired the telephone numbers of the calling parties, revealing the same information as does caller ID. Modern pen registers acquire more detailed information.

171. United and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act, Pub. L. No. 107-56, 115 Stat. 272 (2001) [hereinafter USA PATRIOT Act].

172. For an insightful description of the legislative process that produced the Patriot Act, see generally Beryl A. Howell, *Seven Weeks: The Making of the USA PATRIOT Act*, 72 GEO. WASH. L. REV. 1145 (2004). Ms. Howell was a senior Democratic staffer at the time, and she argues that several Democrats valiantly resisted, sometimes successfully, some of the Administration's demands. *See id.* at 1165–66.

173. *See generally* Mark Eckenwiler, U.S. Dep't of Justice, Field Guidance on New Authorities that Relate to Computer Crime and Electronic Evidence Enacted in the USA PATRIOT Act of 2001, 701 PLI/PAT 1227, 1234 (2002) [hereinafter DOJ Field Guidance] (providing the government's perspective); *see also* Cindy Cohn, EFF Analysis of the Provisions of the USA Patriot Act that Relate to Online Activities, 701 PLI/PAT 1201 (2002) (critiquing several provisions' impact on electronic privacy rights).

174. *See* USA PATRIOT Act § 209, 115 Stat. 272, 283 (2001); DOJ Field Guidance, *supra* note 173, at 1232–33.

attributes, such as electronic mail addressing information, when that was previously unclear.[175]

Other than the Patriot Act, Congress has not significantly altered the statutory scheme just described. In 1994, Congress passed the Communications Assistance for Law Enforcement Act ("CALEA")[176] to ensure that providers of telecommunications services maintained the accessibility of their systems to wiretapping notwithstanding the introduction of digital communications technologies.[177] That Act did not significantly change the substantive restrictions on law enforcement surveillance.[178]

Unlike surveillance to detect terrorist threats in Switzerland,[179] surveillance for foreign intelligence gathering and to prevent terrorism in the United States has significantly fewer constraints.[180] Agents who operate under the Foreign Intelligence Surveillance Act[181] have considerably more discretion and may use all the surveillance tools of traditional law enforcement agents, subject to review only by a secretly impaneled court whose proceedings are not public.[182] Again, in contrast to Switzerland, where the ISA permits only the review of publicly available information, in the United States, extensive and secret surveillance generally proceeds without notice to the targets.[183]

---

175. The Patriot Act established that pen registers could be used to obtain "dialing, routing, addressing or signaling information" associated with electronic communications when it was previously unclear whether pen registers could obtain only the attributes of traditional telephone calls. *See* USA PATRIOT Act § 216, 115 Stat. 272, 288–90 (2001) (amending 18 U.S.C. § 3127(3)); DOJ Field Guidance, *supra* note 173, at 1233–34.

176. Communications Assistance for Law Enforcement Act (CALEA), Pub. L. No. 103-414, 108 Stat. 4279 (1994) (codified at 47 U.S.C. §§ 1001–1010 (2012) and in scattered sections of 18 U.S.C.).

177. *See generally* Freiwald, *supra* note 111 (describing the debates that accompanied the passage of CALEA).

178. *See id.*

179. *See supra* text accompanying notes 146–49.

180. A thorough discussion of foreign intelligence surveillance is beyond the scope of this Article. *See generally* DAVID S. KRIS & J. DOUGLAS WILSON, NATIONAL SECURITY INVESTIGATIONS & PROSECUTIONS (2007) (presenting the law governing investigations for national security rather than domestic law enforcement purposes); Peter Swire, *The System of Foreign Surveillance Law*, 72 GEO. WASH. L. REV. 1306 (2004) (reviewing the history of foreign surveillance laws and practices).

181. Foreign Intelligence Surveillance Act (FISA), 50 U.S.C. §§ 1801–1862 (2012) (covering the use of electronic surveillance and other investigatory techniques to pursue foreign intelligence).

182. *See* KRIS & WILSON, *supra* note 180, § 27; William C. Banks & M.E. Bowman, *Executive Authority for National Security Surveillance*, 50 AM. U. L. REV. 1, 89 (2000).

183. *See* KRIS & WILSON, *supra* note 180, § 31:2 (discussing how FISA applications and orders may not have to be disclosed to surveillance targets if the Attorney General files an

VI. COMMON ELEMENTS IN SURVEILLANCE PROCEDURES

Before detailing Swiss and U.S. surveillance regulations side-by-side, it helps to understand the types of procedures that regulate law enforcement surveillance. The following Sections describe the different procedural mechanisms and the range of choices among them that legislators have to choose from when drafting surveillance regulations. They cover such topics as the depth of judicial scrutiny and the scope of remedies for victims of improper surveillance.

A. LEVELS OF OVERSIGHT

CrimPC, which divides surveillance into six different methods,[184] requires surveillance under it to meet one of three different authorization processes depending on the intrusiveness of the surveillance method. For the most intrusive methods, CrimPC imposes the highest level of scrutiny, under which the Compulsory Measures Court[185] must confirm the propriety of the public prosecutor's order for police surveillance.[186] By contrast, the police may conduct the least intrusive methods of surveillance for up to a month without any prior judicial or prosecutorial authorization.[187] Intermediately intrusive methods require the prosecutor's prior authorization before law enforcement may conduct surveillance.[188]

affidavit stating disclosure would harm national security). In response to controversial large-scale monitoring programs conducted in the wake of the September 11th attacks, Congress amended FISA to provide immunity to service providers who aided such monitoring. *See* FISA Amendments Act of 2007, § 802, Pub. L. No. 110-261, 122 Stat. 2435, codified at 50 U.S.C. § 1885(a) (2012) (granting retroactive immunity to service providers). Recent disclosures of the extensive monitoring of domestic communications in the name of foreign intelligence came out too close to press time for the authors to assess them in this article. *See* The NSA Files, THE GUARDIAN, www.guardian.co.uk/world/the-nsa-files (last visited July 10, 2013) (compiling articles discussing, among other related pieces, the information revealed to the public by Edward Snowden).

184. The six methods are surveillance of post and telecommunications, acquisition of user identification data, use of technical surveillance equipment, surveillance of contacts with a bank, use of undercover agents, and physical observation of people and places accessible to the general public. *See infra* Part VI.

185. CrimPC established independent Compulsory Measures Courts to oversee law enforcement surveillance requests and perform other duties. *See supra* note 143.

186. If the Court does not confirm the prosecutor's order, the surveillance must terminate, and the results obtained from it cannot be used.

187. Police may continue surveillance after a month if they obtain the public prosecutor's authorization.

188. Both the police and the public prosecutor are considered to be law enforcement authorities. CRIMPC arts. 15–16.

U.S. law also requires a law enforcement agent to obtain the approval of a member of the judiciary, such as a trial judge or magistrate judge, before conducting intrusive forms of surveillance.[189] Fourth Amendment cases have noted the importance of having "a neutral magistrate" pre-approve searches and seizures to constrain the executive's zeal for law enforcement.[190]

Various members of the executive branch must also approve some surveillance methods before they may commence. Approval by high-level officials in the executive branch helps to inhibit unjustified investigations.[191] In some cases, the Attorney General himself must initially approve of a surveillance practice, although sometimes lower-level senior officials may approve. The requirement of high-level executive branch approval usually accompanies rather than substitutes for the requirement of judicial approval.

For a large number of surveillance methods, however, agents may conduct surveillance without submitting to any judicial oversight. For example, agents in the United States conduct a great deal of surveillance by issuing subpoenas, or demands for records.[192] In those cases, judges review the surveillance only when the target learns of it and brings a challenge.[193]

As this Article will discuss, ECPA treats some surveillance methods as insufficiently intrusive to require judicial oversight. In addition, surveillance

189. In some emergency situations, agents may conduct surveillance first and then obtain approval afterwards, with the statute specifying how much time the agent has to obtain judicial approval. *See, e.g.*, 18 U.S.C. § 2518(7) (2012) (permitting emergency wiretap orders which last up to forty-eight hours in limited circumstances).

190. *See* Dalia v. United States, 441 U.S. 238, 255–56 (1979).

191. *See, e.g.*, *In re* Sealed Case, 310 F.3d 717, 739 (FISC Ct. Rev. 2002) (noting that the requirement of written approval from senior officials provides an important check on arbitrariness).

192. *See* James X. Dempsey, *Digital Search & Seizure: Standards for Government Access to Communications and Associated Data,* 970 PLI/PAT 687, 702 (2009) (describing how prosecutors can issue subpoenas without any judicial involvement to access a variety of modern communications based on relevance to an investigation); *see also* Christopher Slobogin, *Subpoenas and Privacy*, 54 DEPAUL L. REV. 805, 824–25 (2005) ("The Supreme Court has applied *Miller*'s rationale to phone company records and loan applications, and lower courts have used it to uphold subpoenas for personal records from medical institutions, auditors and accountants, trustees in bankruptcy, and government institutions." (footnotes omitted)).

193. Department of Justice lawyers have argued that the when agents deliver a subpoena or similar order to a service provider, the subject of the records they seek may contest only on the basis that the subpoena or order seeks irrelevant information or that compliance would be too burdensome for the party who has to furnish the records, notwithstanding the subject's privacy interest in the records. *See* Susan Freiwald & Patricia L. Bellia, *The Fourth Amendment Status of Stored Email: The Law Professors' Brief in* Warshak v. United States, 41 U.S.F. L. REV. 559, 579–85 (2007) (describing and responding to the government's argument in the context of the compelled disclosure of stored email).

that proceeds outside of the bounds of ECPA (and related statutes), either by virtue of not being historically covered, or by virtue of being too new to be included, can proceed without any judicial review, so long as a court has not yet held that the Fourth Amendment requires regulation.[194]

B. CONDITIONS

1. Procedural Hurdles

CrimPC requires that agents have some suspicion of criminal activity before they may undertake surveillance; it does not permit preventative monitoring, where government agents use surveillance to prevent crimes from occurring in the first place.[195] Agents cannot use surveillance to create suspicion, as for example in so-called fishing expeditions.[196] Surveillance may not be undertaken unless a criminal offense has already been committed or is currently being committed;[197] it aims to discover the perpetrator or gather evidence related to a committed offense.[198] Swiss law supplies an equivalent to our probable cause standard by forbidding surveillance unless there is a strong suspicion that an offense has been committed. Physical observation, which may proceed according to an intermediate standard lower than strong suspicion but higher than simple suspicion,[199] is the only method that does not proceed according to the strong suspicion standard.[200]

Procedural hurdles in the United States vary considerably in terms of the burden they impose on law enforcement agents and the scope of discretion

194. Note that courts have limited jurisdiction, so only the Supreme Court can issue decisions that affect the entire United States. A Sixth Circuit decision requiring a warrant for access to stored email, for example, affected only investigations taking place in that Circuit. *See infra* text accompanying notes 293–97.

195. *But see* text accompanying notes 145–48 (noting that intelligence monitoring of public information can be used preventatively).

196. Peter Goldschmid, Der Einsatz technischer Überwachungsgeräte im Strafprozess: Unter besonderer Berücksichtigung der Regelung im Strafverfahren des Kantons Bern [Use of Technical Surveillance Equipment for Criminal Investigation: with Particular Attention to the Rules of Criminal Procedure in Canton of Bern] 95 (2001); HANSJAKOB, *supra* note 132, at 145.

197. CrimPC regulates the surveillance law enforcement conducts during an inquiry proceeding, which occurs when a criminal investigation is open and there is an (sometimes unidentified) accused person.

198. Acts in preparation for the commission of some particularly serious offenses are themselves independent offenses. They are intentional homicide (CP art. 111), murder (CP art. 112), serious assault (CP art. 122), robbery (CP art. 140), false imprisonment and abduction (CP art. 183), hostage taking (CP art. 185), arson (CP art. 221), genocide (CP art. 264), crimes against humanity (CP art. 264a) and war crimes (CP art. 264c–264h).

199. "Simple suspicion" is the standard for opening an investigation that does not use surveillance. CRIMPC art. 309.

200. *See infra* Section VII.G.1.

they afford to reviewing judges to deny government applications for surveillance. For the most restricted surveillance methods, judges require government agents to establish probable cause to believe the target "is committing, has committed, or is about to commit" a particular offense and that the surveillance will obtain incriminating communications about that offense.[201]

Some surveillance methods have standards that are much easier to meet than probable cause. One intermediate standard requires that the surveillance will yield information relevant to an ongoing criminal investigation instead of yielding evidence of criminal activity. Another even lower intermediate standard requires that the information sought will be relevant to a law enforcement inquiry. Standards are made less demanding both by using language with a broader scope, as just described, and also by limiting the judge's review to one that checks a surveillance application for completeness rather than conducting an independent review of the facts.[202] The lowest level of judicial review applies when judges review challenges to subpoenas. The recipient of a subpoena may generally challenge it only on the basis that it seeks irrelevant information or that compliance would be too burdensome for the party who has to furnish the records.[203]

Of course procedural standards that judges impose come into play only when judges themselves have a role in the surveillance process. Because a large amount of surveillance proceeds in the United States without any judicial review, or with unlikely and limited judicial review as in the case of subpoenas, judges are much less able to block problematic surveillance in the United States than in Switzerland.

2. Predicate Offenses

Although different methods of surveillance require different levels of seriousness, CrimPC permits law enforcement surveillance to investigate only serious criminal offenses. Agents may use some methods of surveillance only

201. *See* 18 U.S.C. §§ 2516(1), 2518(3)(a) (2012) (establishing the requirement under the Wiretap Act). That hurdle may be raised higher by a requirement that the communications device being surveilled has itself been used in the crime. *See* 18 U.S.C. § 2518(3)(b).

202. 18 U.S.C. § 3122(b) (2012).

203. A target may challenge a subpoena only when it is unreasonable or oppressive. United States v. Comprehensive Drug Testing, Inc., 621 F.3d 1162, 1191 (9th Cir. 2010) (en banc) (Bea, J., concurring in part and dissenting in part); Joshua Gruenspecht, *"Reasonable" Grand Jury Subpoenas: Asking for Information in the Age of Big Data*, 24 HARV. J. L. & TECH. 543, 547 (2011) (listing as "most widely accepted test for [the] reasonableness" of a subpoena: (1) whether the requested information is relevant, (2) whether the request is reasonably particularized, (3) whether the information requested covers a reasonable period of time).

to investigate a specific list of serious crimes,[204] while they may use others to investigate a wider range of crimes.

Similarly, some surveillance methods in the United States may be used only to investigate certain types of offenses, such as particularly serious crimes. Other statutes, however, permit surveillance methods for a wide variety of crimes or place no limit on the types of crimes that justify certain surveillance methods.

3. Other Limits

All Swiss surveillance practices must respect the subsidiarity principle and the need for proportionality between means and end. Subsidiarity requires that other less intrusive investigatory activities already conducted have not been successful or have no prospect of success; surveillance must not be the first investigatory activity.[205] Proportionality requires that the scope and duration of surveillance be as limited as possible. It means that the more invasive the surveillance method, the harder it will be to pass muster.[206] When courts conduct proportionality review they consider the seriousness of the offense, the invasion of privacy, the likelihood of success, and the length and type of the surveillance.

Unlike in Switzerland, where the subsidiarity rules apply to all surveillance covered by CrimPC, only surveillance methods covered by the Wiretap Act (wiretapping and bugging) require that less intrusive methods have failed or been shown to be infeasible.[207] Similarly, only the Wiretap Act requires that agents minimize the collection of non-incriminating conversations.[208] For all other surveillance methods in the United States, such as the vast majority of techniques that apply to modern communication methods, ECPA does not require that agents either minimize the collection of non-incriminating information or exhaust other types of surveillance

204. Several scholars have criticized the lists of offenses for reflecting politics rather than legal analysis. *See, e.g.*, Sträuli, *supra* note 134, at 124–27; HANSJAKOB, *supra* note 132, at 154–76.

205. *See* HANSJAKOB, *supra* note 132, at 152–54; NIKLAUS SCHMID, SCHWEIZERISCHE STRAFPROZESSORDNUNG, PRAXISKOMMENTAR [SWISS CRIMINAL PROCEDURE CODE: PRAXISCOMMENTARY] 505–06 (2009).

206. Other limits restrict surveillance to those set out in the order, see CRIMPC art. 278, and protect professional secrets. *See* CRIMPC art. 271; Sylvain Métille, *Le secret professionnel à l'épreuve des mesures de surveillance prévues par le CPP* [*Privileged information and surveillance ruled by CrimPC*], 03 MEDIALEX 131–37 (2011).

207. *See* 18 U.S.C. § 2518(3)(c) (2012). These requirements also apply to video surveillance in some cases. *See infra* note 365.

208. *See* 18 U.S.C. § 2518(5). Judges in individual cases may impose their own limits, but those appear to be rather rare.

first.[209] Some surveillance methods are, however and like in Switzerland, subject to a time limit that may be renewed upon a sufficient showing.[210]

The United States has no general requirement of subsidiarity or proportionality. As we shall see in the next Section, the lack of any proportionality requirement probably contributes the most to the comparatively lower restrictions on government surveillance in the United States. The other two significant factors are the ability of American agents to conduct surveillance without an authorizing statute and the lack of notice to targets for many types of surveillance.[211]

C. NOTICE

CrimPC requires notice for all methods of surveillance.[212] Swiss commentators view both the Swiss Constitution and the ECHR as mandating that law enforcement notify the targets of surveillance.[213] Notice provides the only official way for a target to learn about surveillance and opens the way for her to defend her rights.[214]

CrimPC requires notice even when surveillance does not provide any usable information, but notice may be postponed or even omitted if necessary for the protection of overriding public or private interests.

---

209. *See* Steve Jackson Games, Inc. v. U.S. Secret Serv., 36 F.3d 457, 463 (5th Cir. 1994) (explaining that only the interception provisions of the federal surveillance statutes have minimization requirements because agents can use keyword searching when going through stored communications). *But see infra* Section VII.D.2 (discussing silent video surveillance which federal appellate courts have found subject to the last resort, minimization, particularity, and limited duration requirements as a matter of constitutional, rather than statutory, law).

210. *See, e.g.*, 18 U.S.C. § 3123(c) (2012) (setting a limit of sixty days for investigations using pen registers unless the orders are renewed).

211. *See, e.g.*, Smith, *supra* note 91 (discussing lack of notice for much electronic surveillance, because of gag orders imposed on service providers, the sealing of judicial orders, and delays in conveying notice even when notice is required).

212. *See* SYLVAIN MÉTILLE, MESURES TECHNIQUES DE SURVEILLANCE ET RESPECT DES DROITS FONDAMENTAUX EN PARTICULIER DANS LE CADRE DE L'INSTRUCTION PÉNALE ET DU RENSEIGNEMENT [SURVEILLANCE MEASURES AND FUNDAMENTAL RIGHTS, WITH PARTICULAR ATTENTION TO CRIMINAL AND INTELLIGENCE INVESTIGATIONS] 182-183 (2011); CRIMPC arts. 279, 298. CrimPC calls notice "communication."

213. *See* HANSJAKOB, *supra* note 132, at 310; PIQUEREZ, *supra* note 134, at 627; Conseil Fédéral, Message relatif à la modification de la Loi fédérale instituant des mesures visant au maintien de la sûreté intérieure [Message related to the modification of the Internal Security Act], FF 4773, 4838 (2007).

214. Sylvain Métille, Mesures de surveillance secrètes: le rôle de l'information dans la protection des droits de l'individu [Secret surveillance measures: Notice as a protection of the rights of the surveilled person], 29 PLAIDOYER (2011).

Typically the court will permit notice to be postponed when notice without delay will ruin another ongoing investigation, but CrimPC requires that recourse to this exception be limited and instructs that courts should rarely permit notice to be omitted altogether.[215] The information obtained from surveillance may not be used if notice of that surveillance has not been provided to the target. After receiving notice, a surveillance target may contest violations of law including misuse or incorrect use of discretion and incomplete or incorrect establishment of the facts of the case before cantonal (trial) courts.[216]

Regardless of its result, the target should be informed of the surveillance by the public prosecutor as soon as possible and at the latest by the conclusion of the preliminary proceedings, which is when the public prosecutor transmits the case to the judge for a trial. Notice must identify the accused person and furnish the list of accused offenses, the reasons for surveillance, the nature and duration of surveillance, the identity of the person who granted the authorization, the conditions imposed on the surveillance, and the rights of the target as a result of the surveillance.[217] CrimPC provides much more extensive notice, and much more often, than does analogous law in the United States. Under American law, evidence obtained from surveillance but not subject to criminal discovery rules, or obtained about those who are not prosecuted, will never come to the target's attention unless an applicable statute requires notification.[218]

ECPA provisions vary in terms of who must receive notice, when agents must provide that notice, and the circumstances under which agents may delay providing notice.[219] ECPA does not require notice for many

215. CRIMPC art. 279.

216. CRIMPC art. 279, para. 3, art. 393, para. 2. Conseil Fédéral, Message relatif à l'unification du droit de la procédure pénale [Message about Unification of Criminal Procedure Law], FF 1057–1296 (2006); André Kuhn, *La procédure pénale suisse selon le futur CPP unifié*, 128 REVUE DE DROIT SUISSE 161–62 (2009).

217. SCHMID, *supra* note 205, at 525; HANSJAKOB, *supra* note 132, at 315–16.

218. *See* Smith, *supra* note 91, at 615–16 n.82 (doubting that criminal defense lawyers will learn of many online surveillance orders and noting that uncharged targets will not learn of much surveillance).

219. Several commentators have recommended that the United States amend its electronic surveillance statutes to provide better notice to targets. *See, e.g.*, Smith, *supra* note 91, at 332 ("ECPA should be amended to require notice to the target of any electronic surveillance order, including the customer, subscriber, or user of a targeted phone or Internet service."); Stephanie Pell & Christopher Soghoian, *Can You See Me Now?: Towards Reasonable Standards for Law Enforcement Access to Location Data that Congress Could Enact*, 27 BERKELEY TECH. L.J. 117, 185–89 (2012) (recommending notice when law enforcement obtains location data); Gruenspecht, *supra* note 203, at 561 (advocating for notice to be given to data creators instead of just third party intermediaries in the context of cloud computing).

surveillance methods and also precludes service providers that are involved in some surveillance methods from notifying targets.[220] Unregulated surveillance methods may, by definition, proceed without notice to targets.

D. CONSEQUENCES OF ILLEGAL SURVEILLANCE

CrimPC entitles the victim of unlawful surveillance to request from the court reasonable compensation and reparation for non-pecuniary loss such as emotional distress. CrimPC provides damages for economic losses but not punitive damages.[221] Both the accused people and third parties are entitled to compensation.[222]

Under CrimPC, data acquired using some surveillance methods without authorization[223] must be completely excluded from trial under what is known as an exclusionary remedy.[224] Under that approach, findings may not be used and data must be destroyed immediately.[225] For less intrusive surveillance methods like physical observation, CrimPC makes the results of unauthorized investigations relatively unusable: findings can be used only if they are necessary to solve serious offenses.[226] If the evidence could have been obtained legally, the court must weigh the competing interests of the prosecution in confirming suspicions and of the accused targets in protecting their personal rights.[227]

---

220. *See* Smith, *supra* note 91, at 610–14. Many orders to conduct surveillance are issued under seal (to be kept secret from the public, including the target), and remain under seal indefinitely. *See* Stephen Wm. Smith, *Kudzu in the Courthouse: Judgments Made in the Shade*, 3 FED. CTS. L. REV. 177 (2009) [hereinafter Smith, *Kudzu in the Courthouse*].

221. CRIMPC arts. 431, 434.

222. *Id.*

223. Surveillance is unauthorized when authorization has not been requested as needed, when the Compulsory Measures Court has refused to authorize it, and when surveillance proceeds past when it is authorized. CRIMPC arts. 277, 281, para. 4, 289, para. 6; TF, May 3, 2005, 131 ATF I 272, 281 (Switz.); HANSJAKOB, *supra* note 132, at 250–53 (2006). Whether or not an authorization would have been granted if requested is irrelevant. *See* TF, Oct. 9, 2007, 133 ATF IV 329, para. 4.4 (Switz.).

224. The ECtHR may opine on the fairness of the proceedings as a whole, including the way in which evidence was obtained. Schenk v. Switzerland, App. No. 10862/84, Eur. Ct. H.R. (1988) (hudoc.echr.coe.int).

225. The ECtHR has held that the exclusion at trial of evidence gained through any unlawful surveillance is a necessary but not sufficient remedy for the violation of the right to private life that may have occurred. Khan v. The United Kingdom, App. No. 35394/97, § 44, Eur. Ct. H.R. (2010) (hudoc.echr.coe.int); Taylor-Sabori v. The United Kingdom, App. No. 47114/99, §§ 22–24, Eur. Ct. H.R. (2002) (hudoc.echr.coe.int).

226. Conseil Fédéral, Message relatif à l'unification du droit de la procédure pénale [Message about Unification of Criminal Procedure Law], FF 1057, 1163 (2006).

227. TF, Sept. 7, 1983, 109 ATF Ia 244, para. 2.3 (Switz.).

In the United States, unlawful surveillance that violates the Fourth Amendment gives rise to a claim for money damages[228] and the protections of the suppression remedy.[229] The latter prohibits any evidence obtained by or derived from the unlawful surveillance from being introduced at the trial of the target of the surveillance. The suppression remedy is designed to deter law enforcement agents from acting unlawfully, but it is not always available.[230]

As discussed earlier, however, the Supreme Court has limited the Fourth Amendment's protection to that subcategory of investigations that intrude upon a target's "reasonable expectations of privacy" and that therefore constitute a "search." So far the Supreme Court has considered only wiretapping, bugging, and the installation and use of a GPS tracking device to be surveillance practices regulated under the Fourth Amendment.[231]

As distinct from the Constitution, the statutes that govern specific surveillance methods provide a range of remedies for noncompliance. Only the Wiretap Act provides a statutory suppression remedy; no such remedy is available for the improper interception of electronic communications.[232] As to damages, ECPA provides varied levels of monetary relief and the possibility of punitive damages and attorney's fees for some surveillance methods.[233] In limited cases, ECPA imposes criminal punishment or administrative discipline on law enforcement agents who conduct unlawful surveillance.[234] The executive branch rarely prosecutes its own agents, however.

---

228. A victim must bring a claim under 42 U.S.C. § 1983 (2012) (state actors) or the authority of *Bivens v. Six Unknown Named Agents of Federal Bureau of Narcotics*, 403 U.S. 388 (1971) (federal actors), to obtain such damages. *See, e.g.*, Warshak v. United States, 532 F.3d 521, 528, 532 (6th Cir. 2008) (expressing disapproval of target's pursuit of injunctive relief rather than a civil damages claim).

229. *See, e.g.*, Kyllo v. United States, 533 U.S. 27, 28 (2001) (reversing appellate court's denial of defendant's motion to suppress after finding that law enforcement agents conducted a "search" without a warrant).

230. *See, e.g.*, United States v. Warshak, 631 F.3d 266, 288–92 (6th Cir. 2010) (denying suppression remedy for constitutional violation when officers relied in good faith on statute that was not plainly unconstitutional).

231. *See supra* Section III.B. The Supreme Court has also treated law enforcement's use of a thermal imaging device to detect the heat emanating from a house as a search under the Fourth Amendment. Kyllo v. United States, 533 U.S. 27 (2001). As we discuss more, *infra* Section VII.G.2, the case's holding is limited. In the United States, moreover, because so few visual investigations require warrants, we tend not to think of them as electronic surveillance.

232. 18 U.S.C. §§ 2515, 2518 (2012).

233. 18 U.S.C. §§ 2520, 2707 (2012).

234. *Id.*

E. REPORTING

CrimPC does not require any particular reports about law enforcement surveillance practices. Information about surveillance practices may be available from the police or other bodies involved in surveillance, including from targets who have been notified of it. Apparently as a voluntary matter, some authorities have published reports about the monitoring of mail and telecommunications.[235] In the United States, Congress receives periodic reports about some surveillance methods. Such reporting facilitates the oversight that may constrain executive branch abuses.[236] Congress may choose to revise surveillance statutes in light of information it receives in surveillance reports. The surveillance statutes vary in how much detail must be provided to Congress, and some surveillance methods require no reporting at all. Compliance with the reporting requirements varies as well.[237]

VII. SURVEILLANCE REGULATION COMPARED

A. INTRODUCTION

Because CrimPC represents a modern and comprehensive statute designed to regulate all surveillance methods in one statute, we have organized the following discussion according to its six categories. CrimPC requires extensive judicial oversight for the most invasive techniques: surveillance of post and telecommunications,[238] use of technical surveillance devices,[239] surveillance of contacts with a bank,[240] and undercover operations.[241] CrimPC treats physical observation[242] as the least invasive method, requiring the least oversight by either a judge or public prosecutor. The acquisition of user identification data[243] is a subcategory of post and telecommunications surveillance and is considered less invasive than that method but more invasive than physical observation. As the following

235. *See Statistical Data*, POST AND TELECOMMUNICATIONS SURVEILLANCE SERVICE, www.li.admin.ch/en/themes/stats.html (last visited Feb. 2, 2013).

236. *See, e.g.*, *In re* Sealed Case, 310 F.3d 717, 741 n.25 (FISC Ct. Rev. 2002) (citing Senate report accompanying FISA).

237. Christopher Soghoian, *The Law Enforcement Surveillance Reporting Gap*, http://ssrn.com/abstract=18066628 (discussing how much modern electronic surveillance takes place without being publicly reported).

238. CRIMPC art. 269ss; *see infra* Section VII.B.

239. CRIMPC art. 280ss; *see infra* Section VII.D.

240. CRIMPC art. 284ss; *see infra* Section VII.E.

241. CRIMPC art. 286ss; *see infra* Section VII.F.

242. CRIMPC arts. 282–283ss; *see infra* Section VII.G.

243. CRIMPC art. 273ss; *see infra* Section VII.C.

discussion will show, ECPA[244] covers only a subset of the methods that CrimPC does. For some methods, such as tracking contacts with a bank, differences in other regulations and practices explain and make relatively uncontroversial why CrimPC but not ECPA covers them.[245] For other methods, however, such as the use of undercover government agents, the utter lack of regulation by U.S. law contrasts sharply with the many restrictions that Swiss law imposes.[246] The most glaring lack of coverage pertains to new methods of surveillance, which law enforcement agents in the United States have free rein to use until a court or legislature acts, but which require specific, legislative authorization in Switzerland. Regarding those methods of surveillance that both countries regulate, CrimPC clearly emerges as much less complex and much more comprehensive in its restrictions on law enforcement surveillance.

B. MONITORING OF POST AND TELECOMMUNICATIONS

1. In Switzerland

Swiss law enforcement agents must follow the most stringent procedures when conducting surveillance of an accused person's mail and telecommunications.[247] The pertinent category under CrimPC has an extremely wide scope due to its technology-neutral wording; it includes the interception of communications made by phone call, email, fax, text, pager, and Voice over IP, as well as the acquisition of any information in letters, parcels, and stored emails.[248] Surveillance conducted under this category may proceed in real time, for example when agents conduct a traditional wiretap

244. Technically, the Wiretap Act, which ECPA amended to cover electronic communications, still regulates the surveillance of traditional telephone calls and the installation of bugs. *See infra* Section VII.B.2.

245. *See infra* Section VII.E.

246. *See infra* Section VII.F. For a discussion of similar strong differences between U.S. surveillance law and that of other European countries, see Christopher Slobogin, *Transnational Law and Regulation of the Police,* 56 J. LEGAL EDUC. 451, 451–53 (2006) ("[T]ransnational law can provide interesting alternatives that might be worthy of adoption in the United States . . . . Denmark requires warrants for *any* undercover activity that requires infiltration, in stark contrast to our law essentially giving the police carte blanche in their undercover work.").

247. CrimPC permits the surveillance of the accused person's mail and calls and, in some cases, those of a third person directly connected to the accused. CRIMPC arts. 269–270ss.

248. August Biedermann, Bundesgesetz betreffend die Überwachung des Post- und Fernmeldeverkehrs (BÜPF) vom 6. Oktober 2000 [Surveillance of Post and Telecommunications Act (SPTA) of October 6, 2000], 120 REVUE PÉNALE SUISSE [SWISS CRIMINAL LAW REVIEW] 106 (2002); PIQUEREZ, *supra* note 134, at 615; HANSJAKOB, *supra* note 132, at 71–72; Sträuli, *supra* note 134, at 95–112.

of a telephone call or intercept an email, or it may proceed retroactively as when the police compel a third party service provider to produce an email from its system or a letter from its facilities. The Swiss recognize that the latter intrudes on the secrecy of communications because it may proceed without the person of interest being aware of it.[249]

The Compulsory Measures Court must approve all surveillance under this category and must confirm that the public prosecutor has a strong suspicion that an offense has been committed.[250] The offense must come from a list of serious predicate offenses.[251] Surveillance requests must be quite detailed[252] and they must establish, under the subsidiarity principle, that other investigatory activities have not been successful or have no likelihood of success.[253] As with all forms of surveillance in Switzerland, in determining whether to authorize surveillance, the court shall ensure that the scope and duration of the surveillance is as limited as possible to respect the principle of proportionality.[254]

The target must receive notice whenever the government conducts the surveillance of his mail or telecommunications.[255] Victims of unlawful monitoring of their post and telecommunications are entitled to damages and violators face criminal prosecution.[256] Victims are also entitled to have any

249. Police acquisition of such stored communications through search of a home, a computer, or a person, rather than from a service provider, or acquisition of computer materials directly from an accused person or his property constitutes a search and seizure. CRIMPC art. 263ss; Rhyner & Stüssi, *Kommentar zu Art. 269–279 StPO*, *supra* note 135, at 443–45; *see* HANSJAKOB, *supra* note 132, at 81–85; Sträuli, *supra* note 134, at 99–100, 107–08.

250. CRIMPC arts. 269, 273–274.

251. CRIMPC art. 269, para. 2.

252. They must include the reasoning supporting the surveillance and must describe the object of surveillance, the identity of the target, the offense being prosecuted, the kind of surveillance proposed, and the date and time of the beginning and end of the surveillance. *Ordonnance sur la surveillance de la correspondence par poste et telecommunication* [Ordinance on the Surveillance of Post and Telecommunications] Arts. 11, 15, 23 (Oct. 31, 2001), RS 780.11; HANSJAKOB, *supra* note 132, at 403–08, 412–24, 443–49.

253. In practice, police officers first recommend that surveillance be undertaken to the public prosecutor, who then makes a written order. Instead of the police, the Post and Telecommunications Surveillance Service ("PTSS") mainly coordinates and transmits the surveillance order from the public prosecutor to the pertinent service providers.

254. Surveillance orders are generally granted for up to three months, though the court may also impose its own requirements.

255. The Compulsory Measures Court may consent to notice being postponed or omitted. In the case of physical observation, the prosecutor may consent to notice being postponed or omitted. If notice is not given, however, the results of surveillance may not be used. *See supra* text accompanying notes 213–15.

256. CP art. 179ss.

evidence obtained from unauthorized surveillance[257] or obtained without their notice of surveillance excluded from trial under the exclusionary rule.[258]

2. In the United States

a) Several Distinctions

For real-time surveillance like that covered by the above category, laws in the United States distinguish between acquisition of the contents of communications made by mail, communications made by wire, and electronic communications. Unlike in Switzerland, ECPA treats the acquisition of electronic communications in electronic storage as less deserving of protection than real-time acquisition and subjects the former to a set of weaker restrictions.[259] Commentators have criticized ECPA for incorporating many distinctions that no longer make sense, if they ever did, and that make the law unduly complex.[260]

As in Switzerland, United States law treats the acquisition of documents and communications directly from a person's home or computer as a search or seizure. Such acquisitions are subject to a standard Fourth Amendment warrant requirement in most cases. The discussion that follows will focus on acquisitions from third parties, which, as in Switzerland, Congress has treated as a form of surveillance.[261]

b) Interception of Postal Mail Contents

First class mail and sealed packages in the United States have long been protected against warrantless interception.[262] To acquire mail and packages,

257. *See supra* Section VI.D.

258. CRIMPC art. 279, para. 2 lit a. Documents and data storage devices must be destroyed immediately and intercepted mail should be delivered.

259. *See supra* Section III.B (discussing the origins of these distinctions in Supreme Court cases from the 1970s).

260. *See, e.g.*, Ohm, *supra* note 4, at 1551 ("First, ECPA is confusing; epically confusing; grand-champion-of-the-U.S. Code confusing . . . . ECPA's complexities confuse judges who then make a mess of our understanding of the Act."); Dempsey, *supra* note 192, at 704–05, 722 (criticizing the complexity of the online surveillance rules and recommending a warrant standard for all stored email).

261. *See* COMPUTER CRIME & INTELLECTUAL PROP. SECTION, DEP'T OF JUSTICE, SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS 126 (3d ed. 2009) [hereinafter CCIPS SEARCH MANUAL], *available at* www.justice.gov/criminal/cybercrime/docs/ssmanual2009.pdf (explaining that ECPA does not apply to emails that "are not stored on the server of a third-party provider" of services).

262. *See* ROBERT ELLIS SMITH, BEN FRANKLIN'S WEB SITE: PRIVACY AND CURIOSITY FROM PLYMOUTH ROCK TO THE INTERNET 49–71 (2000) (reviewing history of protection of

agents must establish probable cause to a judge and also deliver notice to the target of the surveillance.[263] Because of the Fourth Amendment regulation, victims of unlawful acquisition of these items have a suppression remedy available to them.[264] In addition, a federal statute makes tampering with mail a criminal offense.[265] No statute provides other remedies for victims of unlawful mail surveillance, however.

c) Interception of Wire Communications Content

Wiretapping, or the real-time interception of the contents of wire communications,[266] is subject to the highest procedural restrictions, which in the United States are in the Wiretap Act.[267] Under the Act, a member of the judiciary oversees all phases of law enforcement surveillance. Applications for approval, which only high level officials can make,[268] must persuade the reviewing judge of probable cause to believe the target has committed or will commit a particular predicate offense and that the surveillance will obtain incriminating communications about that offense.[269]

The Wiretap Act provides for a U.S. version of subsidiarity, under which the reviewing judge must be convinced that the information sought may not be obtained by normal investigative methods and agents must minimize the interception of non-incriminating communications.[270] Surveillance orders are limited to thirty days, unless renewed, and the wiretapping must end when the information sought is obtained.[271] Together, these attempts to limit the scope and duration of wiretapping parallel the Swiss proportionality principle, although the Wiretap Act does not provide for the explicit

mail); Daniel J. Solove, *Conceptualizing Privacy*, 90 CALIF. L. REV. 1087, 1142–43 (2002) (same).

263. *See* United States v. Jacobsen, 466 U.S. 109, 114 (1984) (describing warrantless searches of sealed packages and letters as "presumptively unreasonable"); *Ex parte* Jackson, 96 U.S. 727, 733 (1877). The warrant requirement does not protect fourth class mail and the information visible on the outside of envelopes. WAYNE R. LAFAVE ET AL., CRIMINAL PROCEDURE § 4.2(a) (3d. ed. 2007).

264. *See* United States v. Villarreal, 963 F.2d 770 (5th Cir. 1992).

265. 18 U.S.C. § 1703 (2012).

266. *See* 18 U.S.C. § 2510(1) (2012) (defining "wire communication").

267. For an overview of the Wiretap Act requirements, see *In re* Sealed Case, 310 F.3d 717, 739–40 (FISC Ct. Rev. 2002).

268. 18 U.S.C. § 2516(1), (2) (2012).

269. 18 U.S.C. §§ 2516(1), 2518(3), (8) (2012). As in Switzerland, applications under the Wiretap Act require detailed information about facts and circumstances that support the request for an order. 18 U.S.C. § 2518(1).

270. 18 U.S.C. § 2518(3)(c).

271. 18 U.S.C. § 2518(5).

balancing incorporated into that principle.[272] As subsequent sections will show, most of the other modern surveillance practices in the United States proceed without consideration of the principles of proportionality or subsidiarity.

The Wiretap Act incorporates significant provisions to ensure transparency. The reviewing judge must provide notice to anyone named in an application and to anyone else the judge deems appropriate.[273] When Congress passed the Wiretap Act, it viewed the notice provision, in combination with civil remedies, as an important check on unlawful practices in that the community would be alerted if wiretaps were not reasonably employed.[274] In addition, Congress provided for detailed reports on the numbers of orders issued under the Wiretap Act and their efficacy in fighting crime.[275] Based on those reports, the Administrative Office of the United States Courts is supposed to make public a Report on Wiretapping each year.[276]

Courts may punish violations of the Wiretap Act with significant fines and jail time.[277] In addition, any person whose communications were intercepted, disclosed, or used in violation of the Act may bring civil claims for damages against those who violated their rights.[278] Under the Wiretap Act, a victim may receive attorney's fees, punitive damages, and actual or statutory damages.[279] The Wiretap Act provides a statutory suppression remedy to victims, which provides a complete exclusionary remedy.[280]

Between the significant procedural hurdles imposed on wiretap surveillance, the high level of judicial oversight, and the severe consequences for illegal investigations, the Wiretap Act sets the high water mark for restrictions on surveillance in the United States. Judicially-guaranteed notice to the target and the transparency of the public and congressional reports encourage victims to exercise their rights and obtain their remedies.

d) Interception of Electronic Communications Content

ECPA regulates the interception of modern communications such as email and cell phone calls the same way it regulates traditional wiretaps with a

272. *See supra* text accompanying notes 204–05.
273. *See* 18 U.S.C. § 2518(8)(d), (9).
274. *See* S. REP. NO. 90-1097, at 105 (1968), *reprinted in* 1968 U.S.C.C.A.N 2112, 2194.
275. *See* 18 U.S.C. § 2519 (2012).
276. *See* 18 U.S.C. § 2519(3); Soghoian, *supra* note 237, at 5.
277. 18 U.S.C. § 2511(4) (2012).
278. *See* 18 U.S.C. § 2520(a) (2012).
279. *See* 18 U.S.C. § 2520.
280. *See* 18 U.S.C. § 2515 (2012).

few significant differences.[281] The most significant difference is that when ECPA extended the Wiretap Act's provisions from "wire communications" to "electronic communications,"[282] it excluded the statutory suppression remedy.[283] Victims of unlawful interceptions of their electronic communications can have evidence obtained thereby excluded from trial only if they succeed in showing a Fourth Amendment violation.[284] The lack of a suppression remedy no doubt reduces the number of cases brought to vindicate rights under ECPA, even when the rights and remedies are otherwise at their height, as they are with the interception of electronic communications contents.[285]

All of the restrictions described above regarding judicial oversight, procedural hurdles, the last resort method, minimization, notice, and time limits apply to the interception of electronic communications, as do the civil remedies, criminal penalties, and reporting requirements. Agents may use electronic communications interceptions for only some crimes[286] and must get executive branch approval before doing so.[287] Government litigators have convinced courts to interpret "intercepts" to mean "acquisitions contemporaneous with transmission" and therefore to exclude the

281. Congress has expressed as its goal in crafting ECPA ensuring the privacy of electronic communications and extending all of the Wiretap Act's protections to new communications media. *See* H.R. REP. NO. 99-647, at 17–19 (1986); S. REP. NO. 99-541, at 25 (1986), *reprinted in* 1986 U.S.C.C.A.N. 3555, 3559.

282. 18 U.S.C. § 2510(12) (2012) (defining "electronic communication").

283. The Senate report reveals that the omission of the statutory suppression remedy was the "result of discussions with the Justice Department." S. REP. NO. 99-541, at 23 (1986), *reprinted in* 1986 U.S.C.C.A.N. 3555, 3577; *see also* Michael S. Leib, *E-Mail and the Wiretap Laws: Why Congress Should Add Electronic Communication to Title III's Statutory Exclusionary Rule and Expressly Reject a "Good Faith" Exception*, 34 HARV. J. ON LEGIS. 393, 409–11 (1997) (describing Justice Department opposition to the suppression remedy and congressional acquiescence due to the need for its support).

284. 18 U.S.C. §§ 2515, 2518(10) (2012); *see* Steve Jackson Games, 36 F.3d 457, 461 n.6 (5th Cir. 1994) (discussing statute and legislative history); *see infra* Section VII.B.2.e (describing United States v. Warshak, 631 F.3d 266, 282 (6th Cir. 2010), which held that an unlawful acquisition of stored email, rather than an interception, violated the Fourth Amendment).

285. *See supra* note 166.

286. 18 U.S.C. § 2516(3) (2012) (providing that electronic communications interceptions may be used in pursuit of any federal felony).

287. The Justice Department has required high level approval as a matter of its own policies. CCIPS SEARCH MANUAL, *supra* note 261, at 167. But ECPA permits any "attorney for the government" to authorize the interception of electronic communications. 18 U.S.C. § 2516(3).

acquisitions of electronic communications out of electronic storage.[288] Because of that narrowed scope, very few cases have been brought under the interception provisions.[289] Agents who choose to wait and acquire electronic communications that have come to rest instead of in real time may comply with the much weaker provisions of the Stored Communications Act ("SCA"),[290] which the next Section describes.

e) Acquisition of Stored Electronic Communications Content

The SCA, which applies when law enforcement agents obtain email and related electronic information stored with third party providers of "electronic communications service[s]" and "remote computing service[s],"[291] is much less restrictive than either the Wiretap Act or CrimPC. The SCA places no limits on who may conduct stored content acquisitions, which may be used to pursue any "ongoing criminal investigation," rather than just felonies or serious crimes.[292] Stored contents do not need to be acquired as a last resort, nor do agents need to minimize non-incriminating stored communications. The SCA places no time limits on stored content acquisitions, which allows investigators to ask for emails received over a span of years.[293] As with the remaining surveillance methods this Article describes, the SCA does not require public reporting on law enforcement's acquisition of stored contents.[294]

The remedies for illegal surveillance are less generous under the SCA provision for acquisition of stored email than they are for the interception of email. The SCA provides for civil damages in some cases, but it does not provide for punitive damages or criminal penalties against law enforcement

288. *See, e.g.*, Konop v. Hawaiian Airlines, 302 F.3d 868, 878 (9th Cir. 2002), *cert. denied*, 537 U.S. 1193 (2003); Steve Jackson Games, Inc. v. U.S. Secret Serv., 36 F.3d 457, 460–63 (5th Cir. 1994).

289. *See, e.g.*, United States v. Councilman, 418 F.3d 67 (1st Cir. 2005) (en banc) (concluding that email may be "intercepted" when it is acquired out of "transient electronic storage that is intrinsic to the communication process").

290. *See* Soghoian, *supra* note 237, at 10 (pointing out that since 1997, federal authorities had obtained only sixty-seven orders to intercept "computer[s] or email (electronic)" reflecting that "law enforcement agencies rarely engage in real-time interception of Internet communications . . . . [I]t is often easier and cheaper for them to do it after the fact rather than in real-time").

291. *See* 18 U.S.C. § 2703(a), (b) (2012).

292. *See* 18 U.S.C. § 2703(d).

293. *See, e.g.*, United States v. Warshak, 631 F.3d 266, 282 (6th Cir. 2010) (government compelled the disclosure of over 27,000 emails); Bellia & Freiwald, *supra* note 108, at 572 (noting Warshak's claim that some of his emails were nine years old).

294. The Attorney General must report to Congress on disclosures that service providers made on a voluntary basis only. *See* 18 U.S.C. § 2702(d) (2012).

officials who violate its provisions.[295] The SCA also provides no statutory suppression remedy, so unless victims of unlawful surveillance have a Fourth Amendment claim, they may not have unlawfully acquired stored contents information suppressed. In late 2010 in *United States v. Warshak,*[296] the Sixth Circuit found a warrantless acquisition of stored email to violate the Fourth Amendment,[297] and became the first federal appellate court to recognize a Fourth Amendment interest in stored email.[298] Until other federal circuits follow suit or Congress amends ECPA to provide a statutory suppression remedy,[299] victims of unlawful stored content acquisitions outside the Sixth Circuit will continue to lack a suppression remedy.

The provisions described above are common to all investigations proceeding under the SCA. But the SCA provides different procedural hurdles, levels of oversight, and rules on notice based on different features of the stored content. The next Sections describe those different rules.[300] If other courts follow *Warshak* and require a warrant, and certainly if Congress amends ECPA to do so as well, then the protections for stored email contents will be more comprehensive and less complex, which will bring them closer to those found in CrimPC.

i) Subject to the Warrant Requirement

Targets of law enforcement investigations that acquire the contents of email in "electronic storage" for 180 days or less benefit from the highest procedural hurdle and greatest oversight—a warrant based on probable cause that a reviewing judge must issue.[301] The 180-day cutoff for the mandatory warrant reflects Congress' view in 1986 that emails stored a relatively short time were likely protected by the Fourth Amendment,[302] while those stored longer than 180 days could be seen to be abandoned by the user and

295. *See* 18 U.S.C. §§ 2707(a)–(c), 2712 (2012). There is the possibility of administrative discipline for willful violations. *Id.* § 2707(d). The SCA provides immunity for private parties who act in good faith. *Id.* § 2707(e).

296. *Warshak*, 631 F.3d 266.

297. *Id.* at 283–88.

298. The court did not grant Warshak a suppression remedy because it found that the officers in his case relied in good faith on the terms of the SCA. *Id.* at 288–92.

299. The current version of the Electronic Communications Amendment Act of 2013, S. 607, would not add a statutory suppression remedy for the unlawful acquisition of stored emails. *See* Electronic Communications Privacy Act Amendments Act of 2013, S. 607, 113th Cong. (2013).

300. The reader will no doubt find the distinctions to be confusing and hard to follow. Table 1, *infra* Appendix, summarizes the differences.

301. *See* 18 U.S.C. § 2703(a) (2012).

302. *See* H.R. REP. NO. 99-647, at 67–68 (1986) (reporting that email in storage less than 180 days as likely protected by the Fourth Amendment).

therefore the business records of the storing company.[303] The Justice Department, whose agents apply for orders under the SCA every day, interprets the statutory language to mean only unopened (unretrieved) emails are entitled to the protection of a warrant requirement, no matter how long they have been stored, because only those emails are in "electronic storage" under the statute.[304] The Ninth Circuit has not accepted the Justice Department's approach, and applies the warrant requirement to all emails stored 180 days or less.[305] In the other jurisdictions, however, the Justice Department accords opened or retrieved emails lesser protections than a warrant, the specific protections depending on the type of server upon which the emails are stored.

Although federal criminal law generally requires notice to the target when a warrant is required,[306] the Justice Department argues that when it is authorized to use a warrant under the SCA it does not have to provide notice.[307] Without notice, of course, targets may never learn of the surveillance or that they have any rights with regard to it. If, as the *Warshak* court held, use of a warrant is constitutionally mandated, it may be that notice is mandated as well. In *Warshak*, however, agents unlawfully delayed providing notice for over a year, and the Sixth Circuit made no definitive statement that the Constitution requires notice.[308]

ii) Subject to a Lesser Standard

The SCA makes it significantly easier to acquire electronic communications contents that have been stored more than 180 days. Law

303. *See also id.* at 23 n.41 (analogizing emails held in long term storage to business records). As practices have changed and many users store their more important emails with their service providers for years, it makes no sense to protect older emails less.

304. *See* CCIPS SEARCH MANUAL, *supra* note 261, at 123–26, 138.

305. Ohm, *supra* note 4, at 1539 (citing Theofel v. Farey Jones, 359 F.3d 1066 (9th Cir. 2004)) (describing the 9th Circuit's rejection of the DOJ's approach and its requirement of a warrant for access to stored email).

306. *See* Smith, *supra* note 91, at 611 n.51 (citing Fed. R. Crim. Pro. 41(f)(1)(C), (f)(3) and noting that traditional search warrants provide notice to the targets while electronic surveillance orders do not); *see also* City of West Covina v. Perkins, 525 U.S. 234, 240 (1999) ("[W]hen law enforcement agents seize property pursuant to a warrant, due process requires them to take reasonable steps to give notice that the property has been taken so the owner can pursue available remedies for its return.").

307. CCIPS SEARCH MANUAL, *supra* note 261, at 133–34. Without any explanation or elaboration, the CCIPS manual asserts that the "search warrant obviates the need to give notice to the subscriber." *See id.* at 134 (citing 18 U.S.C. § 2703(b)(1)(A) (2012)). The Supreme Court has found notice constitutionally required for traditional electronic surveillance like wiretapping and bugging. *See* Berger v. New York, 388 U.S. 41, 73 (1967).

308. United States v. Warshak, 631 F.3d 266, 289 (6th Cir. 2010).

enforcement agents may apply for a special court order, known as a "D order," that a court may issue when the application "offers specific and articulable facts showing that there are reasonable grounds to believe that the . . . information sought [is] relevant and material to an ongoing criminal investigation."[309] When agents acquire stored email contents with a D order, they must give notice to the target, but may delay such notice.[310] In fact, the sample D order in the Justice Department's manual provides for delayed notice until such time as the court determines.[311] Instead of obtaining a D order, agents may obtain the available stored email content without a warrant using an administrative, trial, or grand jury subpoena, so long as they provide notice.[312]

As mentioned above, the Justice Department considers retrieved emails, or those opened, accessed, or read, as subject to the D order standard rather than the warrant requirement, even when they are stored for 180 days or less.[313] According to the DOJ, when emails are stored with a service provider that furnishes email services to the public, that provider is a statutory "remote computing service," and agents may acquire the already-retrieved emails from it pursuant to the lesser statutory standard.[314] If the service provider that stores the email does not furnish email to the public, for example if it is a University or corporate provider, the Justice Department considers the retrieved email to be entirely unprotected by the SCA, as discussed next.[315]

---

309. 18 U.S.C. § 2703(d).

310. *See* 18 U.S.C. § 2705(a)(2)(A)–(E) (2012) (listing reasons that justify the order, such as a concern that evidence will be destroyed or tampered with, the investigation will be jeopardized, or the trial delayed). Apparently agents do not always comply with the requirement that they eventually give notice. *See, e.g.*, *Warshak*, 631 F.3d at 289 (finding that law enforcement delayed giving notice of stored email acquisition for over a year despite only having approval to delay giving notice for ninety days).

311. *See* CCIPS SEARCH MANUAL, *supra* note 261, at 213–23 (App. B and attachment); *cf.* Smith, *Kudzu in the Courthouse*, *supra* note 220, at 208–12 (noting that many electronic surveillance orders remained under seal indefinitely).

312. 18 U.S.C. § 2703(b)(B).

313. *See* Freiwald, *supra* note 14, at 57–59 (criticizing the DOJ's approach).

314. *See* CCIPS SEARCH MANUAL, *supra* note 261, at 127 ("[A] single provider can simultaneously provide ECS [electronic communication services] with regard to some communications and RCS [remote computing services] with regard to others, or ECS with regard to some communications and neither ECS nor RCS with regard to others."). Orin Kerr has praised Congress' foresight in devising ECPA. *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, 1243 (2004) ("It is a particularly remarkable achievement given that its enactment dates back to 1986. The SCA has weathered intervening technological advances surprisingly well.").

315. *See* CCIPS SEARCH MANUAL, *supra* note 261, at 126 (describing how the "SCA no longer regulates access" to an email retrieved from a company provider of email). The

iii) Not Covered by the SCA

The DOJ argues that the SCA does not cover the acquisition of already-retrieved email from a non-public provider.[316] According to the DOJ, agents may compel the disclosure of information that falls outside of the SCA with a simple subpoena without any judicial oversight.[317] Recall that the subpoena can generally be challenged only on the basis that it seeks irrelevant or overbroad information.[318] The process is subject to no statutory restrictions, provides no remedies for unlawful investigations, and proceeds without notice to the subject.[319] Because such "surveillance" is covered and protected under CrimPC, a great disparity exists between U.S. and Swiss surveillance law.

C. ACQUISITION OF USER IDENTIFICATION DATA

1. In Switzerland

User identification data includes information related to communications ("communication attributes") but not the contents themselves. Such data also contains information about the location of the target and when and with which people the target is or was communicating by way of post or telecommunications.[320] Additionally, it includes billing data and traffic data, such as information about the duration of a call, the amount of data downloaded, and the like.[321] CrimPC treats tracking or locating someone using cell site location data as the acquisition of user identification data.[322]

Subject to two exceptions, CrimPC regulates the acquisition of user identification data under the same comprehensive and restrictive standards

Justice Department contends that public systems users qualify for more protection than non-public system users because they are less likely to have a personal relationship with their service providers. *See id.* at 135–36.

316. *See id.* at 125–26, 138.

317. *See id.* at 128 (describing the process for using a subpoena to obtain information beyond the scope of the SCA's protections).

318. *See* Slobogin, *supra* note 192, at 806 (identifying privilege, burdensomeness, and irrelevance as possible grounds for challenging the issuance of a subpoena generally and explaining that those challenges usually prove unavailing); *see also supra* note 203 (discussing ways for recipients to challenge subpoenas).

319. *See also* United States v. Scarfo, 180 F. Supp. 2d 572, 581–83 (D. N.J. 2001) (electronic monitoring by law enforcement that recorded keystrokes as they were typed but purportedly did not operate while the modem was "activated" was not subject to statutory regulation as a wiretap or electronic intercept).

320. CRIMPC art. 273, para. 1a.

321. CRIMPC art. 273, para. 1b.

322. It requires use of a telecommunications installation and involves the secrecy of telecommunications but no access to the contents of communications. *See* TF, Nov. 3, 2011, 132 ATF IV 340 (Switz.).

that apply to the surveillance of post and telecommunications. First, law enforcement agents may acquire user identification data for the investigation of any felony or misdemeanor, but they may only use the surveillance of mail and telecommunications to investigate a limited list of offenses.[323] Second, when judges apply the proportionality principle, they consider the acquisition of non-content user identification information to be less intrusive than interception of the contents of mail, email, and calls.[324]

Unlike ECPA and just as with the interception of post and telecommunications, CrimPC accords the same treatment to acquisition of user identification data in real time as it does to acquisition out of storage.[325] That uniformity of treatment substantially simplifies Swiss law relative to the United States. Agents may request historical user identification data up to six months after the data has been generated and a data retention requirement ensures that mail, telecommunications, and internet service providers will make such data available to them.[326]

As mentioned, the same comprehensive and highly protective procedures that apply to surveillance of post and telecommunications regulate the acquisition of user identification data, with the two exceptions noted. The procedures include several provisions: significant judicial oversight, the principles of subsidiarity and proportionality, the notice requirement, criminal penalties, and the significant remedies of damages and exclusion. These protective provisions all work together to ensure that surveillance under this category will not be overused or abused.

2. In the United States

a) Several Distinctions

The last section introduced the different treatment American law accords to the contents of postal mail, telephone calls, and electronic mail. ECPA has

323. CRIMPC arts. 269, 273. Law enforcement can also acquire user identification data to investigate the misuse of a telecommunications installation, which is an offense less serious than a misdemeanor. *See* CP art. 179septies.

324. ATF 137 IV 340, para 5.5.

325. *See* Sträuli, *supra* note 134, at 98–99.

326. SPTA art. 12, para. 2, art. 15, para. 3. The constitutional courts of the Czech Republic, Germany, and Romania consider the systematic conservation of a log without suspicion as against the constitution. Evaluation Report on the Data Retention Directive (Directive 2006/24/EC), at 5–6, COM (2011), 225 final (Apr. 18, 2011), *available at* http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0225:FIN:EN:PDF; *see also* TF, Jan. 8, 2010, docket no. 6B 766/2009, para. 3.4 (Switz.) (finding that data retention obligation applies to internet service providers). The obligation for service providers to keep logs of user identification data may be extended to twelve months. *See supra* note 73.

not only fallen out of date, but it retains a confusing set of categories that make understanding the applicable legal rules challenging at best. The next sections describe how U.S. law treats the surveillance that CrimPC handles under acquisition of user identification data. Table 2, *infra* Appendix, summarizes the differences.

b) Collection of Postal Mail Attributes

Legislating against the backdrop of the Fourth Amendment, Congress has provided few procedural restrictions on the surveillance of envelope information.[327] U.S. courts have historically distinguished between the contents of a letter that are unreadable until the envelope carrying the letter is opened, and information appearing on the outside of the envelope and therefore observable to postal workers when they process mail.[328] Courts have reasoned that senders of mail can have no reasonable expectation of privacy in information on the outside of envelopes that third party carriers can see.[329]

Under a 1975 Postal Service regulation, law enforcement agents can request that the post office retain "mail cover" information, or information obtained from the outside of postal mail, whenever they "specif[y] . . . reasonable grounds to demonstrate [that] the mail cover is necessary to . . . [o]btain information regarding the commission or attempted commission of a crime."[330] No judge provides oversight of the investigation, no notice needs to be provided, and no remedies are afforded to victims of improper investigations.[331]

c) Collection of Electronic Communication Attributes in Real Time

ECPA's provisions pertaining to pen registers and trap and trace devices provide minimal procedural restrictions comparable to those just described. Modern pen registers acquire the "dialing, routing, addressing and signaling information"[332] associated with wire and electronic communications as well as the date, time, and duration of transmissions, and information in the "cc"

327. Kerr, *supra* note 14, at 631.

328. United States v. Forrester, 512 F.3d 500, 511 (9th Cir. 2008) (describing line of cases finding a constitutional difference between contents and the information on the outside of mail).

329. *See* United States v. Van Leeuwen, 397 U.S. 249, 250–52 (1970); United States v. Hernandez, 313 F.3d 1206, 1209–10 (9th Cir. 2002).

330. 39 C.F.R. § 233.3(e)(2)(iii) (2012); Kerr, *supra* note 14, at 631.

331. Kerr, *supra* note 14, at 631.

332. *See* 18 U.S.C. § 3121(c) (2012).

and "bcc" fields of emails.[333] The Justice Department contends that any electronic communications information that is *not* the content of an electronic mail message or the subject line may be intercepted with a pen register order.[334] Courts have permitted law enforcement agents to acquire IP addresses with a pen register order, but have suggested that more specific URL information could not be acquired solely with a pen register order.[335]

Several courts and commentators have criticized the weak protections afforded by ECPA's pen register provisions.[336] Law enforcement agents who seek a pen register must apply for a special court order but do not need to establish probable cause. Instead, the investigating agent need only certify his belief "that the information likely to be obtained is relevant to an ongoing criminal investigation."[337] A judge asked to grant a pen register order "shall approve it" so long as she "finds that the application is complete."[338] Unlike CrimPC, the pen register provisions do not provide notice to the target or any remedies to the target for unlawful investigations; no statutory

333. *See* CCIPS SEARCH MANUAL, *supra* note 261, at 230 app. D. The Justice Department claims that any email header information may be acquired using a pen register. *See id.* at 154.

334. *See id.* at 154. The manual expresses ambivalence about whether the subject line is content or not by stating that it "*can* contain content." *Id.* at 152–53 (emphasis added). For a thorough discussion of the ambiguity here, see Freiwald, *supra* note 14, at 69–74 (arguing that there should be a third category of information that is neither content nor addressing information). For a different view, see Kerr, *supra* note 14, at 611–16 (arguing that there are only two categories); *see also* Orin S. Kerr, *Applying the Fourth Amendment to the Internet: A General Approach*, 62 STAN. L. REV. 1005, 1019–38 (2010) [hereinafter Kerr, *Applying the Fourth Amendment*] (developing claim that there are only two categories online: content and non-content information).

335. United States v. Forrester, 512 F.3d 500, 510–11 (9th Cir. 2008).

336. *See, e.g.*, Ohm, *supra* note 4, at 1550 ("Congress should amend the Pen Register Act to require at least reasonable suspicion" to "stamp out fishing expeditions"); Daniel J. Solove, *Reconstructing Electronic Surveillance Law*, 72 GEO. WASH. L. REV. 1264, 1289 (2004) (describing the Pen Register Act's protections as "limited and ineffective").

337. 18 U.S.C. § 3122(b) (2012).

338. 18 U.S.C. § 3123(a) (2012). Judges do not conduct independent reviews of the factual support for the applications, and the Justice Department has largely persuaded courts to view their role as "ministerial in nature." *See, e.g.*, United States v. Fregoso, 60 F.3d 1314, 1320 (8th Cir. 1995).

suppression remedy or damages are available.[339] Additionally, the statute does not provide for reports to Congress or the public.[340]

d) Collection of Electronic Communication Attributes from Electronic Storage

Congress afforded electronic communication attributes in electronic storage the lowest level of statutory protection. Law enforcement agents may compel the disclosure of a large set of information called "basic subscriber . . . information" from service providers by presenting an administrative, grand jury, or trial subpoena.[341] Under this provision, law enforcement agents may learn identifying information about a subscriber, including the electronic communication service to which he subscribes, when he used the service to access the Internet, and what IP address he used to do so.[342] In addition, service providers must turn over electronic records that disclose all of the people with whom a person has corresponded online and the "detailed internet address[es] of sites accessed."[343]

Although the size and duration of electronic log files vary by service provider, they can be quite revealing.[344] Service providers keep log files to protect themselves against hacking and fraud; such files can provide the entire history of one's communications and movements through the World Wide Web, down to an astonishing level of detail.[345]

339. Smith, *supra* note 91, at 612. Courts have found no Fourth Amendment right implicated by use of pen registers. *See, e.g.*, United States v. Forrester, 512 F.3d 500, 509–10 (9th Cir. 2008). The statute provides for the possibility of a criminal action against violators, but no known cases have been brought. *See* 18 U.S.C. § 3121(d) (2012) (providing for a penalty of a fine and up to one year of imprisonment).

340. 18 U.S.C. § 3123(a)(3)(A) provides for records to be kept when law enforcement agents use their own devices, but does not require that the reports be sent to Congress or published.

341. 18 U.S.C. § 2703(c)(2) (2012); CCIPS SEARCH MANUAL, *supra* note 261, at 128.

342. For example, the information comprises the subscriber's name, address, length of service, telephone number or IP address, and the means and source of payment. 18 U.S.C. § 2703(c)(2)–(3); *see also* USA PATRIOT Act § 210, 115 Stat. 272, 283 (2001) (adding "records of session times and durations" and "any temporarily assigned network address").

343. CCIPS SEARCH MANUAL, *supra* note 261, at 122.

344. *Id.* at 139 (noting that "some providers retain very complete records for a long period of time," while others retain few if any records). Bills have been proposed to impose a mandatory retention period for service provider logs. *See, e.g.*, Protecting Children from Internet Pornographers Act of 2011, H.R. Res. 1981, 112th Cong. (2011) (imposing obligation to hold identifying information for eighteen months).

345. The sample of a letter an agent may send to a provider to require the preservation of stored information under 18 U.S.C. § 2703(f) lists the following to preserve: all stored communications to and from the target, all files the target has accessed or controlled, all connections logs and records of user activity, including the volume of data transferred, all

Any other records "concern[ing]" electronic communications may be obtained with a D order,[346] but are subject to no other limits (such as subsidiarity or proportionality). Law enforcement agents are specifically excused from giving notice to targets under this section,[347] and are immune from criminal liability. Congress obtains no reports about acquisitions of electronic communications attributes from storage. Targets of unlawful surveillance may bring civil claims for improper investigations, but have no statutory suppression remedy.[348]

e) Cell Site Location Data Acquisition

The legal framework for acquisition of cell phone location data rivals the complexity attendant to acquisition of email. In addition, it is unclear how to apply ECPA rules to this method. Recall that interception of the content of cell phone calls and acquisition of the attributes of cell phone records other than location data are covered in the sections above.

Cell phone location data, however, which refers either to Global Position Satellite ("GPS") data associated with smartphone use or to records of the cell towers with which mobile phones communicate, reside in their own category. Courts have recognized that, while they do not fit under the traditional definition of communications content, such location records raise special concerns because they convey so much information about personal lives and activities. One magistrate judge recently explained that "[t]wo months' worth of hourly tracking data will inevitably reveal a rich slice of the user's life, activities, and associations . . . . If the telephone numbers dialed in *Smith v. Maryland* were notes on a musical scale, the location data sought here is a grand opera."[349] Cases have begun to reach the appellate courts raising

records of files or system attributes accessed, modified, or *added* by the user, and all connection information for other computers to which the user connected. It also includes all correspondence, and other records of contact by the target, the content and connection logs associated with or related to postings, communications or any other activities to or through the target's email or internet connections. *See* CCIPS SEARCH MANUAL, *supra* note 261, at 225–26; *see generally* DANIEL J. SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE (2004) (describing current online information gathering practices in depth).

346. 18 U.S.C. § 2703(c). There are some other limited ways in which government agents may acquire access to such records. *See id.*

347. 18 U.S.C. § 2703(c)(3).

348. 18 U.S.C. § 2707 (2012); *see also* Freedman v. Am. Online, Inc., 412 F. Supp. 2d 174, 181–83 (D. Conn. 2005) (no Fourth Amendment protection for subscriber information disclosed to the service provider's employees in the ordinary course of business).

349. *See, e.g.*, *In re* Application of the U.S. for Historical Cell Site Data, 747 F. Supp. 2d 827, 846 (S.D. Tex. 2010), *vacated*, 724 F.3d 600 (2013).

the issue of whether cell phone location data acquisition is protected by the Fourth Amendment, and if so, just what protections that affords.[350]

In the absence of clear guidance from either appellate courts or Congress, courts vary in the requirements they impose on law enforcement agents who compel disclosure of location data records from service providers. For acquisition of cell phone location data in real time, some courts require a warrant and others require the combination of a D order and a pen register order under what is called the "hybrid theory."[351] For the acquisition of location information out of electronic storage,[352] some courts have required a D order, and some have required a warrant. Because these cases have generally arisen before trial, when the government has requested records as part of its investigation, it is too early to say whether those courts that require a warrant will also require notice to the target and whether they will provide a suppression remedy to those subject to unlawful surveillance.[353] There is currently no reporting of cell phone data acquisitions and no statutory remedies other than civil remedies (but not notice) under the SCA when courts require a D Order.

D. TECHNICAL SURVEILLANCE EQUIPMENT

1. In Switzerland

CrimPC treats the use of technical surveillance devices as sufficiently invasive to be included in the most restricted category and accorded the same comprehensive treatment as the surveillance of mail and telecommunications. Technical surveillance equipment (sometimes called "other surveillance

350. *See* Freiwald, *Cell Phone Location Data*, *supra* note 90, at 732–49 (reviewing a 2010 Third Circuit case in detail and arguing that courts should impose Wiretap Act requirements on acquisition of cell site location data that covers a period of time); Government Brief 5th Circuit, *supra* note 98 (appealing district court case that affirmed Magistrate Judge Smith's opinion cited *supra* note 349).

351. *See, e.g.*, Steven B. Toenisketter, *Preventing a Modern Panopticon: Law Enforcement Acquisition of Real-Time Cellular Tracking Data*, 13 RICH. J.L. & TECH. 19–28 (2007) (describing cases accepting and rejecting the "hybrid theory").

352. In some cases the government purports to seek information out of electronic storage, but actually requests that information be created on an ongoing basis. *See* Susan Freiwald, *The Vanishing Distinction Between Real-time and Historical Location Data*, CONCURRING OPINIONS, (July 17, 2012, 4:50 PM), www.concurringopinions.com/archives/2012/07/the-vanishing-distinction-between-real-time-and-historical-location-data.html (describing how, in a case on appeal to the Fifth Circuit, agents asked for cell site location records to be created in real time and then stored, and then immediately transmitted to law enforcement agents as soon as they were stored).

353. *See, e.g.*, United States v. Muniz, No. H-12-221, 2013 WL 391161 (S.D. Tex. Jan. 29, 2013) (denying motion to suppress to defendant whose historical cell site location records were acquired without a warrant based on good faith rule).

measures") includes listening or audio recording devices, cameras, movie cameras, tracking devices,[354] and the like.[355] Law enforcement agents conduct surveillance using such devices when they observe or record statements or incidents made in non-public places and when they establish the location of people or things in both public and non-public places.[356] While there may appear to be some overlap among the surveillance categories, each technique belongs in only one category. For example, videotaping or photographing a telephone booth constitutes the use of technical surveillance equipment and not the monitoring of telecommunications when there is no access to the content of the phone call.[357] Audio and video recordings of places not accessible to the general public are covered under this category;[358] audio and video recordings in public spaces are not.[359]

As with the surveillance of post and telecommunications, only particular offenses justify the use of technical surveillance devices.[360] In addition, the same oversight, procedural hurdles, notice requirements, and consequences apply to unauthorized surveillance by technical surveillance equipment as apply to unauthorized surveillance by mail, email, and phone.[361]

2. In the United States

Reflecting the relative complexity of U.S. law, no single statute covers technical surveillance equipment. The closest approach to CrimPC in the United States would be the Wiretap Act, which strictly regulates the use of bugs and video surveillance in private areas. The Wiretap Act restricts the recording of spoken words in the same way as it restricts wiretapping, so long as the bugging takes place in an area in which the target has a reasonable expectation of privacy.[362] As described above, the Wiretap Act provides a

354. Including GPS devices and RFID.

355. The Technical Surveillance Equipment category may come to include later developed technologies that fit within its parameters. *See infra* Section VII.H.1.

356. Non-public places are places that are not accessible to the general public. CRIMPC art. 280. Before CrimPC, cantonal law varied a lot with respect to these practices. *See* GOLDSCHMID, *supra* note 196; Sträuli, *supra* note 134, at 112–17.

357. Thomas Hansjakob, Die ersten Erfahrungen mit dem Bundesgesetz über die Überwachung des Post- und Fernmeldeverkehrs [BÜPF], 120 REVUE PÉNALE SUISSE 268 (2002).

358. CRIMPC arts. 272, 281–282.

359. The recording of public spaces is treated as physical observation. CRIMPC art. 282; *see also infra* Section VII.G.1.

360. CRIMPC art. 281, para. 4.

361. *Id.*

362. 18 U.S.C. § 2510(2) (2012) (defining "oral communication" as "any oral communication uttered by a person exhibiting an expectation that such communication is not subject to interception under circumstances justifying such expectation").

comprehensive set of protections, such as approval of high level officials and extensive judicial oversight, subsidiarity and limited proportionality, transparency and notice, and significant remedies and a statutory suppression remedy.[363] Similarly, seven federal courts of appeals have found silent video surveillance, in areas subject to a reasonable expectation of privacy such as a home or office, to also require the highest restrictions of the Wiretap Act.[364] Because the restrictions derive by analogy from the Fourth Amendment rather than from the explicit text of Wiretap Act, however, the provisions for Congressional reporting and some of the other "technical" requirements do not apply to silent video surveillance.[365]

The Supreme Court has restricted similar surveillance methods using the Fourth Amendment. For example, it found law enforcement's use of a thermal imaging device to record the heat emanating from the target's home to be a search under the Fourth Amendment.[366] Though the *Kyllo* case was privacy-protective, its reasoning contains significant limits. The Court's emphasis on the fact that agents used devices not in general public use to search a home suggests that U.S. law would not restrict many of the techniques that CrimPC would.[367] It remains an open question how much the Court will restrict surveillance that does not implicate traditional property rights, especially when that surveillance uses readily available technology.

E. SURVEILLANCE OF CONTACTS WITH A BANK

1. In Switzerland

CrimPC includes surveillance of a target's contacts with a bank or bank-like institution in the most restricted category of surveillance, but relaxes protections by allowing bank surveillance to investigate any felony or misdemeanor, and by providing a slightly weaker exclusionary remedy.[368]

---

363. *See supra* Section VI.B.2.c.

364. *See supra* text accompanying notes 101–02.

365. United States v. Koyomejian, 970 F.2d 536, 542 (9th Cir. 1992) (en banc) (adopting the "last resort rule" for silent video surveillance as one of four Fourth Amendment requirements that also include minimization, particularity, and limited duration).

366. Kyllo v. United States, 533 U.S. 27, 27 (2001).

367. *Id.* at 40 (finding that the "Government use[d] a device that [was] not in general public use, to explore details of the home that would previously have been unknowable without physical intrusion").

368. CRIMPC art. 284; *see also* SYLVAIN MÉTILLE, MESURES TECHNIQUES DE SURVEILLANCE ET RESPECT DES DROITS FONDAMENTAUX EN PARTICULIER DANS LE CADRE DE L'INSTRUCTION PÉNALE ET DU RENSEIGNEMENT [SURVEILLANCE MEASURES AND FUNDAMENTAL RIGHTS, WITH PARTICULAR ATTENTION TO CRIMINAL AND INTELLIGENCE INVESTIGATIONS] 167–70 (2011). The bank itself primarily executes this type of surveillance by following the instructions contained in the surveillance order.

Surveillance of both financial flows and credit card information is available under this category, and authorized techniques include ordering the bank to transmit, in real time, information about every transaction with the bank; information from physical observation; information from communication intercepts; and specific documents relating to the accused person's interactions with a bank.[369] Because an order for real-time transmission of bank transactions requires a bank to transmit information that does not yet exist, it is forward looking.[370] Banks may also be ordered to provide access to their computer systems.[371]

As mentioned, besides the greater number of predicate offenses that can justify surveillance of bank contacts, CrimPC applies the same comprehensive restrictions accorded to surveillance of mail and telecommunications to surveillance under this category. Instead of a complete exclusionary remedy, however, CrimPC treats evidence uncovered by unauthorized surveillance of contacts with a bank as relatively unusable; it can be used only if the evidence could have been obtained legally and if it is necessary to solve serious offenses.[372] More serious committed offenses will increase the weight of the prosecution's interest in the information, tipping the balance against the private interest in not having the illegally obtained evidence used.[373]

369. SCHMID, *supra* note 205, at 538.

370. CrimPC's provision on Surveillance of Contacts with a Bank incorporates into Swiss law Article 4 of the Convention of the Council of Europe on Laundering, Search, Seizure and Confiscation of the Proceeds from Crime of November 8, 1990. Article 4 requires Swiss law to permit the use of special investigative techniques that facilitate the identification and tracking of proceeds and the gathering of evidence related thereto. *See* Conseil Fédéral, Message relatif à l'unification du droit de la procédure pénale [Message about Unification of Criminal Procedure Law], FF 1057, 1236 (2006); DANIEL JOSITSCH, GRUNDRISS DES SCHWEIZERISCHEN STRAFPROZESSRECHTS [OUTLINE OF SWISS CRIMINAL PROCEDURE LAW] 150 (2009);

371. Procedures for acquiring bank records are covered by the rules pertaining to searches and seizures. CPP arts. 241ss, 263ss; STEPHANIE EYMANN, DIE STRAFPROZESSUALE KONTOSPERRE [THE BANK ACCOUNT FREEZE ACCORDING TO CRIMINAL PROCEDURE] 81–90 (2009). However, Rhyner and Stüssi view surveillance of contacts with a bank as both occurring in real time and retroactively. Beat Rhyner & Dieter Stüssi, *Kommentar zu Art. 284–285 StPO*, *in* POLIZEILICHE ERMITTLUNG 484 (2008) (Gianfranco Albertini, et al. eds., 2008).

372. TF, Nov. 4, 1970, 96 ATF I 437, 441 (Switz.).

373. TF, May 3, 2005, 131 ATF I 272, 279 (Switz.). The police may conduct an undercover investigation to establish that the offense has been committed as well as to gather evidence of it.

2. In the United States

Undoubtedly because the United States does not share Switzerland's tradition of bank secrecy and because U.S. bank records are not subject to Fourth Amendment protection, no laws in the United States tailor law enforcement surveillance regulation specifically to the bank context.[374]

F. UNDERCOVER OPERATIONS

1. In Switzerland

CrimPC treats undercover operations as surveillance methods because they analogize the police hiding their official function to obtain evidence of committed offenses[375] to hiding devices like video cameras or wiretaps.[376] In undercover operations, police generally obtain fake identities to engage with suspects.[377] Because undercover operations intrude on privacy, CrimPC subjects them to the highest restrictions. CrimPC also restricts undercover operations to ensure that they are not used to entrap people; agents must restrict their activities to substantiating a preexisting intention to commit a criminal offense and they may not investigate outside of the context of a criminal investigation.[378]

Undercover investigations may be used to investigate a smaller number of serious offenses than surveillance of post and telecommunications or technical surveillance devices.[379] Except for that difference, CrimPC uses the same protective procedures for undercover investigations that it uses for the

374. The United States has a statute providing some secrecy for bank records, but it does not regulate the surveillance of bank contacts as CrimPC does. *See* Right to Financial Privacy Act, 12 U.S.C. §§ 3401–3422 (2012) (requiring a subpoena or warrant for the disclosure of financial information to the government).

375. TF, June 16, 2008, 134 ATF IV 266, 277, para. 3.7 (Switz.).

376. CRIMPC arts. 286–298; Vincent Jeanneret & Roland M. Ryser, *Commentaire ad art. 286-295 CPP* [*Commentary to articles 286–295 CrimPC*], *in* COMMENTAIRE ROMAND DU CODE DE PROCÉDURE PÉNALE [COMMENTARY TO CRIMINAL PROCEDURE CODE] 1315 (André Kuhn & Yvan Jeanneret, eds., 2011); Laurent Moreillon & Miriam Mazou, *Commentaire ad art. 296-298 CPP* [*Commentary to articles 296–298 CrimPC*], *in* COMMENTARY TO CRIMINAL PROCEDURE CODE 1351, *supra*.

377. In some situations a member of a foreign police force or a person temporarily appointed to carry out police work may be deployed as an undercover investigator.

378. If an undercover investigator oversteps the scope of the permissible action, then that shall be taken into consideration in determining the appropriate sentence to be imposed on the person concerned or the court shall refrain from sentencing the person altogether. CRIMPC art. 293, para. 4.

379. CRIMPC art. 286, para. 2 contains the second list pertaining to undercover investigations and contains a smaller number of offenses than the list in CRIMPC art. 269, para. 2 pertaining to post and telecommunications surveillance.

surveillance of post and telecommunications and all of the other methods discussed,[380] apart from the slight variations mentioned.

2. In the United States

In sharp contrast to the Swiss approach, use of undercover agents faces no regulation in the United States. No statute applies, and in a series of cases more than fifty years old, the Supreme Court found no Fourth Amendment search when agents used undercover agents to either record or transmit information divulged by a criminal suspect.[381] As a result, use of undercover agents requires no warrant or judicial oversight. If undercover agents engage in wiretapping or use another restricted surveillance method, however, those restrictions apply.[382]

The difference between the way Switzerland tightly controls undercover agents and the United States does not has tremendous implications for the two countries' systems. First, it illustrates that the Swiss employ a dignity-based approach in which the police do not misrepresent themselves to their people, which is clearly lacking in the United States. Second, the undercover agent rule's assumption of risk approach underlies the third party doctrine.[383] If courts or legislators see the weakness in the doctrine's underpinnings, they will have an easier time in granting more privacy rights in new communications technologies that rely on access to information stored by others.[384]

G. PHYSICAL OBSERVATION

1. In Switzerland

Under CrimPC, use of physical observation is a less invasive category of surveillance. While courts have not yet confirmed that surveillance by physical observation breaches privacy, scholars argue that it does,[385] at least if the observation persists. Accordingly, while CrimPC provides a legal basis for

380. CRIMPC arts. 274, 289.

381. *See* On Lee v. United States, 343 U.S. 747 (1952); United States v. White, 401 U.S. 745 (1971).

382. *See generally* LAFAVE ET AL., *supra* note 263, § 3.1(c); Ross, *supra* note 21, at 533–43.

383. *See* Bellia & Freiwald, *supra* note 108, at 153–56.

384. *See* United States v. Jones, 132 S. Ct. 945, 957 (2012) (Sotomayor, J., concurring) (describing the third party rule as "ill suited to the digital age, in which people reveal a great deal of information about themselves to third parties in the course of carrying out mundane tasks").

385. For the different opinions among commentators, see ROBERTO ZALUNARDO-WALSER, VERDECKTE KRIMINALPOLIZE ILICHE ERMITTLUNGSMASSNAHMEN UNTER BESONDERER BERÜCKSICHTIGUNG DER OBSERVATION [UNDERCOVER LAW ENFORCEMENT INVESTIGATION WITH PARTICULAR ATTENTION TO PHYSICAL OBSERVATION] 50 (1999).

physical observation, it may proceed under a set of procedural requirements that are easier to meet.[386]

Physical observation occurs when, in the course of an investigation, a member of the public prosecutor's office or the police covertly observes people and things in places accessible to the general public and makes audio or video recordings for criminal prosecution.[387] CrimPC regulates focused, systematic physical observation, as well as observation that takes place over time. Surveillance under this category, which does not have to be recorded, is limited to physical observation in public places; CrimPC provides more oversight for surveillance in private places, which constitutes the use of the technical surveillance equipment described in Section VII.D.1, *supra*.

The Swiss Supreme Court recently decided that following a chat in an online (public) forum and focusing on some participants constitutes observation. Just following the conversation without focusing on someone in particular does not constitute surveillance but is instead comparable to when an officer patrols the street. If the observation develops to the point that the officer takes part in a conversation without identifying himself as a police officer, then it will have become an undercover investigation and be subject to further restrictions.[388]

CrimPC permits the public prosecutor or police to authorize physical observation, rather than requiring independent judicial review.[389] It may proceed so long as there are concrete reasons to assume that crimes or offenses have been committed and may be used to investigate any felony or misdemeanor.[390] The procedural hurdle is lower than the strong suspicion required of the other surveillance methods, but higher than the standard of simple suspicion used to open investigations.[391] Similar to surveillance of contacts with a bank, CrimPC provides a modified rather than a complete exclusionary remedy for targets of unauthorized physical observation.[392] Notwithstanding the lower level of oversight, lower procedural hurdles, and

386. *See* Conseil Fédéral, Message relatif à l'unification du droit de la procédure pénale [Message about Unification of Criminal Procedure Law], FF 1057, 1235 (2006).

387. *See* CRIMPC art. 282, para. 1; Conseil Fédéral, Message about Unification of Criminal Procedure Law, FF 1057, 1235.

388. Observation occurs at a distance, while undercover investigation requires an officer designated for this purpose to infiltrate a given environment. TF, June 16, 2008, 134 ATF IV 266 (Switz.).

389. Physical observation that continues for longer than one month requires the authorization of the public prosecutor. CRIMPC art. 282, para. 2. CrimPC does not require that the authorization be in writing, but that is obviously recommended.

390. CRIMPC art. 282, para. 1a.

391. *See supra* note 199.

392. *See supra* Section VI.E.1.

modified exclusionary remedy, CrimPC still requires that those targeted by physical observation receive notice.[393]

2. In the United States

While CrimPC provides reduced regulation for surveillance in public, U.S. law has traditionally provided no regulation at all. The understanding has been that one has no privacy from government surveillance in public. As Christopher Slobogin has written, "[t]he advent of sophisticated technology that allows the government to watch, zoom in on, track, and record the activities of anyone, anywhere in public, twenty-four hours a day, demands regulation. Yet to date no meaningful constraints on this type of surveillance exist."[394] According to Orin Kerr, "[t]he distinction between government surveillance outside and government surveillance inside is probably the foundational distinction in Fourth Amendment law . . . . According to this distinction, the government does not need any cause or order to conduct surveillance outside."[395] Although some have criticized the notion that people assume the risk of unobserved surveillance when they venture outside,[396] courts have largely accepted it.

The Supreme Court's decision in *United States v. Jones*[397] may indicate a shift. The *Jones* case found the use of a specialized GPS device attached to a car to be a search under the Fourth Amendment, but the case has broader implications. The Court could have disposed of the defendant's constitutional claim on the ground that law enforcement agents observed him while he was outside. The Court's failure to do so paves the way for future cases to revisit the assumption that movements out of doors cannot be subject to Fourth Amendment protection.[398]

---

393. But the public prosecutor may decide to postpone or omit giving notice. Defendants may challenge the surveillance when they learn of it by submitting an objection to the decision of the public prosecutor or to a cantonal court. CRIMPC art. 393, para. 1a.

394. CHRISTOPHER SLOBOGIN, PRIVACY AT RISK: THE NEW GOVERNMENT SURVEILLANCE AND THE FOURTH AMENDMENT 79 (2007).

395. *See* Kerr, *Applying the Fourth Amendment*, *supra* note 334, at 1010 (citing cases); Philadelphia Yearly Meeting of the Religious Society of Friends v. Tate, 519 F.2d 1335 (3d Cir. 1975) (finding that no privacy right was violated by police observations of public meetings and activities).

396. *See, e.g.*, HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY AND THE INTEGRITY OF SOCIAL LIFE 113–26 (2011); SLOBOGIN, *supra* note 394, at 79–136.

397. United States v. Jones, 132 S. Ct. 945 (2012).

398. *See, e.g.*, Montana State Fund v. Simms, 2012 MT 22 (Mont. 2012) (Nelson, J., specially concurring) (asserting that "Montanans do retain expectations of privacy while in public" particularly in light of the Justice's statements in *Jones*), *available at* http://goo.gl/GSL1f.

H. NEW TECHNIQUES

1. In Switzerland

It seems likely that as new techniques are developed, Swiss law will consider them to be covered under rules pertaining to technical surveillance devices. Indeed the legislature drafted the Technical Surveillance Equipment category to cover techniques used to listen, record, observe, or locate, but those categories are considered illustrative rather than exhaustive.[399]

If a new surveillance technique appears to have fundamentally different means or goals, however, a specific new rule or amendment would be needed. The federal Constitution and the ECHR require that a law be clear and foreseeable as to its effects,[400] which prohibits interpreting CrimPC to permit surveillance techniques that could not have been imagined when the law was passed. A new rule would also be needed for any techniques that the legislature considered when drafting CrimPC and specifically decided not to cover.

When law enforcement agents want to use a new surveillance technique, they have to discern if the legislature deliberately excluded that technique from CrimPC, even without explicitly saying so. If so, the technique could be used only after CrimPC had been modified to address it. On the other hand, if the legislature merely forgot to mention a technique in the explanatory reports or hearings and if the technique fits a specific category of CrimPC by analogy, the technique may be usable.[401]

For example, surreptitious installation of a government monitoring software, though not mentioned explicitly in CrimPC, may be covered under the rules pertaining to Post and Telecommunications when it targets electronic communications content, the rules pertaining to User Identification Data when it targets communication attributes, and rules pertaining to Technical Surveillance Equipment when it is used to control a webcam or microphone.[402] However, the Federal Council decided a court

399. CRIMPC art. 280; Tribunal administratif fédéral [TAF] [Federal Administrative Court], June, 23, 2011, RECUEIL OFFICIEL DES ARRÊTS DU TRIBUNAL FÉDÉRAL ADMINISTRATIF SUISSE [ATAF] A-8267/2010, § 3.2.

400. *See supra* note 61.

401. *See* SYLVAIN MÉTILLE, MESURES TECHNIQUES DE SURVEILLANCE ET RESPECT DES DROITS FONDAMENTAUX EN PARTICULIER DANS LE CADRE DE L'INSTRUCTION PÉNALE ET DU RENSEIGNEMENT [SURVEILLANCE MEASURES AND FUNDAMENTAL RIGHTS, WITH PARTICULAR ATTENTION TO CRIMINAL AND INTELLIGENCE INVESTIGATIONS] 220–24 (2011).

402. *See* Sylvain Métille, *Les mesures de surveillance prévues par le CPP*, WEBLAW JUSLETTER (Dec. 19, 2011), *available at* http://jusletter.weblaw.ch/_645.

may not consider the legal basis for such use to be sufficiently clear and foreseeable and proposed that Parliament amend CrimPC to permit government monitoring software to monitor communications.[403] Any other use of government monitoring software (e.g., distant search and seizure, monitoring of the environment of the computer, etc.) is deemed illegal.[404]

Similarly, IMSI-Catchers, which mimic cell towers to acquire cell site location data,[405] have never been mentioned by courts or legislators, but they are sometimes used to intercept communications and communications attributes by using communications infrastructures.[406] As such, courts should treat IMSI-Catchers under the rules pertaining to Post and Telecommunications and Acquisition of User Identification Data when they collect electronic communications and their attributes.[407] For the sake of clarity and foreseeability of the law, the Federal Council also proposed that Parliament amend CrimPC to explicitly allow the use of IMSI-Cacthers.[408.]

2. In the United States

Because law in the United States generally provides negative rights (restrictions on government behavior) rather than positive rights (rules that must be in place to authorize government behavior), law enforcement agents have generally used new surveillance methods during the period before their treatment under existing statutes or the Fourth Amendment was clear.

---

403. *See* Conseil Fédéral, Message concernant la loi fédérale sur la surveillance de la correspondance par poste et télécommunication [LSCPT] [Message About the Modification of the Surveillance of Post and Telecommunications Act] FF 2013 2379, 2464–74 (2013), *available at* www.admin.ch/opc/fr/federal-gazette/2013/2379.pdf.

404. *Id.*

405. *See EPIC v. FBI—Stingray / Cell Site Simulator*, ELECTRONIC PRIVACY INFORMATION CENTER, http://epic.org/foia/fbi/stingray/ ("A StingRay is a device that can triangulate the source of a cellular signal by acting 'like a fake cell phone tower' and measuring the signal strength of an identified device from several locations. With StingRays and other similar 'cell site simulator' technologies, Government investigators and private individuals can locate, interfere with, and even intercept communications from cell phones and other wireless devices."); *see, e.g.*, United States v. Rigmaiden, No. CR08-0814, 2012 WL 1038817 (D. Ariz. Mar. 28, 2012) (involving the government's use of StingRay to locate defendant).

406. *See* Sophie de Saussure, *Le IMSI-Catcher: fonctions, applications pratiques et légalité*, WEBLAW JUSLETTER (Nov. 30, 2009), *available at* http://jusletter.weblaw.ch/_547.

407. New articles may be added to CrimPC to authorize the use of Government-Software (Trojans) and IMSI-Catchers and to extend to twelve months from six the obligation for service providers to keep logs of user identification data. *See* Conseil Fédéral, Message concernant la loi fédérale sur la surveillance de la correspondance par poste et télécommunication [LSCPT] [Message About the Modification of the Surveillance of Post and Telecommunications Act], FF 2379, 2393-4, 2397-8, 2426-7, 2436-7, 2464-72 (2013).

408. *Id.*

For example, some courts have found that acquisition of cell phone location data falls outside the scope of ECPA.[409] But if so, it remains unclear whether the technique is covered by the Fourth Amendment, and if not, whether there are any constraints at all upon the use of that method of surveillance.[410] As another example, some law enforcement agencies have started the widespread use of license plate readers to match captured data from parked cars with state databases of stolen vehicles and wanted criminals. Because no regulation currently addresses what can be done with the information or how long it can be retained, one privacy advocate complained, "the infrastructure to protect individuals' privacies and rights doesn't exist, particularly on the legislative and the judicial side."[411]

VIII. CONCLUSION

In the United States, traditional wiretapping (of wire, oral, and electronic communications) and some video surveillance is subject to most of the restrictions imposed by CrimPC in Switzerland: notice, a remedy, subsidiarity, and proportionality.[412] The rest of what CrimPC treats as surveillance is subject to significantly less protection. Law enforcement agents in the United States may use undercover agents, collect stored communications contents and attributes, intercept communication attributes in real time, track location data, and use other modern surveillance techniques subject either to no regulation at all or to the anemic protections afforded by ECPA and a few related statutes.[413]

CrimPC, which brought unity and comprehensive treatment to Swiss surveillance law, dramatically contrasts with the incomplete, confusing, and ineffective laws that regulate surveillance in the United States. It seems clear that the substantive requirements in both the European Convention on Human Rights and the Swiss constitution have yielded significantly stronger

409. *See In re* Application of the U.S. for an Order Directing a Provider of Elec. Commc"ns Servs. to Disclose Records to the Gov"t, 534 F. Supp. 2d. 585, 602 n.44 (W.D. Pa. 2008) (collecting cases), *aff"d*, No-524M, 2008 WL 4191511 (W.D. Pa. 2008), *vacated*, 620 F.3d 304 (3d Cir. 2010).

410. *See supra* Section VII.C.2.e.

411. Eric Roper, *Police Cameras Quietly Capture License Plates, Collect Data,* STAR TRIBUNE, Aug. 10, 2012, www.startribune.com/local/minneapolis/165680946.html.

412. The significant exception is that the unlawful interceptions of electronic communications are not subject to a statutory suppression remedy.

413. This Article has not covered a few minor surveillance statutes, such as the Video Privacy Protection Act. 18 U.S.C. § 2710 (2012), *amended by* Video Privacy Protection Act Amendments Act of 2012, 18 U.S.C.A. § 2710, Pub. L. No. 112-258, 126 Stat. 2414 (amended 2013).

restrictions on law enforcement surveillance. The limited coverage of the Fourth Amendment, and the fact that it exerts no real influence absent a ruling, shifts the default rule in the United States in favor of using new surveillance methods that the legislature has not yet regulated. The opposite rule applies in Switzerland, where techniques that CrimPC does not cover, either explicitly or by analogy, cannot be used. It would represent a significant and likely unattainable shift in our jurisprudence to prohibit law enforcement agents from using new surveillance techniques until Congress explicitly authorizes those techniques. It should be possible, however, for Congress to design a set of surveillance rules that abandon arbitrary distinctions, provide sufficient procedural hurdles and oversight to constrain invasive practices, furnish meaningful remedies to deter abuse, and provide notice and transparency to ensure that the system works as designed. In drafting such an overhaul, American legislators should look to CrimPC for guidance.

APPENDIX

Table 1.

Comparison of U.S. and Swiss Laws for Interception/Acquisition of Communications Content

	Notice Requirement		Suppression Remedy		Level of Judicial Review	
	Switz.	U.S.	Switz.	U.S.	Switz.	U.S.
Mail	Yes	Yes	Yes	Yes	Strong suspicion**	Probable cause
Wire and Phone Communications	Yes	Yes	Yes	Yes	Strong suspicion**	Probable cause with add'l requirements
Electronic Communications	Yes	Yes	Yes	No	Strong suspicion**	Probable cause
Communications Stored ≤ 180 days	Yes	*	Yes	No	Strong suspicion**	Probable cause
Communications Stored > 180 days	Yes	*	Yes	No	Strong suspicion**	Relevant and material to an ongoing investigation

* Notice requirement varies depending on the procedures used and where the data is stored.

** of any enumerated felony

Table 2.

Comparison of U.S. and Swiss Laws Regarding Acquisition of User Identification/Non-Content Data

	Notice Requirement		Suppression Remedy		Level of Judicial Review	
	Switz.	U.S.	Switz.	U.S.	Switz.	U.S.
Mail	Yes	No	Yes	No	Strong suspicion*	None
Real Time Interception of Electronic and Phone Data	Yes	No	Yes	No	Strong suspicion*	Relevant to an ongoing criminal investigation
Stored Electronic Data	Yes	No	Yes	No	Strong suspicion*	Relevant and material to an ongoing investigation
Cell Site Location Data	Yes	No	Yes	No	Strong suspicion*	Varies by Jurisdiction

* of any felony or misdemeanor

PRIVACY BY DESIGN: A COUNTERFACTUAL ANALYSIS OF GOOGLE AND FACEBOOK PRIVACY INCIDENTS

Ira S. Rubinstein[†] *& Nathaniel Good*[††]

ABSTRACT

Regulators here and abroad have embraced "privacy by design" as a critical element of their ongoing revision of current privacy laws. The underlying idea is to "build in" privacy—in the form of Fair Information Practices or ("FIPs")—when creating software products and services. But FIPs are not self-executing. Rather, privacy by design requires the translation of FIPs into engineering *and* usability principles and practices. The best way to ensure that software includes the broad goals of privacy as described in the FIPs and any related corporate privacy guidelines is by including them in the definition of software "requirements." And a main component of making a specification or requirement for software design is to make it concrete, specific, and preferably associated with a metric. Equally important is developing software interfaces and other visual elements that are focused around end-user goals, needs, wants, and constraints.

This Article offers the first comprehensive analysis of engineering and usability principles specifically relevant to privacy. Based on a review of the technical literature, it derives a small number of relevant principles and illustrates them by reference to ten recent privacy incidents involving Google and Facebook. Part I of this Article analyzes the prerequisites for undertaking a counterfactual analysis of these ten incidents. Part II presents a general review of the design principles relevant to privacy. Part III turns to ten case studies of Google and Facebook privacy incidents, relying on the principles identified in Part II to discover what went wrong and what the two companies might have done differently to avoid privacy violations and consumer harms. Part IV of the Article concludes by arguing that all ten privacy incidents might have been avoided by the application of the privacy engineering and usability principles identified herein. Further, we suggest that the main challenge to effective privacy by design is not the lack of design guidelines. Rather, it is that business concerns often compete with and overshadow privacy concerns. Hence the solution lies in providing firms with much clearer guidance about applicable design principles and how best to incorporate them into their software development processes. Regulators should provide greater guidance on how to balance privacy with business interests, along with appropriate oversight mechanisms.

---

† Adjunct Professor of Law and Senior Fellow, Information Law Institute, New York University School of Law.

†† Principal, Good Research LLC.

This Article was presented at the NYU Privacy Research Group and at the 2012 Privacy Law Scholars Conference, and we are grateful for the comments of workshop participants. Ron Lee, Paul Schwartz, and Tal Zarsky provided valuable suggestions on an earlier draft. Thanks are also due to Jeramie Scott and Mangesh Kulkarni for excellent research assistance and to Tim Huang for his help with citations. A grant from The Privacy Projects supported this work.

TABLE OF CONTENTS

I. BACKGROUND

Regulators have embraced privacy by design.[1] Both the European Commission ("EC") and the Federal Trade Commission ("FTC") have recently called for a new approach to data protection and consumer privacy in which privacy by design plays a key role.[2] However, the details of what this means in practice will remain unclear until the EC completes its work on the delegated acts and technical standards anticipated by the proposed Regulation,[3] or until the FTC refines the meaning of "unfair design" through enforcement actions[4] and/or develops guidelines based on its ongoing dialogue with private firms.[5] Indeed, despite the strong expressions of support for privacy by design, its meaning remains elusive.

Presumably, the regulatory faith in privacy by design reflects a commonsense belief that privacy would improve if firms "designed in" privacy at the beginning of any development process rather than "tacking it on" at the end. And yet there is scant relevant data in support of this view. A few firms have adopted privacy guidelines for developing products and services;[6] however, a search of the literature reveals no before-and-after studies designed to determine if such firms have achieved better privacy

1. *See* Ira S. Rubinstein, *Regulating Privacy by Design*, 26 BERKELEY TECH. L.J. 1409, 1410–11 (2012) (describing statements by regulators in Canada, Europe, and the United States).

2. *See Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)*, Recital 61, art. 23, COM (2012) 11 final (Jan. 25, 2012), *available at* http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf [hereinafter *Proposed E.U. Regulation*] (requiring data controllers to implement mechanisms ensuring "data protection by design and by default"); FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS (2012), http://www.ftc.gov/os/2012/03/120326 privacyreport.pdf [hereinafter FTC FINAL REPORT] (urging companies to "build in privacy at every stage of product development").

3. *Proposed E.U. Regulation*, *supra* note 2, art. 23(3)–(4).

4. *See, e.g.*, Complaint for Permanent Injunction and Other Equitable Relief at 13, 19, F.T.C. v. Frostwire LLC, No. 1:11-CV-23643, 2011 WL 9282853 (S.D. Fla. 2011) (describing default setting of Android application that allowed sharing of all existing files on the device in terms of "unfair design").

5. *See* Kenneth A. Bamberger & Deirdre K. Mulligan, *Privacy on the Books and on the Ground*, 63 STAN. L. REV. 247, 287–89 (2011) (describing various "deliberative and participatory processes promoting dialogue with advocates and industry").

6. *See The Role of Privacy by Design in Protecting Consumer Privacy*, CTR. FOR DEMOCRACY & TECH. (Jan. 28, 2010), https://www.cdt.org/policy/role-privacy-design-protecting-consumer-privacy [hereinafter *Role of Privacy by Design*] (explaining that IBM, Sun Microsystems, Hewlett-Packard, and Microsoft have adopted privacy by design into their business models and product development procedures).

results. We propose to examine this question in a different fashion—not by gathering empirical data but rather by conducting and reporting on case studies of ten major Google and Facebook privacy incidents.[7] We then consider whether the firms in question would have averted these incidents if they had implemented privacy by design.

This is a counterfactual analysis: we are asking a "what if?" question and will try to answer it by discussing what Google and Facebook might have done differently to better protect consumer privacy and thereby avoid these incidents. The proposed analysis has two prerequisites. First, we need ready access to a great deal of information about the selected incidents so that we have a reasonably clear idea of what happened as well as how and why the firms responded as they did (for example, by modifying certain features or even withdrawing a service entirely). Absent such information, it would be impossible to consider what the firm might have done differently if it had adopted privacy by design. Second, we need to identify a baseline set of *design* principles that will inform our discussion of alternative outcomes.

The first task is easy because there are so many well-documented major Internet privacy incidents. A non-exhaustive list would include privacy gaffes by AOL, Apple, DoubleClick, Facebook, General Motors, Google, Intel, Microsoft, MySpace, Real Networks, Sony, and Twitter.[8] This Article focuses on a series of related incidents—five each from Google and from Facebook—for several reasons. To begin with, both firms have experienced serious privacy incidents and suffered major setbacks ranging from negative publicity and customer indignation to government scrutiny, regulatory actions, and law suits. Second, their travails have been well documented by investigative journalists, privacy advocates, and various regulators. And, third, both firms have all of the necessary resources—engineering talent, financial wherewithal, and business incentives—to prevent future incidents by implementing a leading-edge program of privacy by design. Moreover, studying a range of incidents at each company—Gmail, Search, Street View, Buzz (and Google+), and changes in privacy policies for Google; and News Feed, Beacon, Facebook Apps, Photo Sharing, and changes in privacy

7. As used here, the term "incident" is descriptive rather than normative. Thus, a "privacy incident" is no more than an episode or event that raises privacy concerns. Not every privacy incident results from a design failure or causes harm. However, because privacy is highly cherished and causes anxiety if violated, many privacy incidents are associated with negative press coverage, reputational harm, regulatory investigations, and/or enforcement actions.

8. We identified these incidents based on general knowledge and by reviewing the websites of leading privacy organizations for discussion of privacy issues; we also conducted a LexisNexis® search.

policies and settings for Facebook—makes it possible to observe patterns and compare how the two companies think about privacy, especially in similar services such as social networking.[9]

The second task—identifying design principles to rely on for purposes of a counterfactual analysis—is far more difficult. An obvious starting point for understanding what it means to design products and services with privacy in mind is the set of internationally recognized values and standards about personal information known as the Fair Information Practices ("FIPs").[10] The FIPs define the rights of data subjects and the obligations of data controllers; most privacy laws throughout the world rely on FIPs.[11] This Article argues that although the FIPs allocate rights and responsibilities under applicable legal standards, the present task requires something different, namely, *design* principles and related practices.

Another possible source of guidance is the work of Ann Cavoukian, the Information and Privacy Commissioner ("IPC") of Ontario, Canada. Cavoukian is a tireless champion of privacy by design (or "PbD" to use her preferred acronym) and has authored or coauthored dozens of papers describing both its origins and its business and technology aspects.[12] In 2009, Cavoukian advanced the view that firms may accomplish privacy by design by practicing seven "foundational" principles:

1. Proactive not Reactive; Preventative not Remedial;
2. Privacy as the Default Setting;
3. Privacy Embedded into Design;
4. Full Functionality—Positive-Sum, not Zero-Sum;
5. End-to-End Security—Full Lifecycle Protection;
6. Visibility and Transparency—Keep it Open; and
7. Respect for User Privacy—Keep it User-Centric.[13]

9. *See infra* Part III.

10. The FIPs are a set of internationally recognized privacy principles that date back to the 1970s. They have helped shape not only the main U.S. privacy statutes but also European data protection law. *See infra* Section II.A; *see generally Fair Information Practice Principles*, FED. TRADE COMM'N, http://www.ftc.gov/reports/privacy3/fairinfo.shtm (last visited Mar. 15, 2013).

11. *See, e.g.*, Marc Rotenberg, *Fair Information Practices and the Architecture of Privacy (What Larry Doesn't Get)*, 2001 STAN. TECH. L. REV. 1, ¶ 44 (2001).

12. These publications are available on the IPC website. *Discussion Papers*, IPC, http://www.ipc.on.ca/english/Resources/Discussion-Papers (last visited Mar. 6, 2013).

13. ANN CAVOUKIAN, PRIVACY BY DESIGN: THE 7 FOUNDATIONAL PRINCIPLES (2011), www.privacybydesign.ca/content/uploads/2009/08/7foundationalprinciples.pdf.

Although Cavoukian's many publications offer valuable lessons in how the public and private sector might apply the "PbD approach" to new information systems and technologies, it is not at all clear for present purposes that her seven principles are of any greater assistance than the FIPs.

To begin with, Cavoukian's seven principles are more aspirational than practical or operational. Principles 1–3 provide useful, if somewhat repetitive, guidance about the importance of considering privacy issues early in the design process and setting defaults accordingly, but they stop far short of offering any design guidance. Granted, Cavoukian offers more practical advice in several of her technology-specific papers,[14] but she makes little effort to systematize or even summarize the design principles found therein.[15] Principle 4 seems unrealistic in an era when some view personal data as the "new oil" of the Internet and privacy controls only tend to limit the exploitation of this valuable commodity.[16] Principle 5 emphasizes lifecycle management, which is a key aspect of privacy engineering. Principle 6 resembles the familiar transparency principle found in all versions of FIPs, while Principle 7 functions primarily as a summing up of the earlier principles. Moreover, Cavoukian associates PbD with many other concepts,

14. Among the topics covered are smart grids, Radio Frequency Identification ("RFID"), biometric systems, mobile communications, Wi-Fi positioning systems, and mobile near field communications ("NFC"). *See Publications: Papers*, PBD, http://www.privacybydesign.ca/index.php/publications/papers (last visited Mar. 15, 2013).

15. Instead, many of the papers merely restate or elaborate the seven foundational principles. *See, e.g.*, ANN CAVOUKIAN, OPERATIONALIZING *PRIVACY BY DESIGN*: A GUIDE TO IMPLEMENTING STRONG PRIVACY PRACTICES (Dec. 4, 2012), http://www.ipc.on.ca/images/Resources/operationalizing-pbd-guide.pdf; ANN CAVOUKIAN, ACCESS BY DESIGN: THE 7 FUNDAMENTAL PRINCIPLES (May 10, 2010), http://www.ipc.on.ca/images/Resources/accessbydesign_7fundamentalprinciples.pdf; ANN CAVOUKIAN & MARILYN PROSCH, PRIVACY BY REDESIGN: BUILDING A BETTER LEGACY (May 20, 2011), http://www.ipc.on.ca/images/Resources/PbRD-legacy.pdf.

16. *See* Meglena Kuneva, European Consumer Commissioner, Keynote Speech, Roundtable on Online Data Collection, Targeting and Profiling 2 (Mar. 31, 2009), http://europa.eu/rapid/press-release_SPEECH-09-156_en.htm; *see also* Julia Angwin & Jeremy Singer-Vine, *Selling You on Facebook*, WALL ST. J. ONLINE (Apr. 7, 2012), http://online.wsj.com/article/SB10001424052702303302504577327744009046230.html. Angwin and Singer-Vine wrote:

> This appetite for personal data reflects a fundamental truth about Facebook and, by extension, the Internet economy as a whole: Facebook provides a free service that users pay for, in effect, by providing details about their lives, friendships, interests and activities. Facebook, in turn, uses that trove of information to attract advertisers, app makers and other business opportunities.

Id.

including accountability,[17] risk management,[18] FIPs,[19] and privacy impact assessments ("PIAs").[20] This breadth tends to dilute, rather than clarify, Cavoukian's definition of PbD. As several European computer scientists recently concluded, the principles as written do not make it clear "what 'privacy by design' actually is and how it should be translated into the engineering practice."[21]

Of course, various commentators have taken different approaches to privacy by design. Some see PbD as an offshoot of privacy-enhancing technologies ("PETs");[22] others in terms of a life cycle approach to software development and/or data management (i.e., one that considers privacy at all stages of product design and development);[23] and still others in terms of implementing "accountability based mechanisms" such as risk-based privacy impact assessments.[24] Some regulators combine all of these ideas under the

17. *See* ANN CAVOUKIAN, SCOTT TAYLOR & MARTIN ABRAMS, PRIVACY BY DESIGN: ESSENTIAL FOR ORGANIZATIONAL ACCOUNTABILITY AND STRONG BUSINESS PRACTICES 3 (Nov. 2009), http://www.privacybydesign.ca/content/uploads/2009/11/2009-11-02-pbd-accountability_HP_CIPL.pdf (describing accountability as a business model wherein "organizations tak[e] responsibility for protecting privacy and information security appropriately and protecting individuals from the negative outcomes associated with privacy-protection failures").

18. *See* ANN CAVOUKIAN, INFO. & PRIVACY COMM'N, PRIVACY RISK MANAGEMENT: BUILDING PRIVACY PROTECTION INTO A RISK MANAGEMENT FRAMEWORK TO ENSURE THAT PRIVACY RISKS ARE MANAGED, *BY DEFAULT* 17 (Apr. 2010), http://www.ipc.on.ca/images/Resources/pbd-priv-risk-mgmt.pdf (asserting that privacy risks may be "[m]anaged in a fashion similar to conventional risks . . . by employing the principles of privacy by design").

19. *See* ANN CAVOUKIAN, THE 7 FOUNDATIONAL PRINCIPLES: IMPLEMENTATION AND MAPPING OF FAIR INFORMATION PRACTICES (2011), http://www.ipc.on.ca/images/Resources/pbd-implement-7found-principles.pdf (comparing FIP principles with privacy by design principles).

20. *See* PAT JESELON & ANITA FINEBERG, A FOUNDATIONAL FRAMEWORK FOR A *PBD*–PIA (Nov. 2011), http://privacybydesign.ca/content/uploads/2011/11/PbD-PIA-Foundational-Framework.pdf (offering a framework for a privacy by design privacy impact assessment).

21. Seda Gürses et al., *Engineering Privacy by Design*, International Conference on Privacy and Data Protection ("CPDP") (2011), http://www.dagstuhl.de/mat/Files/11/11061/11061.DiazClaudia.Paper.pdf (arguing that many of the seven principles include the term "privacy by design" in the explanation of the principle itself resulting in recursive definitions).

22. *See generally* Rubinstein, *supra* note 1, at 1414–26.

23. *See* FTC FINAL REPORT, *supra* note 2, at 46–47.

24. *See* E.U. ARTICLE 29 DATA PROTECTION WORKING PARTY, OPINION 3/2010 ON THE PRINCIPLE OF ACCOUNTABILITY (WP 173) 3 (July 2010) [hereinafter WP 173], http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp173_en.pdf; *see also* Paula J. Bruening, *Accountability: Part of the International Public Dialogue About Privacy Governance*, BNA INT'L WORLD DATA PROTECTION REP. 2 (October 2010) (describing the work of an

umbrella of privacy management programs that include policies, procedures, and systems architecture; several recent FTC consent decrees have required companies like Google, Facebook, Twitter, and MySpace to adopt identical five-part programs combining accountability, risk assessment, design processes, due diligence in selecting vendors, and ongoing program adjustments.[25] But the FTC offers firms no guidance about how to implement such programs.

Fortunately, a few private sector firms have developed more detailed privacy guidelines, explaining how to integrate privacy into the several stages of the software development process (requirements, design, implementation, verification, and release).[26] For example, in 2006 Microsoft published a comprehensive set of guidelines that explores nine specific development scenarios and identifies over 120 required and recommend practices for "creating notice and consent experiences, providing sufficient data security, maintaining data integrity, offering customers access [to their data], and supplying [other privacy] controls."[27] Although the guidelines are full of sound advice and would benefit both established and start-up firms, they also have several shortcomings. First—and this is not a problem limited to Microsoft—the tools and techniques concerning "privacy by design" are quite immature, especially as compared with those relied upon for "security by design."[28] Second, the guidelines have not kept up with the transition from client-server products to social media and Web 2.0 services and largely omit this topic, which makes them badly outdated. Finally, the guidelines

expert group convened by the Irish Data Protection Commissioner for the purpose of defining the essential elements of accountability).

25. *See, e.g.*, Agreement Containing Consent Order, Google, Inc., F.T.C. No. 102-3136, 4–5 (Mar. 30, 2011) [hereinafter Google Settlement], http://www.ftc.gov/os/caselist/1023136/110330googlebuzzagreeorder.pdf; Agreement Containing Consent Order, Facebook, Inc., F.T.C. No. 092-3184, 5–6 (Nov. 29, 2011) [hereinafter Facebook Settlement], http://www.ftc.gov/os/caselist/0923184/111129facebookagree.pdf. The third element specifically requires firms to engage in "the design and implementation of reasonable controls and procedures to address the risks identified through the privacy risk assessment." *Id.*

26. *See Role of Privacy by Design*, *supra* note 6.

27. *Privacy Guidelines for Developing Software Products and Services, v. 3.1*, MICROSOFT, 5 (Sept. 2008), http://www.microsoft.com/en-us/download/details.aspx?id=16048 [hereinafter *Microsoft Privacy Guidelines*]. Ira Rubinstein was an Associate General Counsel at Microsoft when these guidelines were first developed but did not contribute to them.

28. In security engineering, there is consensus on the meaning of key concepts and there are tried-and-true design principles and canonical texts, international standards, and a large cadre of certified security experts. Additionally, security professionals may draw upon a variety of technical resources including sophisticated threat-modeling processes, secure coding practices, and automated development and testing tools. Privacy professionals enjoy few of these advantages or resources.

allow business units within Microsoft to balance privacy requirements against business purposes but offer limited guidance on this delicate task.[29] For example, while "essential" actions such as processing of real-time location data, waiver of certain notice requirements, and transfer of sensitive personal information require "Company Approval,"[30] there is little discussion of the relevant factors for granting or withholding such approval. Similarly, the guidelines state that when data transfers or updates are "essential" to the functioning of a product (as defined by Microsoft), this justifies a weaker "all-or-nothing" form of user controls.[31] More generally, Microsoft's internal decision-making process under the guidelines remains opaque to customers and policy makers, which has led to accusations that business or competitive considerations sometimes overwhelm privacy requirements.[32]

All of these varied attempts at fleshing out the meaning of privacy by design are valuable and we have no wish to disparage them. This Article takes a different approach, however. We contend that although FIPs underlie privacy by design, they are not self-executing. Rather, privacy by design requires the translation of FIPs into engineering and design principles and practices. An example helps illustrate what we have in mind. One of the FIPs, the purpose specification principle, is the basis for limits on how long a company may retain personal data. But there is a vast difference between a company promising to observe reasonable limitations on data retention and designing a database that automatically tags personal and/or sensitive information, keeps track of how long the information has been stored, and deletes it when a fixed period of time has expired. To adapt a familiar distinction, one is just words, while the other is action realized through code.

We argue that FIPs must be translated into principles of privacy engineering and usability and that the best way to accomplish this task is to

29. *Microsoft Privacy Guidelines*, *supra* note 27, at § 1.2; *see also infra* notes 456, 461–63 and accompanying text (discussing balancing).

30. The Microsoft Privacy Guidelines define "Company Approval" as "[t]he consent of the authorized privacy council or privacy decision makers within the Company, which may include legal counsel." *Microsoft Privacy Guidelines*, *supra* note 27, at 26.

31. *Id.* at 30, 33, 36.

32. *See* Nick Wingfield, *Microsoft Quashed Efforts to Boost Online Privacy*, WALL ST. J. ONLINE (Aug. 1, 2010), http://online.wsj.com/article/SB10001424052748703467304575383530439838568.html (describing an internal debate in 2008 over privacy features in Microsoft's Internet Explorer ("IE") 8 browser that the advertising division feared would undermine both Microsoft's and its business partners' targeted advertising abilities). Microsoft later reversed this decision and added a very similar feature to IE 9. *See* Nick Wingfield & Jennifer Valentino-DeVries, *Microsoft To Add 'Tracking Protection' to Web Browser*, WALL ST. J. ONLINE (Dec. 7, 2010), http://online.wsj.com/article/SB10001424052748703296045760055422015345 46.html.

review the relevant technical literature and distill the findings of computer scientists and usability experts.[33] This is a departure from most discussions of privacy by design, which tend to slight the small but significant design literature in favor of advocating broad discourse on policy principles and business practices. We seek to remedy this omission and put the design back into privacy by design.

This Article proceeds as follows: In Part II, we present a general review of the design principles relevant to privacy. This requires a brief analysis of the strengths and weaknesses of FIPs as a source of privacy design principles. Here we mainly focus on the failure of the notice-and-choice model of FIPs and the shortcomings of all versions of FIPs insofar as they rely primarily on a control conception of privacy. Next, we closely examine what it means to design for privacy, defining "design" in terms of two broad and at times overlapping ideas: back-end software implementations of networking and related systems infrastructure, which are generally hidden from the user but drive the heart of any system; and front-end user interfaces, which (in the privacy setting) handle tasks such as notification, consent, access, preference management, and other user experiences.[34] We therefore analyze privacy by design from two complementary perspectives: *privacy engineering*, which refers to the design and implementation of software that facilitates privacy, and *usable privacy design*, which refers to design tasks involving human-computer interaction ("HCI"). The former focuses on building software to satisfy the abstract privacy requirements embodied in the FIPs (in some cases overlapping with security engineering), the latter on ensuring that users understand and benefit from well-engineered privacy controls. Our discussion of privacy engineering draws mainly on four key papers in the technical design literature and the works cited therein.[35] In contrast, our discussion of usable privacy design looks at a rather different body of work

33. We also suggest that FIPs must be extended to address the "social dynamics" of privacy. *See infra* notes 59–68, 85–86 and accompanying text.

34. This distinction is by no means absolute. Back-end systems may come with user interfaces and front-end interfaces may rely on sophisticated engineering techniques. For more on this distinction, see generally Rubinstein, *supra* note 1, at 1418, 1422.

35. *See* George Danezis & Seda Gürses, *A Critical Review of 10 Years of Privacy Technology*, *in* PROCEEDINGS OF SURVEILLANCE CULTURES: A GLOBAL SURVEILLANCE SOCIETY? (2010), http://homes.esat.kuleuven.be/~sguerses/papers/DanezisGuersesSurveillancePets2010.pdf; Joan Feigenbaum et al., *Privacy Engineering for Digital Rights Management Systems*, *in* REVISED PAPERS FROM THE ACM CCS-8 WORKSHOP ON SECURITY AND PRIVACY IN DIGITAL RIGHTS MANAGEMENT 79 (Tomas Sander ed., 2002), *available at* http://dl.acm.org/citation.cfm?id=760739; Gürses et al., *supra* note 21; Sarah Spiekermann & Lorrie Faith Cranor, *Engineering Privacy*, 35 IEEE TRANSACTIONS ON SOFTWARE ENGINEERING 67 (2009).

that finds inspiration in the writings of Irwin Altman, a social psychologist, and Helen Nissenbaum, a philosopher of technology—both of whom analyze privacy in terms of social interaction.

Subsequently, in Part III, we offer ten case studies of Google and Facebook privacy incidents and then rely on the principles identified in Part II to discover what went wrong and what the two companies might have done differently to avoid privacy violations and consumer harms. We conclude in Part IV by considering what lessons regulators might learn from this counterfactual analysis.

II. DESIGN PRINCIPLES

A. FAIR INFORMATION PRACTICES ("FIPS") AS THE BASIS OF DESIGN PRINCIPLES

FIPs describe the rights of individuals and the obligations of institutions associated with the transfer and use of personal data.[36] There are many different formulations and they vary in crucial respects.[37] The different versions coalesce around the following nine principles:

1. Defined limits for controllers and processors of personal information on the collection, processing, and use of personal data (often referred to as data minimization);
2. Data quality (accurate, complete, and timely information);
3. Limits on data retention;
4. Notice to individual users;
5. Individual choice or consent regarding the collection and subsequent use of personal information;
6. Reasonable security for stored data;
7. Transparent processing systems that affected users can readily understand and act on;
8. Access to one's personal data; and
9. Enforcement of privacy rights and standards (including industry self-regulation, organizational measures implemented by individual firms, regulatory oversight and/or enforcement, and civil litigation).[38]

36. *See* DANIEL J. SOLOVE & PAUL M. SCHWARTZ, INFORMATION PRIVACY LAW 699 (4th ed. 2010) (explaining that "personal data" generally refers to any data that relates or is linkable to an identifiable individual, including aggregations of data).

37. *See* Fred H. Cate, *The Failure of Fair Information Practice Principles*, *in* CONSUMER PROTECTION IN THE AGE OF THE 'INFORMATION ECONOMY' 341, 341–53 (Jane K. Winn ed., 2006) (discussing six different versions of FIPs).

38. This formulation draws on the work of Paul Schwartz & William Treanor. *See* Paul M. Schwartz & William M. Treanor, *The New Privacy*, 101 MICH. L. REV. 2163, 2181 (2003).

FIPs have many strengths. First, FIPs are universally recognized as the foundation of international privacy law.[39] Second, they are open-ended and, therefore, permit data controllers to take account of all relevant factors.[40] For example, the scope and content of notices depend on a business's specific data processing practices. Similarly, data security measures must be appropriate to a company's size and complexity, the nature and scope of its activities, and the sensitivity of the personal information it holds. Finally, FIPs are both flexible, allowing for social and technological change,[41] and technology neutral, thereby permitting a wide range of solutions.[42] While regulators on both sides of the Atlantic are busy reinterpreting or supplementing FIPs, no one rejects them outright or seriously proposes replacing them.[43]

FIPs have two main weaknesses. First, some versions of FIPs are less comprehensive than others, which may result in a weak foundation for privacy engineering efforts.[44] Second, FIPs mainly reflect a control conception of privacy and therefore provide limited guidance on how to address privacy issues associated with Web 2.0 services in which users generate content and voluntarily share personal data about themselves and their associates.[45]

---

39. *See* Rotenberg, *supra* note 11.

40. *See* ORG. FOR ECON. CO-OPERATION & DEV. ("OECD"), THE EVOLVING PRIVACY LANDSCAPE: 30 YEARS AFTER THE OECD PRIVACY GUIDELINES 12 (2011), *available at* http://dx.doi.org/10.1787/5kgf09z90c31-en [hereinafter OECD, EVOLVING PRIVACY LANDSCAPE] (describing the eight principles of the OECD Guidelines (which parallel FIPs) as "remarkably adaptable to the varying government and legal structures of the implementing countries and the changing social and technological environment"). For the purpose of this discussion, we frequently examine FIPs through the lens of the OECD Guidelines, which evolved out of the FIPs and share many of the same attributes. *See* Cate, *supra* note 37, at 346 ("Fair Information Practices . . . played a significant role in the development of the [OECD Guidelines] . . . .").

41. *Id.* at 12.

42. *Id.*

43. *See* WP 173, *supra* note 24, § 4, at 10; FTC FINAL REPORT, *supra* note 2; *see also* OECD, EVOLVING PRIVACY LANDSCAPE, *supra* note 40; WHITE HOUSE, CONSUMER DATA PRIVACY IN A NETWORKED WORLD: A FRAMEWORK FOR PROTECTING PRIVACY AND PROMOTING INNOVATION IN THE GLOBAL DIGITAL ECONOMY 9–22 (2012) (incorporating FIPs into a "Consumer Privacy Bill of Rights").

44. *See* Cate, *supra* note 37, at 355 (discussing how "the FTC first narrowed the OECD's eight principles down to five—notice, choice, access, security, and enforcement—and then later abandoned enforcement as a 'core' principle").

45. *See* OECD, EVOLVING PRIVACY LANDSCAPE, *supra* note 40, at 27. This is true of most Web 2.0 services, but especially of a social network service ("SNS") such as Facebook.

1. *FIPs or FIPs-Lite?*

It is the received wisdom among most privacy scholars[46] that U.S. privacy law relies on a scaled-down version of FIPs as compared to the more robust version adopted in Europe and other nations that base their national privacy laws directly on the OECD Privacy Guidelines[47] or the E.U. Data Protection Directive.[48] In the United States, both regulators and firms tend to think of FIPs primarily in terms of a notice-and-choice model of online privacy, which requires that businesses post clear and accurate privacy policies describing how they handle consumers' personal information, thereby enabling them to make informed decisions "as to whether and to what extent to disclose personal information."[49] This approach mainly emphasizes procedural requirements over substantive obligations such as fair processing, data minimization, or data quality.[50] As a result, privacy advocates often deride the U.S. version of FIPs as "FIPs-lite."[51]

Obviously, if privacy engineering were premised on FIPs-lite, this would severely limit its value. Under this model, firms may collect whatever data they wish to as long as they provide consumers with notice and obtain opt-

46. *See* Bamberger & Mulligan, *supra* note 5, at 256–57; Cate, *supra* note 37, at 353–54.

47. ORG. FOR ECON. CO-OPERATION & DEV., OECD GUIDELINES ON THE PROTECTION OF PRIVACY AND TRANSBORDER FLOWS OF PERSONAL DATA (1980), http://www.oecd.org/sti/interneteconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldatabackground.htm.

48. Directive 95/46/EC, 1995 O.J. (L 281) 31 [hereinafter E.U. Directive], *available at* http://ec.europa.eu/justice/policies/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf; *see also* Article 29 Data Protection Working Party, *The Future of Privacy: Joint Contribution to the Consultation of the European Commission on the Legal Framework for the Fundamental Right to Protection of Personal Data* at 2 (Dec. 1, 2009), *available at* http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2009/wp168_en.pdf (explaining "that the main principles of data protection are still valid despite the new technologies and globalisation").

49. *Fair Information Practice Principles*, FED. TRADE COMM'N (Nov. 23, 2012), http://www.ftc.gov/reports/privacy3/fairinfo.shtm (limiting FIPs to five core principles—"(1) Notice/Awareness; (2) Choice/Consent; (3) Access/Participation; (4) Integrity/Security; and (5) Enforcement/Redress"—and stating that "the most fundamental principle is notice").

50. *See* Cate, *supra* note 37, at 353.

51. *See Privacy Today: A Review of Current Issues*, PRIVACY RIGHTS CLEARINGHOUSE, http://www.privacyrights.org/ar/Privacy-IssuesList.htm (last visited Apr. 10, 2013); *see also* Bamberger & Mulligan, *supra* note 5, at 254. There are two main criticisms of FIPs-lite: first, there is overwhelming evidence that privacy notices are largely futile because so few people read or understand them properly, with the result that most individuals are unaware of the choices available to them; and second, individual choice, even if achieved, does not equate with privacy protection. *See* Bamberger & Mulligan, *supra* note 5, at 256–58; Cate, *supra* note 37, at 356–67.

out consent. Nothing in the FIPs-lite version obligates firms to build systems that minimize data collection and use, discard (or anonymize) personal data once it has served its purpose, ensure data quality, or provide extensive access rights.[52]

And yet, recent developments suggest that over the past decade, U.S. privacy standards have evolved a great deal since the days of FIPs-lite. For example, in its recent enforcement actions, the FTC has begun to embrace a broader notion of privacy based on "consumer expectations."[53] Indeed, the preliminary FTC staff report took issue with the notice-and-choice model quite explicitly.[54] Similarly, in identifying privacy by design as one of its three key recommendations, the FTC Final Report indicates that companies should incorporate "substantive privacy protections" into their practices such as "data security, reasonable collection limits, sound retention practices, and data accuracy."[55] Finally, the White House framework on consumer data privacy abandons FIPs-lite entirely in favor of a new formulation of FIPs consisting of seven principles,[56] which match up quite well with both the OECD Privacy Guidelines and the E.U. Directive.

In short, the gap between the U.S. and E.U. versions of FIPs is beginning to close, although it will not close entirely until Congress enacts new privacy legislation. Unless and until they are equivalent, however, the applicable version of FIPs will make a significant difference to what it means to build FIPs into products and services. For purposes of this Article, we will discuss

52. *See* Bamberger & Mulligan, *supra* note 5, at 273. In comparison, Article 6 of the E.U. Directive codifies the full set of FIPs by requiring that all personal data must be processed fairly and lawfully; processed in a way strictly limited to the purpose for which such data was collected; adequate, relevant, and not excessive in relation to these purposes; accurate and up-to-date; and kept in a form that permits identification of individuals for no longer than is necessary for the purposes for which the data were collected or for which they are further processed. Article 7(a) establishes high standards for obtaining informed consent. Articles 10 and 11 set detailed, minimum transparency requirements for collecting personal data from individuals. Article 12 grants strong rights of access. Article 17 requires that confidentiality and security of processing be guaranteed. *See* E.U. Directive, arts. 6, 7(a), 10–12 & 17, *supra* note 48.

53. *See* Bamberger & Mulligan, *supra* note 5, at 284–92.

54. BUREAU OF CONSUMER PROTECTION, FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE (PRELIMINARY FTC STAFF REPORT) 19–20 (2010), www.ftc.gov/os/2010/12/101201privacyreport.pdf ("Additionally, the emphasis on notice and choice alone has not sufficiently accounted for other widely recognized fair information practices, such as access, collection limitation, purpose specification, and assuring data quality and integrity."). This report from December 2010 was largely incorporated into the FTC Final Report, published March 2012. *Compare id.*, *with* FTC FINAL REPORT, *supra* note 2.

55. *See* FTC FINAL REPORT, *supra* note 2, at 23.

56. *See* WHITE HOUSE, *supra* note 43, at 1, 9–22.

engineering and usability principles in the context of a robust conception of FIPs, not FIPs-lite.

2. Privacy as Control

Most privacy scholars also agree that at the heart of FIPs is an understanding of privacy as control over personal information. This idea is expressed in Alan Westin's canonical definition of privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."[57] More generally, individual control underpins the protections offered by FIPs, and this cuts across any differences in national privacy laws.

The control paradigm has a major shortcoming: namely, it seems highly unsuited to address a new class of privacy risks associated with social media and Web 2.0 services. When individuals use Facebook or any other social networking service ("SNS"), they voluntarily disclose personal—and often very sensitive—information to their friends and acquaintances. Recent scholarship on social network sites rejects the view that users (especially young users) willingly share personal information because they do not care about privacy[58] in favor of a more nuanced approach based on what James Grimmelmann calls the "social dynamics" of privacy.[59] According to Grimmelmann, the reason that so many Facebook users entrust Facebook with so much personal information is that "people have *social* reasons to participate on *social* network sites, and these social motivations explain both why users value Facebook notwithstanding its well-known privacy risks and why they systematically underestimate those risks."[60]

Grimmelmann's rich and detailed analysis of the social dynamics of privacy leads him to an important insight: many privacy violations on social networks are not caused by the SNS operator; rather they are peer-produced.[61] For example, unwanted disclosures occur when the wrong person sees something intended for a different audience.[62] Users create profiles that are widely available but feel violated by the "snooping" of

57. ALAN F. WESTIN, PRIVACY AND FREEDOM 7 (1967).

58. *See* Lilian Edwards & Ian Brown, *Data Control and Social Networking: Irreconcilable Ideas?*, *in* HARBORING DATA: INFORMATION SECURITY, LAW, AND THE CORPORATION 202 (Andrea Matwyshyn ed., 2009).

59. James Grimmelmann, *Saving Facebook*, 94 IOWA L. REV. 1137 (2009).

60. *Id.* at 1151.

61. *Id.* at 1164 ("Users' privacy is harmed when *other users* learn sensitive personal information about them. Facebook enters the picture as catalyst; it enables privacy violations more often than it perpetrates them." (emphasis added)).

62. *Id.* at 1164–66.

college administrators, legal investigators, or potential employers.[63] Multiple users may tag group photos of embarrassing events but—in the sober light of day—the tagger and the subject of the tag may disagree when the latter asks the former to remove the tag.[64] And the very fact that the structure of one's social network is visible to others may cause spillover harms in which patterns of association leak sensitive information.[65]

Not surprisingly, then, Grimmelmann takes issue with any proposal to "fix" Facebook that disregards the social dynamics of privacy.[66] This includes technical controls, which are ineffective precisely because if they "get in the way of socializing, users disable and misuse them."[67] There is an "irreconcilable tension" between ex ante controls and unplanned social interactions for the simple reason that privacy controls, especially if a user sets them when first registering for an account, utterly fail to capture the nuances of evolving social relationships.[68] Moreover, redistribution of information is inevitable in a social network whose very purpose is to make information accessible to others. And it is these other people who ultimately decide what to do with this shared information irrespective of any privacy controls.[69]

Grimmelmann's argument is compelling but we endorse it with two caveats. The first is to emphasize that not all privacy violations involving social network sites are peer-produced. Rather, as demonstrated by several recent regulatory investigations, many Facebook privacy incidents reflect more traditional privacy concerns such as changing privacy practices without obtaining approval from users, inadequate disclosure and lack of consent to sharing of personal information with third-party applications, insufficient user access to their personal information, and inadequate disclosure about the information made available to advertisers.[70] Granted, these are not the

63. *Id.* at 1164–68.

64. *Id.* at 1171–72.

65. *Id.* at 1174–75.

66. Grimmelmann analyzes and rejects a range of policy options all of which fail because they miss the social dynamics of privacy. *See id.* at 1178–95 (discussing market forces, privacy policies, technical controls, commercial data collection rules, user restrictions, and "data 'ownership' ").

67. *Id.* at 1140.

68. *Id.* at 1185–86 (noting that many users do not understand or ignore Facebook's extensive privacy controls or never modify the default privacy settings).

69. *Id.* at 1186–89.

70. *See* ELIZABETH DENHAM, OFFICE OF THE PRIVACY COMM'R OF CAN., REPORT OF FINDINGS INTO THE COMPLAINT FILED BY THE CANADIAN INTERNET POLICY AND PUBLIC INTEREST CLINIC (CIPPIC) AGAINST FACEBOOK INC. UNDER THE PERSONAL INFORMATION PROTECTION AND ELECTRONIC DOCUMENTS ACT (July 16, 2009),

issues that animate Grimmelmann's analysis but they are problems nonetheless, for which privacy controls may provide adequate solutions. Second, the ex ante technical controls that Grimmelmann rejects do not exhaust the range of possible design-based solutions. In Section II.B, we discuss a set of design practices that are much better suited to address the social dynamics of privacy than are technical controls premised on FIPs.

B. AN ALTERNATIVE APPROACH TO PRIVACY BY DESIGN

Software design encompasses multiple interests and expertise and hence the coordination of multiple parties, each with their own set of concerns, such as business, engineering, marketing, legal, and policy to name a few. Designing a product is also about managing risk, which has both internal sources (engineering, security, compliance) and external sources (market response, press coverage, competition).[71] In a nutshell, good product design creates something the market wants while minimizing these assorted risks. As researchers in behavioral economics have pointed out, however, part of the challenge of designing products to account for privacy risks is that they are not well understood.[72] Privacy risk management is an ongoing but still unsettled area of inquiry for international privacy regulators. For present purposes, we assume that regulators and companies alike agree that it should be incorporated into the design process. But the question is "how?" We turn, now, to the heart of this Article—saying what it means to translate privacy into design practices.

1. Multiple Meanings of Design

For background we will walk through a highly generalized design process. A software product or service typically begins as an idea. Through a series of brainstorming sessions, analysis of feedback, requirements, and iterations, this idea achieves some concrete form, which depends on the user goals and

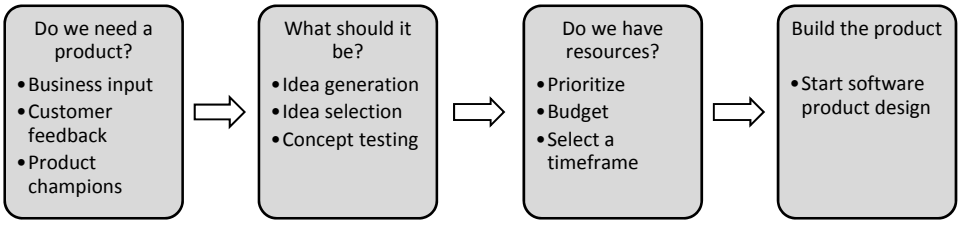
http://www.priv.gc.ca/cf-dc/2009/2009_008_0716_e.pdf; Facebook Settlement, *supra* note 25; OFFICE OF THE DATA PROT. COMM'R, FACEBOOK IRELAND LTD.: REPORT OF AUDIT, § 3.6 (Dec. 21, 2011) [hereinafter IRISH AUDIT], http://dataprotection.ie/documents/facebook report/final report/report.pdf.

71. *See* Sarah Spiekermann, *The Challenges of Privacy by Design*, 55 COMM. ACM 38 (2012).

72. For a good overview of the dichotomy of expressed privacy preferences and user actions, see Alessandro Acquisti & Jens Grossklags, *Privacy Attitudes and Privacy Behavior: Losses, Gains, and Hyperbolic Discounting*, *in* THE ECONOMICS OF INFORMATION SECURITY 171 (L. Jean Camp & Stephen Lewis eds., 2004) [hereinafter Acquisti & Grossklags, *Privacy Attitudes and Privacy Behavior*]. Follow-up work in 2007 explores how behavioral economics can be used to learn about these issues. *See* Alessandro Acquisti & Jens Grossklags, *What Can Behavioral Economics Teach Us About Privacy?*, *in* DIGITAL PRIVACY: THEORY, TECHNOLOGIES, AND PRACTICES 364–65 (Alessandro Acquisti et al. eds., 2007).

the technical processes relied upon to realize them. Initially, these ideas exist as lists of requirements, flow diagrams, wireframes, and related concepts. They are shared with other team members who build out the software design, decide how they are going to define and solve problems, and begin coding the actual product. Figure 1 illustrates this process. This conceptualization phase may seem trivial, but it is an important step in the process, as it motivates the overall way the product will work. This process includes determining the look and feel, the functional requirements, the product goals, and the software architecture of the final product. In some cases, more time is spent on the idea phase than on actual software development. And from a "privacy by design" perspective, influencing the conceptualization phase is essential to ensuring that privacy concepts are taken into account throughout the product development cycle.[73]

Figure 1: High Level Overview of Product Conceptualization



After completing the conceptual phase and deciding to build a product, the next phase consists of "design." The elements of design vary widely by project and timeframe. Factors that typically influence design include the maturity of a company, the motivation behind the design task (i.e., building a new product or updating an existing one), the intended audience, available resources, and so on.[74] Software development methodologies also vary with the nature of the product and constantly evolve.[75] For example, software built for an enterprise may have longer development cycles and use the "waterfall" model, which follows design stages in order (requirements,

73. This is central to Cavoukian's thinking. *See* CAVOUKIAN, *supra* note 13; *see also* Spiekermann, *supra* note 71.

74. *See* Michael Keeling, *Choosing a Software Design Strategy*, REFLECTIONS ON SOFTWARE ENG'G (Aug. 2, 2010), http://neverletdown.net/2010/08/choosing-a-software-design-strategy.

75. *See, e.g.*, FREDERICK BROOKS, THE MYTHICAL MAN MONTH: ESSAYS ON SOFTWARE ENGINEERING 7–8 (2d ed. 1995) (describing the ongoing evolution of software development and the difficulty of predicting or increasing the pace and quality of software development).

design, implementation, testing, release).[76] Software developed by a startup or for fast-moving Internet markets is more likely to rely on the "agile" development processes, which allows small teams to make changes very quickly and measures iterations in days or hours, rather than years or months.[77] Not surprisingly, waterfall or similar top-down approaches are well suited for regulatory compliance (including security and privacy requirements), whereas most agile and lightweight development approaches tend to be more feature-focused. Consequently, the latter methodologies tend to overlook security and privacy requirements at the outset and address them only over the course of several iterations—and sometimes neglect them entirely.[78]

Regardless of which methodology is appropriate to a given project, most programmers have come to rely on a host of software toolkits to assist them with the complex tasks associated with coding and testing software. As software has become more modular, programmers also borrow freely from code libraries, with the overall result that much of the code a programmer uses to build a product or service originates with a variety of third parties. Additionally, business and marketing managers, lawyers, and other non-engineers are now heavily involved in the design of software. Most importantly for present purposes, the growing needs and demands of consumers have encouraged software developers to pay far more attention to the applied art and science of user experience ("UX") design in order to improve the aesthetics, ergonomics, and usability of a product. Thus, a software development team for a popular Web 2.0 service would now typically include industrial designers, graphic artists, visual designers, and usability experts.

76. The waterfall method is one of the oldest in software engineering and is covered in standard textbooks. For a good description, see *Software Process Models*, TARGET: THE SOFTWARE EXPERTS, http://www.the-software-experts.de/e_dta-sw-process.htm (last visited July 23, 2012). This approach emphasizes advance planning but tends to be inflexible in the face of evolving requirements. *Id.*

77. The agile method was first mentioned in the *Manifesto for Agile Software Development*, http://agilemanifesto.org/ (last visited July 23, 2012), as a way to describe an emerging trend in rapid iterative software development. It has since become very popular, especially among start-ups, and has spawned a very large literature.

78. *See* MICROSOFT, SECURITY DEVELOPMENT LIFECYCLE FOR AGILE DEVELOPMENT (June 30, 2009), http://www.blackhat.com/presentations/bh-dc-10/Sullivan_Bryan/BlackHat-DC-2010-Sullivan-SDL-Agile-wp.pdf (defining a way to embrace lightweight software security practices when using Agile software development methods).

a) Front-End Versus Back-End Design: The New Challenges of Designing for Privacy

The design of systems that meet privacy requirements as described in the FIPs has traditionally relied on back-end implementations and system protections, which have largely been the domain of security engineers and legal teams. In fact, some consider privacy design not an engineering discipline at all but merely an adjunct of security engineering "used to ensure that privacy rights are protected to the extent specified by law and organizational policy."[79] We reject this position for two reasons.

First, privacy engineering is an emerging discipline with its own structure and topics, and it is not reducible to security engineering.[80] Second, not all privacy concerns are resolvable by reference to the FIPs or use of associated security-based controls. Several of the privacy incidents that arise in the Google and Facebook case studies illustrate this emerging trend.[81] They are not the result of failures in back-end security engineering but rather nuanced violations of users' perceptions of privacy and of their choices regarding the *context* in which to share and post personal information.[82]

These developments in turn raise an interesting and still-unanswered question: Which part of an organization should be responsible for designing in privacy and addressing these context-based requirements? More importantly, how should we define these requirements or measure them for engineering purposes? In what follows, we offer some preliminary answers to these questions but for the moment merely wish to emphasize the importance of including UX designers in the conversations on privacy and product design.[83] We maintain that UX designers should have a strong role in defining privacy requirements. By working closely with legal and security engineers early on in the development and design process, they help ensure that privacy expectations as understood by end users are taken into account in a nuanced way.

79. DEBRA S. HERRMANN, A COMPLETE GUIDE TO SECURITY AND PRIVACY METRICS: MEASURING REGULATORY COMPLIANCE, OPERATIONAL RESILIENCE AND ROI 523 (2007).

80. *See* Spiekermann & Cranor, *supra* note 35, at 67–68.

81. *See infra* Part III.

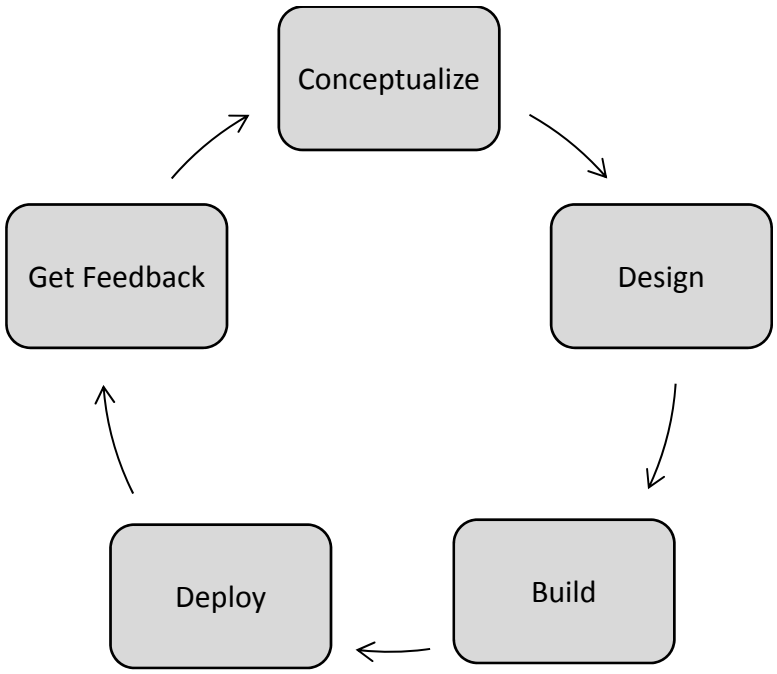
82. For an overview of Nissenbaum's theory of privacy as contextual integrity, see *infra* notes 204–17 and accompanying text.

83. On the value of incorporating HCI factors into privacy by design, see Deirdre Mulligan & Jennifer King, *Bridging the Gap Between Privacy and Design*, 14 U. PA. J. CONST. L. 989, 1019–26 (2012).

b) Putting Design into Privacy by Design

The most reliable way to incorporate privacy design into product development is to include privacy considerations in the definition of software "requirements" or specifications. This "roadmap" or product blueprint typically guides the creation of a software product, codifying what needs to be implemented as part of the product iteration, whether it is a short sprint or a longer term multi-phase development cycle, as depicted in Figure 2. If the privacy concerns are addressed by the FIPs, defining and implementing the relevant requirements are fairly straightforward tasks, and may even lend themselves to engineering metrics.[84]

Figure 2: A Schema for Software Development



In many of the cases discussed below, however, the privacy concerns have less to do with FIPs than with the social dynamics of privacy, which makes defining software requirements a fuzzier and hence more difficult task. Whereas legal teams play a significant role in advising development

84. For example, measurements can track how long a search engine stores personal data and whether it is deleted in a timely fashion in accordance with applicable legal requirements or corporate privacy guidelines. For a discussion of data retention issues, see *infra* notes 132–38 and accompanying text.

teams on how to implement the FIPs, UX designers are much better suited than lawyers at interpreting HCI requirements (and they contribute to back-end design tasks as well).[85] For example, a startup may want to implement functionality for requesting access to a user's address book or contact list for use with a mobile service. A UX designer would respond by researching consumer expectations about the use of such private data in a mobile application and develop insights about what approaches work best to balance competing goals such as transparency and a high rate of user interest and acceptance. UX designers perform these tasks in tandem with engineering teams and business managers and play an increasingly important role in developing consumer software, making them highly sought after by top companies.[86]

In what follows, we flesh out the meaning of "privacy by design" by separately discussing privacy engineering and usable privacy design. This is mainly an expository convenience and does not necessarily reflect the nature or goals of the software development process, which ideally incorporates both engineering and design in a unified manner. Another reason to discuss engineering principles separately from design principles is that they take their inspiration from different sources and have been explored in different research literatures.

2. Privacy Engineering

a) Background

We suggested earlier that privacy by design requires a translation of FIPs into engineering principles. Ten years ago, Joan Feigenbaum and her colleagues attempted to do just that for digital rights management ("DRM") systems.[87] Their paper argues that blind signatures, zero knowledge proofs, selective disclosure of credentials, and other sophisticated cryptographic protocols that form the basis of most PETs have not solved the privacy

85. Policy makers have been hesitant to make specific design recommendations because they lack the relevant expertise and are wary of stifling innovation. In the absence of any legally enforceable design principles analogous to the FIPs, designers have considerable leeway in determining best practices. FTC decisions generally support this flexible approach, as long as a user interface is not "unfair or deceptive" and users are given "clear and prominent notice." *See* F.T.C. v. Frostwire LLC, No. 1:11-CV-23643-DLG at 9, 11 (S.D. Fla. 2011), http://www.ftc.gov/os/caselist/1123041/111012frostwirestip.pdf.

86. *See* Lance Whitney, *Facebook Hires Former Apple Design Manager*, CNET (June 22, 2012), http://news.cnet.com/8301-1023_3-57458870-93/facebook-hires-former-apple-design-manager.

87. *See* Feigenbaum et al., *supra* note 35. We use the term "privacy engineering" partly because it suggests the entire range of software engineering and design techniques as applied to the protection of personal data and partly because it overlaps with many aspects of better-known security engineering practices.

problems raised by DRM.[88] Indeed, they reject cryptographic PETs—citing a number of shortcomings that we believe remain true today[89]—and instead put forward the thesis that if DRM were "properly designed, implemented, and used," it could provide "reasonable user-privacy protection and simultaneously supply businesses with information necessary for their basic functionality at a fair cost."[90]

This is the approach we adopt here. Before developing our own set of privacy engineering principles, however, we need to clarify two points. The first is that the privacy research community has by no means abandoned PETs based on cryptographic protocols.[91] Indeed, a recent and important paper by Seda Gürses et al. strongly reaffirms the cryptographic approach.[92] We have no quarrel with this analysis or with the attempt to generalize the lessons learned from them. Rather, we reject the binary choice of Gürses et al. (strong cryptography or no privacy) in favor of the more nuanced analysis of Feigenbaum et al., which concludes that in spite of the apparent profusion of cryptographic technologies, few are in widespread use and even if they were, they would not necessarily overcome the technical and economic barriers blocking their deployment.[93] In fact, exposure of sensitive information remains a problem even when systems utilize encryption techniques.[94] That said, we also agree with the view of Feigenbaum et al. that

88. For a review of these technologies as of 2002, see IAN GOLDBERG, PRIVACY-ENHANCING TECHNOLOGIES FOR THE INTERNET, II: FIVE YEARS LATER pt. 3, at 4–7 (2002), *available at* http://www.cypherpunks.ca/~iang/pubs/pet2.pdf.

89. Feigenbaum et al., *supra* note 35, at 82–88 (citing both technical issues such as overdependence on abstract models as opposed to real-world uses, insecure implementations, ease-of-use issues, difficulties integrating PETs with legacy systems, and excessive technical costs; and a variety of economic issues).

90. *Id.* at 78 (citations omitted).

91. For a review of contemporary privacy technology, see Danezis & Gürses, *supra* note 35.

92. *See* Gürses et al., *supra* note 21, § 2.2 (arguing that data minimization with strong, technical guarantees "must be the foundational principle in applying privacy by design" to systems that collect data in massive databases).

93. *Compare id.*, *with* Feigenbaum et al., *supra* note 35, at 81–88. While cryptographic techniques might find their way into government-managed systems, there is little evidence that businesses will adopt them absent much stronger incentives or a government mandate as in the *Proposed E.U. Regulation*. *See* Rubinstein, *supra* note 1, at 1431–44.

94. *See* Nathaniel Good & Aaron Krekelberg, *Usability and Privacy: A Study of KaZaA P2P File Sharing*, *in* SECURITY AND USABILITY: DESIGNING SECURE SYSTEMS THAT PEOPLE CAN USE 651, 652–53 (Lorrie Faith Cranor & Simon Garfinkel eds., 2005) [hereinafter SECURITY AND USABILITY] (discussing inadvertent sharing of personal information in P2P software as an early example of usability issues resulting in loss of private information). For a more recent incident in which address books were uploaded over a secure channel but without users' consent, see Joshua Topolsky, *Privacy Controversy over Path for iPhone, iPad Should*

even if cryptography cannot by itself solve the privacy problems raised by businesses collecting, storing, and sharing data and using it for profitable purposes, "it *can* play a role in various solutions."[95]

The second point we need to clarify is that the term "privacy engineering" encompasses not only the work of Feigenbaum et al. but also a variety of other approaches.[96] These include the analysis of data minimization of Gürses et al.,[97] in addition to works on requirements engineering,[98] privacy policy languages and user preference tools,[99] privacy-aware access controls,[100] privacy rights management,[101] identity management,[102] and privacy threat modeling.[103] In what follows, we give no or only passing mention to most of these alternatives, not because they lack value but rather because the FIPs-based approach of Feigenbaum et al., supplemented by the "architectural" approach of Spiekermann and Cranor,[104] better suit our purposes.

Be a Wake-Up Call, WASH. POST (Feb. 15, 2012), http://www.washingtonpost.com/business/technology/privacy-controversy-over-path-for-iphone-ipad-should-be-a-wake-up-call/2012/02/15/gIQA8oHVGR_story.html ("[W]hen you logged into the app on an Apple iOS device—an iPhone or iPad—it automatically uploaded your entire address book to its servers. Without asking.").

95. Feigenbaum et al., *supra* note 35, at 76 (emphasis added).

96. *See* Paolo Guarda & Nicola Zannone, *Towards the Development of Privacy-Aware Systems*, 51 INFO. & SOFTWARE TECH. 337 (2009).

97. *See supra* note 92 and accompanying text.

98. *See* Travis D. Breaux & Annie I. Antón, *Analyzing Regulatory Rules for Privacy and Security Requirements*, 34 IEEE TRANSACTIONS ON SOFTWARE ENGINEERING 5 (2008) (describing formal methods for extracting descriptions of rules from the policies and regulations that govern stakeholder actions).

99. *See infra* notes 146–52 and accompanying text (discussing P3P).

100. *See* Guarda & Zannone, *supra* note 96, at 343.

101. *See* Larry Korba & Steve Kenny, *Towards Meeting the Privacy Challenge: Adapting DRM* (Nat'l Research Council of Can., Paper No. 44,956, 2002), http://crypto.stanford.edu/DRM2002/KorbaKennyDRM20021.pdf.

102. *See generally* DIGITAL PRIVACY: PRIME—PRIVACY AND IDENTITY MANAGEMENT FOR EUROPE (Jan Camenisch et al. eds., 2011).

103. *See* M. Deng et al., *A Privacy Threat Analysis Framework: Supporting the Elicitation and Fulfillment of Privacy Requirements*, 16 REQUIREMENTS ENGINEERING 3 (2011) (describing a threat modeling approach that maps privacy threats to elements in a system model and identifies countermeasures based on existing PETs).

104. *See* Spiekermann & Cranor, *supra* note 35, at 67 (distinguishing "architecture" from "policy" and suggesting that if a firm implements a privacy architecture, it should be exempted from providing notice and choice). We reject this sharp distinction, which stems, in part, from thinking of policy in terms of FIPs-lite. Instead, we rely on a more robust version of FIPs, and argue that the FIPs-based approach to privacy engineering described here bridges the gap between architecture and policy.

b) FIPs-Based Privacy Engineering

We organize the following discussion around the FIPs. Not all of the FIPs are equally relevant; hence, we focus especially on data avoidance and minimization, data retention limits, transparency, individual choice and access, and accountability.[105] As a preliminary matter, FIPs only apply to personally identifiable information ("PII"). Although there is no uniform definition of PII,[106] privacy laws "all share the same basic assumption—that in the absence of PII, there is no privacy harm."[107] It follows that the two most basic privacy engineering principles are protecting PII against unauthorized access and limiting the linkability of data to personal identifiers. This may entail encrypting PII in transit and in storage and/or the use of anonymity services that delink users from all traces of their online activity,[108] or of user-centric identity management systems that enable anonymous or pseudonymous credentials.[109] Other techniques for limiting linkability are better characterized as data avoidance or minimization techniques. These include not recording IP addresses and/or not enabling User ID cookies, or using a third party proxy server to strip out an IP address; and a variety of techniques that protect, shield, and minimize location data,[110] from which identity is readily inferred.

In practice, few companies build services that implement these techniques, and those that do labor in obscurity. Rather, most companies treat privacy (like security) as primarily a compliance task best left to lawyers, not product developers.[111] As a result, the vast majority of Internet services collect information about web activity and link it to IP addresses or other identifiers.[112] Of course, consumer advocates encourage users to rely on a

105. Security also belongs on this list but we omit it here because the topic requires a separate paper of which there are many.

106. *See* SOLOVE & SCHWARTZ, *supra* note 36, at 1828–36 (identifying three approaches to defining PII).

107. *Id.* at 1816. Although Schwartz and Solove limit this observation to U.S. privacy law, it holds true for E.U. data protection law as well. *See id.*

108. This can be done, for example, by using proxies or "mix" systems that shield or hide a user's IP address and other identifiers. *See* Danezis & Gürses, *supra* note 35, at 2–3.

109. *Id.* at 5–6.

110. *See* KIM CAMERON & ANN CAVOUKIAN, WI-FI POSITIONING SYSTEMS: BEWARE OF UNINTENDED CONSEQUENCES 16 (June 2011), http://www.ipc.on.ca/images/Resources/wi-fi.pdf (identifying techniques such as "location fuzzing or obfuscation, ambient notices, two-way communication for privacy notices and consent, user-generated identifiers, substitution of MAC addresses, cloaking, changing identifiers in mix zones, etc." (citations omitted)).

111. *See* Spiekermann, *supra* note 71.

112. *Id.*

variety of self-help strategies that would prevent companies from associating their browsing activity with identifiers or linking their browsing activity across different websites.[113] But our focus here is not on what *users* can do to defeat privacy invasions, but on what *companies* can do to build privacy protections into their own systems by implementing privacy engineering principles. It follows that these self-help strategies are outside the scope of this Article and that where we discuss anonymization techniques, we will do so under the broader heading of data avoidance and minimization.

c) Data Avoidance and Minimization

Data avoidance and minimization are central tenets of the FIPs, the E.U. Directive, and certain U.S. privacy laws and play a key role in the work of Feigenbaum et al. as well. For example, she recommends that DRM systems enable users to "easily configure the system to accommodate their preferred information-collection and handling procedures," which she refers to as "customizable privacy."[114] In order for configurable systems to support data minimization, however, they must by default be set to avoid or minimize the collection of PII, which in turn requires that engineers analyze information needs and flows at the outset of any design project, and consider techniques for disassociating functionality that requires PII (e.g., paying for music or movies with a credit card) from activation, recommendation services, and other functionality, for which pseudonyms should suffice. As Feigenbaum et al. points out, this also requires that businesses determine at the outset which information is necessary for different business practices and whenever possible build systems that achieve business purposes without collecting PII.[115] It also requires serious attention to database architecture and management. "Data may be segmented," Feigenbaum et al. suggest, "according to the different groups that are interested in it—a principle of *split databases* and *separation of duty*."[116]

113. *See, e.g.*, *EPIC Online Guide to Practical Privacy Tools*, ELEC. PRIVACY INFO. CTR. (EPIC), http://epic.org/privacy/tools.html (last visited Feb. 27, 2013) (describing various ways to enhance personal privacy online).

114. Feigenbaum et al., *supra* note 35, at 91.

115. *Id.* at 16–17; *cf.* Gürses et al., *supra* note 21, § 3.1 (discussing the use of advanced cryptographic protocols to minimize data collection and maintain anonymity).

116. Feigenbaum et al., *supra* note 35, at 92 (discussing, for example, separating accounting and customer service data requirements from those of marketing and risk management). For a similar approach, see MICROSOFT, MICROSOFT'S PRIVACY PRINCIPLES FOR LIVE SEARCH AND ONLINE AD TARGETING (July 2007), http://www.reallyfirst.com/ad-trageting/microsofts-privacy-principles-for-live-search-and-online-ad-targeting.shtml ("We will store our Live Search service search terms separately from account information

Spiekermann and Cranor offer the most comprehensive discussion of the importance of architectural choices to privacy engineering.[117] They argue that "engineers typically can make architectural choices on two dimensions: network centricity and identifiability of data."[118] Network centricity refers to "the degree to which a user's system relies on a network infrastructure to provide a service, as well as the degree of control a network operator can exercise over a client's operations."[119] The extent of privacy protection lies on a continuum between network-centric systems and client-centric systems. Not surprisingly, businesses prefer network-centric architectures, which gives them much greater control over how their systems work, and competitive advantages if they can design a better system than others. Unfortunately, privacy risks are also greater with network-centric systems (which must collect and store personal data in providing a service to users). In contrast, privacy problems on client-centric systems are greatly reduced because these systems have less or no need to transfer personal data to a web server, thereby eliminating data retention issues and/or unwanted secondary use.[120]

User identifiability refers to "the degree to which data can be directly attributed to an individual."[121] The authors note that many service providers are already familiar with this approach to reducing privacy concerns and offer their users pseudonymous access to their services. But pseudonyms alone are insufficient because they allow service providers to reidentify users.[122] This occurs in either of two ways: the service combines a user's pseudonymous profile with PII stored in a billing or shipping database (i.e., it fails to follow the principle of Feigenbaum et al. of split databases and separation of duty);[123] or the service applies data mining techniques to pseudonymous

that personally and directly identifies the user, such as name, email address, or phone numbers . . . .").

117. Spiekermann & Cranor, *supra* note 35.

118. *Id.* at 74.

119. *Id.*

120. *Id.* Spiekermann and Cranor offer two examples: a collaborative filtering system (for use in various recommendation services) in which users have control over and access to all data recorded about their preferences, and a location-based service enabling smart phones and other clients to calculate their own positions without having to share location-information with a central server. *Id.*; *see also* Mikhail Bilenko et al., Targeted, Not Tracked: Client-Side Solutions for Privacy-Friendly Behavioral Advertising (Sep. 25, 2011) (Telecommunications Policy Research Conference Accepted Paper Series), http://petsymposium.org/2011/papers/hotpets11-final3Bilenko.pdf (discussing "methods that facilitate behavioral targeting while providing consumer privacy protections").

121. Spiekermann & Cranor, *supra* note 35, at 74.

122. *Id.* at 75.

123. *See supra* note 116 and accompanying text.

transaction logs or otherwise links user data to personal identifiers by pattern matching.[124]

Arguing that "the degree of privacy friendliness of a system is inversely related to the degree of user data identifiability," the authors discuss concrete steps that engineers can take to specify the degree of user identifiability in a given system.[125] They describe four privacy stages and their corresponding system characteristics as illustrated in their framework for privacy-friendly system design. At one end of the spectrum, Stage 0, privacy is limited and identifiability is easy because systems utilize unique identifiers across databases and store contact information with profile information, thus linking data to personal identifiers. Stage 1 provides a minimal degree of privacy by eliminating unique identifiers and common attributes across databases while storing contact information separately from profile or transaction information (which is stored under a pseudonym).[126] But reidentification of users is still possible with a reasonable effort (as in the AOL case)[127] and may even be automated, rendering it cost effective.[128]

At the other end of the spectrum, Stage 2 systems are "actively designed for non-identifiability of users" and thereby achieve what the authors denote as "privacy-by-architecture." These systems have all of the characteristics of Stage 1 but also take steps to generate identifiers randomly to prevent future databases from reintroducing common identifiers and endeavor to collect long-term personal characteristics at a low level of granularity (e.g., year of birth where possible rather than date of birth).[129] Even if these steps are taken, however, data-mining techniques may still be used to reidentify a user based on comparisons of an anonymous database and a subset of similar, identified data, although such linking attempts would require a relatively high level of effort compared with Stage 1.[130] Thus, the coauthors also describe

124. Spiekermann & Cranor, *supra* note 35, at 75.

125. *Id.*

126. *Id.* at 76.

127. *See infra* note 249 and accompanying text; *see also* Complaint, Myspace LLC, F.T.C. No. 102-3058 (May 8, 2012), *available at* http://ftc.gov/os/caselist/1023058/120508 myspacecmpt.pdf (charging MySpace with misrepresenting whether advertisers could link user IDs to their broader web-browsing activities).

128. As a result, Stage 1 systems must be supplemented by policies that prohibit or restrict reidentification and give users adequate notice of these policies and other steps they might take to protect their privacy. *See* Spiekermann & Cranor, *supra* note 35, at 75–76.

129. *Id.* at 76.

130. *See* Arvind Narayanan & Vitaly Shmatikov, *Robust De-anonymization of Large Sparse Datasets (How to Break Anonymity of the Netflix Prize Dataset)*, *in* PROCEEDINGS OF 29TH IEEE SYMPOSIUM ON SECURITY AND PRIVACY 111 (2008), *available at* http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4531148 (discussing successful efforts at reidentifying

Stage 3, in which privacy is very well protected and users remain anonymous, either because there is no collection of contact information or of long-term personal characteristics, or profiles are deleted and anonymized via more sophisticated techniques.[131]

d) Data Retention Limits

As noted above, Article 6(1)(e) of the Directive specifically limits the period of time a controller may retain identifiable data to a period "no longer than is necessary" for the purposes for which they were collected or processed.[132] Consequently, this implies that data must be erased or de-identified as soon as they are no longer needed. Both in Europe and the United States, questions over the appropriate length of retention periods and the technique used for anonymization or de-identification play out mainly in the context of search engines and targeted advertising,[133] and, more recently, SNSs.[134] Feigenbaum et al. argues that the practice of data erasure should be addressed in the context of database architecture and management and recommends that the PII can and should be removed from usage records on

some poorly anonymized data records of Netflix movie rankings by comparing this data with public, identifiable information in the Internet Movie Database, IMDB, http://www.imdb.com (last visited Mar. 25, 2013)).

131. *See* Spiekermann & Cranor, *supra* note 35, at 76 (noting that stage 2 "does not guarantee unlinkability; rather, it ensures that the process of linking a pseudonym to an individual will require an extremely large effort"). For the ongoing debate over the effectiveness of even the most sophisticated anonymization techniques, compare Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010) (arguing that anonymization fails to protect privacy due to the threat of reidentification of anonymized data sets), with Jane Yakowitz, *Tragedy of the Data Commons*, 25 HARV. J. LAW & TECH. 1 (2011) (arguing that properly de-identified data is not only safe, but has high social utility). *Cf.* Felix T. Wu, *Privacy and Utility in Data Sets*, U. COLO. L. REV. (forthcoming 2013), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2031808 (staking out a middle ground and explaining that the debate over anonymization turns on society's goals for privacy and utility in specific contexts).

132. *See supra* note 52.

133. This debate included security experts objecting to the weak anonymization techniques relied on by certain firms. *See, e.g.*, Chris Soghoian, *Debunking Google's Log Anonymization Propaganda*, CNET (Sep. 11, 2008), http://news.cnet.com/8301-13739_3-10038963-46.html. Soghoian took issue with Google's method of removing the last eight bits in the IP address and changing the cookie information by pointing out that:

> Since each octet (the numbers between each period of an IP) can contain values from 1–255, Google's anonymization technique allows a user, at most, to hide among 254 other computers. In comparison, Microsoft deletes the cookies, the full IP address and any other identifiable user information from its search logs after 18 months.

Id.

134. *See supra* note 70.

a massive scale, "before those records are inserted into a long-lived data warehouse."[135] Spiekermann and Cranor's privacy framework takes account of recent developments in pattern matching techniques[136] but shifts the debate from how to reduce the risks of reidentification to how to design systems that avoid identification of users in the first place.[137] As to data retention limits, they recommend not only the deletion of PII after its purpose has been fulfilled but also the "purging [of] nonidentified data as well, to minimize the risk of reidentification based on pattern matching."[138]

e) Notice, Choice, and Access

We have already rehearsed the limits of the notice-and-choice model, but the fact remains that whatever their shortcomings as a standalone privacy strategy, notice and choice, together with access, are not going away.[139] Despite the serious design challenges presented by this model, this section very briefly considers a few design options.

Most commentators agree that adequate notice must be understandable, timely, and widely disseminated (not only to consumers but also to other systems that must respect the assumptions under which consumers make privacy decisions).[140] The Microsoft Privacy Guidelines contain a useful discussion of different types of notice[141] and different notice mechanisms.[142] This guidance remains highly relevant today, especially in light of the recent controversies over Google's decision to consolidate its many separate privacy

135. Feigenbaum et al., *supra* note 35, at 93.

136. *See* Narayanan & Shmatikov, *supra* note 130.

137. *See* Spiekermann & Cranor, *supra* note 35, at 76. Spiekermann and Cranor wrote:

> Separate databases for profile and contact information must be created in such a way that common attributes are avoided. In addition, steps should be taken to prevent future databases from reintroducing common identifiers. Identifiers should therefore be generated at random and any information that is highly specific to an individual (e.g., birth dates or contact data) should be avoided whenever possible.

Id.

138. *Id.*

139. For example, notice, choice, and access remain central to both the Proposed E.U. Regulation and the White House's proposed Consumer Privacy Bill of Rights. *See Proposed E.U. Regulation*, *supra* note 2; WHITE HOUSE, *supra* note 43.

140. *See* Feigenbaum et al., *supra* note 35, at 93; Spiekermann & Cranor, *supra* note 35, at 77–79.

141. *See Microsoft Privacy Guidelines*, *supra* note 27, § 1.3.1 (distinguishing prominent notice, discoverable notice, and layered notice).

142. *Id.* § 1.4 (distinguishing just-in-time notice, first-run notice, installation-time notice, and "out of the box" notice).

policies into a single, comprehensive policy.[143] There is a large literature on how to improve privacy policies, describing a number of different approaches.[144] Here we distinguish and briefly discuss both an engineering approach, which relies on the Platform for Privacy Preferences ("P3P") standard for specifying and handling privacy policies in an automated and integrated fashion, and a usability approach, which seeks to redesign cookie handling in browsers based on a model of informed consent.[145]

P3P is the oldest and best-known specification of privacy policies.[146] This W3C standard enables websites and services to encode their privacy practices in machine-readable XML format and allows user agents "to make automated privacy choices based on a user's stored privacy preferences."[147] In practice, P3P has been sharply criticized on technical, legal, and policy grounds.[148] It also has difficult user interface problems.[149] Microsoft's adoption of the P3P framework in Internet Explorer ("IE") has very broad reach but limits P3P functionality to merely signaling whether a website meets a user's cookie preferences; even so, few users are likely even to be aware of the "Privacy Report" icon on the IE status bar.[150] In contrast, more robust P3P implementations, such as Privacy Bird, have a very small audience.[151] In short, P3P has yet to fulfill its supposed potential, although research and experimentation continues with P3P-based privacy tools.[152]

In 2000, a group of researchers developed a model of informed consent for information systems based on six components: disclosure, comprehension,

143. *See infra* notes 320–37 and accompanying text.

144. Much of this literature originates at the Carnegie Mellon CyLab, in the work of Lorrie Cranor and her colleagues. CYLAB, CARNEGIE MELLON UNIV., http://www.cylab.cmu.edu/index.html (last visited Feb. 27, 2013).

145. *See* LORRIE CRANOR ET AL., THE PLATFORM FOR PRIVACY PREFERENCES 1.0 (P3P1.0) SPECIFICATION, W3C RECOMMENDATION (Apr. 16, 2002), http://www.w3.org/TR/P3P.

146. *Id.*

147. *See* Spiekermann & Cranor, *supra* note 35, at 78 (citations omitted).

148. For technical issues, see Guarda & Zannone, *supra* note 96, at 342–43. For legal and policy issues, see William McGeveran, *Programmed Privacy Promises: P3P and Web Privacy Law*, 76 N.Y.U. L. REV. 1812 (2001).

149. *See* Mark S. Ackerman, *The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility* 15 HUM.-COMPUTER INTERACTION 179, 184–87 (2000).

150. *See* Tom Spring, *First Look at Microsoft IE 6.0*, PCWORLD (Aug. 28, 2001), http://www.pcworld.com/article/59928/article.html (describing the implementation of P3P into Microsoft's IE and noting some of the privacy concerns of privacy advocates).

151. PRIVACY BIRD, http://www.privacybird.org (last visited Mar. 17, 2013).

152. *See* Aleecia M. McDonald, et al., *A Comparative Study of Online Privacy Policies and Formats*, *in* PRIVACY ENHANCING TECHNOLOGIES: 9TH INTERNATIONAL SYMPOSIUM, PETS 2009, SEATTLE, WA, USA, AUGUST 5–7 2009, PROCEEDINGS 37 (Ian Goldberg & Mikhail Atallah, eds., 2009).

voluntariness, competence, agreement, and minimal distraction.[153] In subsequent papers, they explored how cookie technology and web browser design have responded to concerns over informed consent and found significant design problems, which they sought to remedy through the development of new technical mechanisms for cookie management using a value-sensitive design methodology.[154] A later paper reviews the design possibilities for implementing informed consent not only in web browsers but also in other widely deployed technologies, such as secure web connections and a web email service, and proposed ten design principles.[155] Interestingly, these principles overlap to some extent with Cavoukian's emphasis on proactive design and attention to default,[156] while also raising classic HCI concerns such as the system interactions with both direct and indirect stakeholders, the use of icons that support an accurate mental model of information flows, and extensive field testing to validate and refine initial designs.[157] Unfortunately, few companies follow this approach in developing their information and computer systems.

Finally, turning to access, it is now commonplace for both e-commerce and Web 2.0 services to provide users with direct online access to their PII by means of a password-protected account. The scope of access varies with the service but may include the ability to view or edit user profiles, billing and account information, and privacy settings, as well as data sharing and communications preferences. For example, Google allows users to review data associated with their Google Accounts via Dashboard and to remove or edit their interests and inferred demographics associated with their cookies via Ads Preferences.[158] Additionally, when Facebook was inundated with over 40,000 access requests from European users within a period of weeks, it quickly developed technical means to expand the range of data it made

153. *See* BATYA FRIEDMAN ET AL., INFORMED CONSENT ONLINE: A CONCEPTUAL MODEL AND DESIGN PRINCIPLES 1–4 (2000), *available at* ftp://ftp.cs.washington.edu/tr/2000/12/UW-CSE-00-12-02.pdf.

154. *See* LYNETTE I. MILLETT ET AL., COOKIES AND WEB BROWSER DESIGN: TOWARD REALIZING INFORMED CONSENT ONLINE 7–8 (2001), *available at* ftp://ftp.cs.washington.edu/tr/2000/12/UW-CSE-00-12-03.pdf; Batya Friedman, et al., *Informed Consent in the Mozilla Browser: Implementing Value-Sensitive Design*, *in* PROCEEDINGS OF THE THIRTY-FIFTH HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES (IEEE 2002), http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=994366.

155. *See* Batya Friedman et al., *Informed Consent by Design*, *in* SECURITY AND USABILITY 495.

156. *See* CAVOUKIAN, *supra* note 13.

157. *See* Friedman et al., *supra* note 155.

158. *See* Dennis O'Reilly, *How to Prevent Google from Tracking You*, CNET (Jan. 30, 2012, 12:57 PM), http://howto.cnet.com/8301-11310_39-57368016-285/how-to-prevent-google-from-tracking-you.

available via a user's activity log profiles, a user-accessible database, and a new download tool.[159]

f) Accountability

We noted previously that many regulators think of privacy by design as one way of achieving accountability, defined as data governance by organizations for the purpose of demonstrating compliance with FIPs.[160] Here we observe that companies may also adopt technical measures to audit and enforce their data privacy practices. Both Feigenbaum et al., and Spiekermann and Cranor, make this point: the former favor a combination of privacy notices and audits but emphasize that effective auditing requires stronger tools,[161] while the latter identify some new tools that help automate the evaluation of data access requests according to a set of privacy rules.[162] Both recognize that even if such accountability measures control who can access PII for legitimate purposes, they cannot control whether such data will be misused once accessed. In other words, auditing offers no cryptographic guarantees of privacy, but it does provide practical solutions that are both familiar and cost-effective.[163] Auditing can also "help[] protect against unintentional privacy violations" and assist management in determining "who may be responsible should a breach occur."[164]

3. Designing for Privacy: A UX Approach

a) Background

As distinct from translating FIPs into engineering principles and practices, a complementary approach consists of embedding privacy into UX design processes, which generally handle both usability and visual and aesthetic design. Design, of course, is a discipline unto itself, with strong roots in the creative arts, and it plays an important role in today's modern engineering practice. Apple is often cited as a proponent of excellence in the design of consumer goods, and its enormously popular devices and software systems seem to confirm the wisdom of obsessive attention to design details.[165] Furthermore, Apple's success tends to support the notion that a

159. *See* IRISH AUDIT, *supra* note 70, at 63–68.
160. *See supra* note 24 and accompanying text.
161. *See* Feigenbaum et al., *supra* note 35, at 96.
162. *See* Spiekermann & Cranor, *supra* note 35, at 79.
163. *See* Feigenbaum et al., *supra* note 35, at 96–97.
164. Spiekermann & Cranor, *supra* note 35, at 79.
165. *See, e.g.*, Matt Brian, *Apple Wins Prestigious UK Design Awards, Flies Entire Design Team to London to Pick Them Up*, THE NEXT WEB (Sept. 19, 2012 9:24 AM), http://thenextweb.com/apple/2012/09/19/apple-wins-prestigious-uk-design-awards-flies-entire-design-team-london-pick

design-centric approach helps avoid usability problems that otherwise undermine many desirable products—everything from medical software and devices, to enterprise software, to security and even encryption software.[166]

Arguably, good design in software is more prevalent than ever. For example, content management systems such as WordPress[167] and Drupal[168] make it easy for novice web users to construct entire websites with a click of a mouse and thereby incorporate good aesthetics and design principles. Indeed, one could claim that the proliferation of well-designed and easy-to-use consumer products, coupled with the wide availability of alternatives in most product categories, has led consumers to take good design for granted. However, this design boom, as it relates to software aesthetics and usability, has created many different, and sometimes conflicting, approaches to incorporating design elements into the software development process. Design is now the domain of several different disciplines including visual designers, illustrators, content and copy designers, and both user interface ("UI") and UX designers, along with creative directors and project managers tasked with pulling together these diverse elements. While the design process can vary significantly from one organization to another, a few generally accepted practices and processes have emerged.

For example, user-centered design seeks to develop software and software interfaces that are focused around end-user goals, needs, wants, and constraints. This methodology depends on learning as much about the end-user as is necessary to create the best user experience for the specific software product. The process begins with a UX researcher who may create ethnographic and field studies, interviews, surveys, heuristic evaluations,[169] user tests, and related methods to generate data regarding user requirements, pain points, and expectations. This data forms the basis for the creation of narratives (known as use cases or use scenarios) that help drive software engineering requirements, which are then incorporated into the overall development plan.[170] Beyond this initial stage, the typical software

(describing the ceremony in which Apple won an award for being the "best design studio of the last 50 years").

166. *See* Alma Whitten & J. D. Tygar, *Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0*, *in* SECURITY AND USABILITY, *supra* note 94, at 669 (arguing that effective computer security measures require enhanced and particularized user interface standards).

167. WORDPRESS, http://wordpress.org (last visited Mar. 17, 2013).

168. DRUPAL, http://drupal.org (last visited Mar. 17, 2013).

169. *See* Jakob Nielsen, *How to Conduct a Heuristic Evaluation*, NIELSEN NORMAN GROUP (Jan. 1, 1995), http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation.

170. For background on the traditional software engineering process, see generally JAKOB NIELSEN, USABILITY ENGINEERING (1993).

development process is iterative, alternating between testing, design tweaks, and coding changes, the extent of which depends on the complexity of the project. In the simplest case, a software developer receives requirements from the design team, devises an architecture and system design, and then relies on this design to develop the software in a matter of days. See *supra*, Figures 1 and 2. In more complex cases, several iterations and multiple departments and stakeholders contribute to a project that may take months or even years to complete.

UX design is growing in importance as a design discipline, with almost every major software company incorporating it into their product design process.[171] Recently cited examples of privacy by design such as the Google+ user interface were at least partially the result of work done by UX designers examining socialization in the "real world."[172] Indeed, privacy by design—insofar as it seeks to anticipate and address potential privacy problems that customers may have in using any product—may be understood as an extension of existing UX design. However, making this a reality would require adjusting current user research protocols to include probes on the relationship of privacy and consumer expectations. This is not always a straightforward task.[173] In many cases, privacy is a latent concern, and it may be difficult to recognize without additional training or awareness. However, it should not be an insurmountable task for UX designers to develop a better

171. *See, e.g.*, Aaron Marcus, *The ROI of Usability*, USABILITY PROF'LS' ASS'N, http://www.upassoc.org/usability_resources/usability_in_the_real_world/roi_of_usability.html (last visited Apr. 10, 2013) (illustrating the added value for companies implementing usability principles and practices).

172. *See* Paul Adams, The Real Life Social Network, Remarks at the Voices that Matter: Web Design Conference (June 29, 2010) (presentation available at http://www.slideshare.net/padday/the-real-life-social-network-v2) (describing research in social engagement in the offline world and how this maps to online social networks).

173. Research in behavioral economics reveals the highly contextual and nuanced nature of consumer decisions and how stated preferences related to privacy are not necessarily adhered to in practice. *See generally* Acquisti & Grossklags, *Privacy Attitudes and Privacy Behavior*, *supra* note 72; *see also* Somini Sengupta, *Letting Down Our Guard*, N.Y. TIMES, Mar. 31, 2013, at BU1 (describing Alessandro Acquisti's research).

sense of privacy issues by extending existing concepts.[174] Indeed, Lederer et al. and other HCI experts have already have begun doing this.[175]

As institutional knowledge of privacy expectations and related social norms develop, it is likely that UX practitioners will get better at recognizing and incorporating them into their user research protocols. Initially, they may have to rely on trial and error to determine what interpretations of privacy values work best for UX privacy research. User experience design was founded on work from pioneers such as Jakob Nielsen and has strong roots in academia that persist to this day. According to Nielsen, usability is a "quality attribute" for determining the ease-of-use of any user interface.[176] Usability in the privacy (and security) domains has additional, unique aspects. First, users consider usable privacy and security controls secondary to completing some primary task (like searching the Internet), so they must be accessible without getting in the way; second, they must accommodate a broad range of users with different skill levels and not only be designed for technical elites; and, third, if sophisticated security and privacy systems lack usability, they may put users at a higher risk than less sophisticated, but more easily used systems. The increased risk of error provides an even greater incentive to ensure that privacy and security are more usable than in many other domains.[177]

Usability may be studied throughout the several stages of the software development cycle (requirements, design, release), but it is a large topic and well beyond the scope of this Article.[178] Here, we focus on a narrow slice of the relevant literature that is directly concerned with the interface design aspects of social networks generally, their implications for privacy, and the usability of privacy features in Google+ and Facebook.

---

174. Many examples of UX guidelines exist. *See, e.g.*, UI WIZARDS, http://www.uiwizards.com (last visited Feb. 27, 2013) (offering services, classes, and books for designing user interfaces); Steve Krug, ADVANCED COMMON SENSE, http://www.sensible.com (last visited Feb. 27, 2013). Many companies have their own guidelines as well. *See, e.g.*, *What Makes a Design "Googley"?*, GOOGLE OFFICIAL BLOG (Apr. 23, 2008), googleblog.blogspot.com/2008/04/what-makes-design-googley.html (listing Google's ten design guidelines for interfaces).

175. *See infra* notes 183–202 and accompanying text.

176. *See* Jakob Nielsen, *Usability 101: Introduction to Usability*, NIELSEN NORMAN GROUP, http://www.nngroup.com/articles/usability-101-introduction-to-usability (last visited Feb. 27, 2013).

177. *See* Claire-Marie Karat et al., *Usability Design and Evaluation for Privacy and Security Solutions*, *in* SECURITY AND USABILITY, *supra* note 94, at 47, 48–50.

178. For key reference books in the field, see SECURITY AND USABILITY, *supra* note 94, at 49. For an overview of HCI as it relates to privacy, see Mark S. Ackerman & Scott D. Mainwaring, *Privacy Issues and Human-Computer Interaction*, *in* SECURITY AND USABILITY, *supra* note 94, at 381.

In reviewing the strengths and weaknesses of FIPs, we previously argued that the control paradigm on which FIPs are based has limited relevance to the social dynamics of privacy.[179] This line of thinking is borne out by the fact that when usability experts analyze the privacy implications of user interfaces, they do not turn to FIPs as a source of understanding.[180] Rather, they rely on the writings of Irwin Altman—a social psychologist who studied personal space and territoriality and conceptualized privacy as a dynamic process of negotiating personal boundaries in intersubjective relationships[181]—and Helen Nissenbaum—a philosopher of technology, who understands privacy in terms of norms governing distinct social contexts, a framework that she refers to as "contextual integrity."[182] Both reject the view that privacy is solely concerned with control over personal information or that the notion of "privacy in public" is somehow an oxymoron. We will briefly describe Altman's views and their influence on two important essays on the design implications of understanding privacy as a dynamic process.[183] We then turn to a group of researchers who have sought to analyze and suggest remedies for interface design flaws in Facebook by reference to both Altman's work and Nissenbaum's contextual integrity framework.[184]

b) Altman

Altman views privacy as an "interpersonal boundary process" by which individuals become more or less accessible and open to others through a variety of behavioral mechanisms.[185] These include verbal and para-verbal behavior (what we say and how we say it—i.e., tone, intensity, pitch, and inflection of voice), personal spacing (distance and angle of orientation to

179. *See supra* Section II.A.

180. *See* Benjamin Brunk, *A User-Centric Privacy Space Framework*, *in* SECURITY AND USABILITY, *supra* note 94, at 401, 407 (explaining that the "primary drawback" of using FIPs for understanding user experiences and privacy solutions is "that none of them are particularly user centered").

181. IRWIN ALTMAN, THE ENVIRONMENT AND SOCIAL BEHAVIOR: PRIVACY, PERSONAL SPACE, TERRITORY, AND CROWDING (1975).

182. *See* HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE (2010) [hereinafter NISSENBAUM, PRIVACY IN CONTEXT]; Helen Nissenbaum, *Privacy as Contextual Integrity*, 79 WASH. L. REV. 101 (2004) [hereinafter Nissenbaum, *Privacy as Contextual Integrity*].

183. *See* Scott Lederer et al., *Personal Privacy Through Understanding and Action: Five Pitfalls for Designers*, 8 PERS. & UBIQUITOUS COMPUTING 440 (2004); Leysia Palen & Paul Dourish, *Unpacking "Privacy" for a Networked World*, 5, *in* CHI 2003: NEW HORIZONS: PROCEEDINGS OF THE SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS 129 (2003), *available at* http://dl.acm.org/citation.cfm?id=642635.

184. *See* notes 218–24 and accompanying text.

185. ALTMAN, *supra* note 181, at 6.

others), and additional forms of non-verbal behavior such as facial expression and body language, territorial behavior (i.e., use, possession, and ownership of places or objects), and cultural norms that regulate contact with others.[186] For example, if we are conducting an intimate conversation in public, we achieve the desired level of privacy by relying on familiar mechanisms such as speaking softly to make the conversation inaudible to others, standing with our backs to the crowd, and avoiding eye contact with anyone who might approach. Analogous mechanisms are generally lacking in online settings, even in SNSs, despite the fact they are all about social interactions.

Clearly, Altman rejects traditional views of privacy as a form of social withdrawal that goes on only in "private" spaces.[187] Rather, privacy is a process that is dynamic (i.e., shaped by personal and collective experiences and expectations),[188] dialectical (i.e., informed by a continuous balancing act over what to disclose or conceal),[189] and optimizing.[190] Finally, privacy is less a matter of an individual's unilateral control over the disclosure of information than it is a bidirectional process, involving control over both inputs from others (being looked at, approached, or called on the telephone) and outputs to others (staring, seeking out friends, initiating a telephone call).[191] In short, privacy is a process of regulating the boundaries by which people make themselves more or less accessible and open to others.[192]

Altman is primarily concerned with how people manage face-to-face interactions occurring in physical space and mediated by the environment in which we live.[193] Building on Altman's work, Palen and Dourish explain how

186. *Id.* at 33–42.

187. In this stance, Altman anticipates Nissenbaum's rethinking of the "public-private" distinction. *See id.*, NISSENBAUM, PRIVACY IN CONTEXT, *supra* note 182, at 113–25.

188. ALTMAN, *supra* note 181, at 43–45.

189. *Id.* at 11 (noting that such balancing depends not only on the time and circumstances but also on individual character and group cultural norms).

190. Altman notes that at any given moment, individuals may achieve more privacy than they desire (resulting in boredom, loneliness, or social isolation), less privacy than they desire (resulting in feelings of crowding, intrusion, or invasion), or the optimum level, where the achieved level of interaction and contact with others matches the desired level. *Id.* at 25–27.

191. *Id.* at 27–28.

192. *Id.* at 10. Goffman offers a similar analysis of the complex behavioral decisions people engage in when deciding whether to release or withhold specific information to a given person at a given time depending on their view of themselves, the current situation, and the consequences of disclosure. *See generally* ERVING GOFFMAN, RELATIONS IN PUBLIC: MICROSTUDIES OF THE PUBLIC ORDER (1972); ERVING GOFFMAN, THE PRESENTATION OF SELF IN EVERYDAY LIFE (1959).

193. Sociologist Christena Nippert-Eng has extended Altman's work to privacy management in a variety of everyday settings and tasks such as visiting the beach; keeping

privacy as a boundary process works in a networked world mediated by information technology, which simultaneously enables social interaction with large, distant, and even unknown audiences, but also eliminates most of the familiar physical, psychological, and social cues we rely on to manage our interpersonal relationships.[194] Whether we know it or not, every time we "go online," we disclose information about ourselves.[195] Common activities like web surfing or searching create data trails that are collected, aggregated, and analyzed, often without our knowledge or consent.[196] SNSs introduce new opportunities for social interaction and sharing but offer very limited means to convey demeanor and intent or otherwise establish the context of self-expression.[197]

Privacy management in a networked world therefore involves "combinations of social and technical arrangements that reflect, reproduce, and engender social expectations, guide the interpretability of action, and evolve as both technologies and social practices change."[198] These new privacy mechanisms should enable individuals to intervene in the flows of existing data about them in relation to others and to renegotiate the boundaries of disclosure, identity, and temporality.[199]

What, then, are the tools that support both strategic concealment and revelation of data in the various contexts of our networked lives? Lederer et al. suggest improving privacy practices in technical systems through a combination of understanding and action, and offer design guidelines in the form of "five pitfalls" for designers to avoid.[200] They are:

---

and revealing secrets; assembling and categorizing the contents of one's wallet or purse; managing the receipt of email, cell phone, and other interruptions; and controlling accessibility at the "porous perimeters" of one's home (windows, doors, yards, and curbs). CHRISTENA NIPPERT-ENG, ISLANDS OF PRIVACY 2–3 (2010) (defining privacy as "selective concealment and disclosure" and as a daily activity of trying to "deny or grant varying amounts of access to our private matters to specific people in specific ways").

194. *See* Palen & Dourish, *supra* note 183.

195. *Id.* at 131–32.

196. *Id.*

197. *Id.* at 132 (highlighting specifically the concern of friends sharing photographs online without consent or input from the subjects).

198. *Id.* at 133.

199. *Id.*; *see also* danah boyd, *Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications*, *in* NETWORKED SELF: IDENTITY, COMMUNITY, AND CULTURE ON SOCIAL NETWORK SITES 49 (Zizi Papacharissi ed., 2010) (describing three dynamics that play a role in shaping what she calls "networked publics": invisible audiences, collapsed contexts, and the blurring of public and private).

200. *See* Lederer et al., *supra* note 183, at 445–49.

1. Designs should not obscure potential information flow (because informed use of a system requires that users understand the scope of its privacy implications);
2. Designs should not conceal actual information flow (because users need to understand what information is being disclosed to whom);
3. Designs should not require excessive configuration to manage privacy but rather should enable users to practice privacy as a natural consequence of their normal engagement with the system;
4. Designs should not forgo an obvious, coarse-grain mechanism for halting and resuming disclosure; and
5. Designs should not inhibit users from transferring established social practice to emerging technologies.[201]

According to the authors, tools that combine such feedback (understanding) and control (action) mechanisms "make their consequences known and do not require great effort to use," resulting in socially meaningful privacy practices.[202]

c) Nissenbaum

Nissenbaum's theory of privacy as contextual integrity begins with the observation that norms govern the flow of information in highly specific social contexts. Familiar social contexts include health care, education, employment, religion, family, and the commercial marketplace.[203] Each of these contexts may be more fully understood in terms of the roles people play within them (e.g., doctor, nurse, patient), the activities and practices they engage in within such roles (e.g., asking about symptoms, administering medicines, describing an ailment), the norms that define acceptable and unacceptable behaviors within a given context (e.g., respecting patient privacy), and the values around which activities in a context are defined (e.g., prescribing medicine for the good of the patient or applying measures that benefit the sick while avoiding overtreatment).[204] People move from one context to another throughout the day, and they implicitly understand what norms apply and act accordingly. For example, we expect a physician treating us for hepatitis to inquire about our consumption of alcohol and drugs but not to share this information with our employer. And we share our joys and anxieties with spouses or partners but not with the clerk at the convenience

201. *Id.* at 441.
202. *Id.* at 450.
203. NISSENBAUM, PRIVACY IN CONTEXT, *supra* note 182, at 129–30.
204. *Id.* at 132–37.

store. Each of these everyday social contexts has distinctive sets of rules governing information flows. These informational norms define contextual integrity, which is preserved when informational norms are respected, and violated when they are breached.

Nissenbaum posits two fundamental types of informational norms: appropriateness and distribution.[205] The former prescribe what personal data is (or is not) allowable, expected, or even required to be revealed in a given context.[206] These norms vary greatly and may be more or less restrictive, explicit, or complete. But the key point is that there is no area of life not governed by some informational norms.[207] The latter prescribe how and with whom data may be shared in a given context.[208] Distributional norms are also highly variable and rather complex.[209] For example, information sharing between friends is bidirectional but they expect that what they say to each other will be held in confidence and not arbitrarily spread to others. In contrast, information flows in only one direction in the doctor-patient relationship; doctors expect (and even may demand) that patients reveal their physical and/or mental condition, while patients expect that what they say is confidential, subject to exceptions when a disease poses a public health risk.[210]

Nissenbaum proposes contextual integrity as a "benchmark" of privacy insofar as in any given situation, a privacy violation may be understood as a violation of informational norms.[211] Her work thus sheds much light on recent privacy controversies associated with new information technologies and systems. In a nutshell, information technologies worry and alarm us—and in more extreme cases result in privacy incidents—when they "flout entrenched informational norms and hence threaten contextual integrity."[212] In her later work, Nissenbaum argues more generally that information norms are characterized by four parameters—context, actors, attributes, and transmission principles—which are also key parameters for determining whether a new practice resulting from the deployment of a novel technical device or system violates contextual integrity, for example, photo tagging on

205. Nissenbaum, *Privacy as Contextual Integrity*, *supra* note 182, at 138.
206. *Id.* at 137–40.
207. *Id.*
208. *Id.* at 140–43.
209. *Id.*
210. *Id.*
211. NISSENBAUM, PRIVACY IN CONTEXT, *supra* note 182, at 140; *see also* Nissenbaum, *Privacy as Contextual Integrity*, *supra* note 182, at 138 (referring specifically to violations of norms of appropriateness and distribution).
212. NISSENBAUM, PRIVACY IN CONTEXT, *supra* note 182, at 127.

a SNS.[213] She offers a "decision heuristic" for detecting such violations, which involves five steps:[214] establish the prevailing context (e.g., junior high school students sharing photos by posting them to Facebook); identify key actors (e.g., parents, minor children, friends, friends of friends, Facebook); analyze whether the novel technology (SNSs) affect the types of information transmitted (e.g., not only visual images but links to status updates, photos, tags, etc.);[215] establish whether transmission principles have changed (e.g., school children sharing a racy photograph with each other in a very guarded fashion versus one of them uploading the photo to a SNS and tagging their friends, thereby sharing and distributing the now tagged photo to a potentially large audience of classmates, teachers, parents, and all of their social networks); and "flag" violations.[216]

For Nissenbaum, contextual integrity is not only a sound benchmark for describing and predicting how people respond to privacy violations but also a prescriptive guide. We will not explore the moral elements of Nissenbaum's theory[217] but instead focus on how the framework of contextual integrity assists firms seeking to design new systems that avoid privacy incidents. For example, Heather Richter Lipford builds on Nissenbaum's work to propose

---

213. *Id.* at 140–47.

214. NISSENBAUM, PRIVACY IN CONTEXT, *supra* note 182, at 148–50.

215. As Facebook explains:

> **Tag people in your posts:** Add tags to anything you post, including photos and updates. Tags can point to your friends or anyone else on Facebook. Adding a tag creates a link that people can follow to learn more.
>
> **Tell people about stuff they're in:** Adding tags can let people know when they're in photos or other things you share. People you tag can receive a notification so they can see your post. The post may also go on the person's profile and appear in their friends' news feeds.
>
> **Help tag things other people missed:** You can tag other people's photos and posts to help them add details. Your name appears with the tag, so it's always clear where it came from.

How Tagging Works, FACEBOOK, https://www.facebook.com/about/tagging (last visited Feb. 27, 2013). For further discussion of tagging on Facebook, see *supra* Section III.B.4.

216. On the topic of "red flags," Nissenbaum says, "[i]f the new practice generates changes in actors, attributes, or transmission principles, the practice is flagged as violating entrenched informational norms and constitute[s] a prima facie violation of contextual integrity." NISSENBAUM, PRIVACY IN CONTEXT, *supra* note 182, at 150.

217. The moral element goes beyond assessing information flows in terms of entrenched norms and instead asks whether "new practices are demonstrably more effective at achieving contextual values, ends, and purposes." *Id.* at 179. This addition of a moral element results in what Nissenbaum refers to as the "augmented" decision heuristic. *Id.* at 181–82.

two specific interface modifications for managing privacy issues on Facebook: Restrict Others[218] and AudienceView.[219]

As discussed more fully below,[220] photo sharing on Facebook reduces the photo subject's control over her image and its distribution, resulting in now-commonplace stories of embarrassment, humiliation, discrimination, and even arrest.[221] Restrict Others relies on the analytic frameworks of Nissenbaum (and Altman) to develop a new tool for enhancing a user's ability to control who sees photos uploaded and tagged by users other than the subject of the photo.[222] Quite simply, "[i]t works by allowing tagged users to send a request to the owner [the person who uploaded the photo] asking that a photo be hidden from certain people."[223] When Lipford et al. first proposed AudienceView in 2008, the then-current Facebook interface offered limited visual feedback regarding a user's audience and a "poor mental model" of how changes in the privacy settings affect the sharing of profile data with different audiences.[224] As a result, many users unwittingly revealed profile data more broadly than they intended.[225] While Facebook enabled users to configure their profile privacy settings in advance through the usual "wall of checkboxes,"[226] the default privacy settings were permissive and users rarely changed them. AudienceView offers a modified interface:

> [U]sers view pages of their profiles from the point of view of different audiences, such as their different groups of friends, networks, public, etc. This interface provides a more visual and accurate mental model of what different people can view of them, allowing users to more explicitly and concretely consider the

218. *See* Andrew Besmer & Heather Richter Lipford, *Moving Beyond Untagging: Photo Privacy in a Tagged World*, CHI 2010: PRIVACY: PROC. SIGCHI CONF. ON HUM. FACTORS COMPUTING SYS. 1563 (2010), *available at* http://dl.acm.org/citation.cfm?id=1753560&bnc=1.

219. *See* Heather Richter Lipford et al., *Visible Flows: Contextual Integrity and the Design of Privacy Mechanisms on Social Network Sites*, *in* PROC. 12TH IEEE INT'L CONF. ON COMPUTATIONAL SCI. & ENGINEERING 985 (2009), *available at* http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5283751.

220. *See infra* notes 399–403 and accompanying text.

221. *See* NISSENBAUM, PRIVACY IN CONTEXT, *supra* note 182, at 59–61 (citing various incidents).

222. Besmer & Lipford, *supra* note 218, at 1567.

223. *Id.*

224. *See* Heather Richter Lipford et al., *Understanding Privacy Settings in Facebook with an Audience View*, UPSEC '08: PROC. FIRST CONF. ON USABILITY, PSYCHOL., & SECURITY, art. 2 (Elizabeth Churchill & Rachna Dhamija eds., 2008), *available at* http://static.usenix.org/event/upsec08/tech/full_papers/lipford/lipford.pdf (describing the relationship between user privacy concerns and the shortcomings in Facebook's user interface).

225. *Id.* at 2.

226. *See* Nathan Good, *The Deadly Sins of Security User Interfaces*, *in* THE DEATH OF THE INTERNET 290, 300–02 (Markus Jakobsson ed., 2012).

> context of their information and adjust that information flow as desired.[227]

To remind users of information flows as they view friends' profiles and post information, Lipford et al. proposed a "message box on the user's home page . . . show[ing] the most recent information access[ed by others], . . . or summariz[ing] the number of accesses in a certain time period."[228] And to prevent the "flattening" of a user's social contexts and ensure that Facebook privacy settings better reflect the more nuanced and varied social contexts of offline relationships, they suggested that Facebook automatically determine a user's social spheres by analyzing every user's "social network graph" and that it make this information available to users through the AudienceView interface.[229]

Finally, Lipford et al. offer six design guidelines based on the contextual integrity framework for making information flows more visible in SNSs.[230] They are as follows:

1. Make information flows more transparent, so that users know what information they are sharing and with whom;
2. Increase user awareness of information flows as they make decisions about sharing profile data, photos, and the like, both with other users and/or third parties;
3. Increase user awareness of how much information is archived and still available to others;
4. Make information and context concrete by providing specific examples of who will see what;
5. Provide more granular controls over information flows; and
6. Do not abruptly modify the flow of information.[231]

This completes our brief overview of the design process, the multiple factors that influence design, and the engineering and usability principles that firms must follow in order to translate a robust understanding of FIPs into well-engineered and highly usable privacy designs. We began by distinguishing back-end implementation and security protections from front-

227. *See* Lipford et al., *supra* note 219, at 987.

228. *Id.* at 988.

229. *Id.* at 987–88. For a related idea, see Lauren Gelman, *Privacy, Free Speech, and "Blurry-Edged" Social Networks*, 50 B.C. L. REV. 1315, 1342–44 (2009) (proposing a "tool for users to express . . . privacy preferences over uploaded content . . . by tagging any uploaded content with [a visual] icon" and perhaps machine-readable metadata).

230. *See* Lipford et al., *supra* note 219, at 987. The guidelines are highly reminiscent of the five "pitfalls" of Lederer et al. *See supra* note 200 and accompanying text.

231. Lipford et al., *supra* note 219, at 987.

end design issues centered on users' privacy expectations and, hence, on the disciplines of Human-Computer Interactions ("HCI"), including both user-interface ("UI") and user-experience ("UX") design. Next, relying mainly on the works of Feigenbaum, et al., and Spiekermann and Cranor, we identified a short list of FIPs-based engineering principles including data avoidance and minimization, data retention limits, notice, choice, access, and accountability. Finally, relying on Grimmelmann's idea of the "social dynamics" of privacy, we fleshed out the UX approach to privacy design by discussing "five pitfalls" for designers to avoid as identified by Lederer et al., and six design principles of Lipford et al., which we traced back, respectively, to the pioneering work of Altman and Nissenbaum. With the completion of this preliminary work, we may now take up the counterfactual analysis, which applies these privacy engineering and design principles to the ten case studies.

III. CASE STUDIES AND COUNTERFACTUAL ANALYSES

A. GOOGLE

Google manages the most popular Internet search engine,[232] which generates revenue when users click or view advertising related to their searches.[233] Advertising revenue accounted for about 95% of Google's approximately $46 billion in annual revenues in 2012.[234] The company has a long history with privacy issues, and we review four major Google services—Gmail, Search, Street View, Buzz (and its successor, Google+)—as well as its controversial new revisions to the company's privacy policies.

1. Gmail

Gmail is Google's free, web-based and advertising-supported email service.[235] When launched in early 2004 as an invitation-only beta release, it was an immediate success, offering users unprecedented storage capacity in exchange for receiving contextual ads.[236] Gmail's ad engine automatically scans header information and the content of incoming and outgoing

232. *See* Press Release, comScore, comScore Releases January 2013 U.S. Search Engine Rankings (Feb. 13, 2013), http://www.comscore.com/Insights/Press_Releases/2013/2/comScore_Releases_January_2013_U.S._Search_Engine_Rankings (showing Google accounts for two-thirds of U.S. web searches).

233. *See Investor Relations: 2012 Financial Tables*, GOOGLE, http://investor.google.com/financial/tables.html (last visited Mar. 8, 2013).

234. *Id.*

235. *See Google Gets the Message, Launches Gmail*, NEWS FROM GOOGLE (Apr. 1, 2004), http://googlepress.blogspot.com/2004/04/google-gets-message-launches-gmail.html.

236. *Id.*

messages for key words provided by advertisers in advance.[237] Despite this privacy-sensitive design, Google's decision to fund free storage by serving contextual ads proved quite controversial: users and consumer advocacy groups raised concerns over the lack of consent by non-subscribers, the impact of storage capacity on data retention (and hence government requests for data), and the prospect of Google someday modifying its approach and creating highly detailed user profiles based on the correlation of users' Gmail identities with their Google search behavior.[238] Despite numerous government investigations, no adverse actions were taken, and the controversy gradually faded without forcing any major change in Gmail's handling of ads.[239] Gmail's success allowed Google to branch out into new products and create a more individualized relationship with hundreds of million users, managing their email accounts and contact lists, and laying the foundation for its later foray into social networking.[240]

Gmail is a design success: it offered users a very clear value proposition, served ads while avoiding both profiling and disclosure of PII to advertisers, and did a reasonably thorough job of ensuring informed consent.[241] And yet many greeted Gmail with anxiety and suspicion, which persisted notwithstanding its design strengths.[242] Prior to Gmail, email was conceived of as a personal and inviolate form of communication between a sender and

237. *See Gmail Privacy FAQ: How Does Google's "Content Extraction" Work?*, EPIC, http://epic.org/privacy/gmail/faq.html (last visited Mar. 8, 2013).

238. *See* Stefanie Olsen, *Google's Web Mail No Joke*, CNET (Apr. 2, 2004), http://news.cnet.com/2100-1032-5184090.html (describing how government could subpoena archived information or how Google could mine the data); Donna Wentworth, *Gmail: What's the Deal?*, ELEC. FRONTIER FOUND. (Apr. 5, 2004), https://www.eff.org/deeplinks/2004/04/Gmail-whats-deal (listing various news stories regarding concerns about Gmail).

239. *See* Jane Perrone, *Google Free Email Faces Legal Challenge*, THE GUARDIAN (Apr. 12, 2004), http://www.guardian.co.uk/technology/2004/apr/13/internationalnews.onlinesupplement.

240. *See* Dante D'Orazio, *Gmail Now Has 425 Million Total Users*, THE VERGE (June 28, 2012), http://www.theverge.com/2012/6/28/3123643/gmail-425-million-total-users; Robert Charette, *Google Expands In Two More Directions: Social Media and Broadband Service*, IEEE SPECTRUM: RISK FACTOR BLOG (Feb. 11, 2010), http://spectrum.ieee.org/riskfactor/telecom/internet/google-expands-in-two-more-directions-social-media-and-broadband-service.

241. *See* Friedman et al., *Informed Consent by Design*, *supra* note 155, at 521–26. Friedman et al., however, also raised two related privacy concerns regarding Gmail: (1) whether using a machine to read email violates a person's privacy expectations, and (2) whether email senders actually consent to automatic scanning. *Id.* at 524.

242. *See* Perrone, *supra* note 239.

a receiver; contextual ads disrupted these informational norms by treating a private communication as the basis for a commercial offer.[243]

Are there additional design steps that Google might have taken to allay users' privacy concerns? First, Google might have been more transparent about whether Gmail served ads related to the content of one's emails and also tracked users in other ways or shared information with other services for advertising or other purposes.[244] Second, Google might have translated these assurances into architectural choices by designing a web mail service that segregated and separated any personal data about subscribers or their message content from any data collected by other Google services.[245] Third, and more radically, Google might have considered a simultaneous release of both an ad-supported free web mail service and an ad-free paid version.[246] By providing consumers with a range of choices from the outset, Google could have facilitated "privacy by experimentation" and set a sound precedent for designing future services with privacy in mind.[247]

2. Search

Unlike Gmail, Google Search attracted more sustained interest from privacy officials. Beginning in the final months of 2006, European and U.S. regulators challenged Google and its search engine competitors regarding the amount, sensitivity, and retention periods of the data collected for search ads and other purposes.[248] Both consumer and regulatory concerns were spurred in part by two widely read news stories alerting the public to the data processing practices of their favorite search engines.[249] Over the next several

243. *See id.* ("[I]t's an absolute invasion of privacy. It's like having a massive billboard in the middle of your home." (quoting former California State Senator Liz Figueroa)).

244. The original Gmail privacy policy stated, "Google may send you information related to your Gmail account or other Google services." *Google Gmail Privacy Policy*, TOSBACK (Sep. 12, 2008), http://www.tosback.org/version.php?vid=1030.

245. *See supra* notes 116, 137 and accompanying text.

246. Google later provided options for paying customers of Google Apps for Business or Education to turn off ads for a given domain. *See Disable Advertisements*, GOOGLE, http://support.google.com/a/bin/answer.py?hl=en&answer=60758 (last visited July 23, 2012).

247. *See* Betsy Masiello, *@betsymas*, TWITTER (Aug. 19, 2010), https://twitter.com/betsymas/status/21615739700 ("two concepts from today: privacy by experimentation in contrast to privacy by design; data driven policy by design as a way fwd").

248. *See* Verne Kopytoff, *Google Comes Under Scrutiny*, SFGATE (May 30, 2007), http://www.sfgate.com/business/article/Google-comes-under-scrutiny-FTC-European-2590461.php.

249. In one, the DOJ subpoenaed millions of search records from leading firms and Google challenged the request in court, winning concessions on the scope of the final order. *See* Verne Kopytoff, *Google Says No to Data Demand*, SFGATE (Jan. 20, 2006), http://www.sfgate.com/news/article/Google-says-no-to-data-demand-Government-wants-

years, regulators and advocates called upon all search firms to offer greater transparency regarding their data practices, shorter data retention periods, and improved methods for anonymizing data after the retention period expired.[250] In response, Google, Yahoo!, and Microsoft shortened data retention periods, sought to improve anonymization techniques, and began developing new compliance mechanisms.[251] Soon, all of the major search firms were competing on privacy features for their search engine and browser offerings.[252] Despite this heated competition, Google remained the leading search engine and moved ahead with a $3.1 billion acquisition of DoubleClick, overcoming objections on both antitrust and privacy grounds.[253]

With Search, the public grew alarmed when it learned that leading search engines were tracking their searches and collecting and storing sufficient information to attract the attention of law enforcement agencies and to permit inquisitive journalists to discover their "real-life" identities.[254] Two interrelated design issues emerged: (1) how long search data should be retained before being deleted; and, (2) if it was anonymized instead of deleted, the proper method of anonymization. Google sought to achieve what it deemed the "right balance" between "privacy and other goals (like security, fraud prevention, and search improvements)" by retaining search logs for eighteen months and then "anonymizing" any data linking search terms to IP addresses by erasing the last octet of the IP address.[255] To be

2523692.php. In the other, AOL shared "anonymized" search logs with researchers but the released data enabled curious journalists to identify a sixty-two-year-old woman in Georgia. *See* Michael Barbaro & Tom Zeller Jr., *A Face is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES, Aug. 9, 2006, at A1.

250. *See, e.g.*, E.U. ARTICLE 29 DATA PROTECTION WORKING PARTY, OPINION 1/2008 ON DATA PROTECTION ISSUES RELATED TO SEARCH ENGINES (WP 148) (Apr. 2008) http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2008/wp148_en.pdf; FED. TRADE COMM'N, SELF-REGULATORY PRINCIPLES FOR ONLINE BEHAVIORAL ADVERTISING: TRACKING, TARGETING, AND TECHNOLOGY (Feb. 2009), http://www.ftc.gov/os/2009/02/P085400 behavadreport.pdf.

251. *See, e.g.*, Peter Fleischer, *Data Retention: The Right Balance Between Privacy and Security*, GOOGLE PUB. POL'Y BLOG (July 11, 2007), http://googlepublicpolicy.blogspot.com/2007/07/data-retention-right-balance-between.html.

252. *See* Gregg Keizer, *Mozilla Adds Privacy Mode to Firefox 3.1 Plans*, MACWORLD (Sept. 12, 2008), http://www.macworld.co.uk/macsoftware/news/?newsid=22767 (describing Mozilla's efforts to compete with Microsoft and Google by adding a private search mode to browser).

253. *See* Louise Story & Miguel Helft, *Google Buys an Online Ad Firm for $3.1 Billion*, N.Y. TIMES, Apr. 14, 2007, at C1.

254. *See* Barbaro & Zeller Jr., *supra* note 249.

255. *See* Fleischer, *supra* note 251. For criticisms of Google's anonymization method, see Soghoian, *supra* note 133. Other search engines have experimented with shorter periods. *See*

sure, there is a trade-off between retaining data to improve search results and maintain security, and deleting or anonymizing search data to protect user privacy. That being said, are there other steps Google might have taken to address privacy concerns? First, it might have assisted users in conducting searches or browsing the web anonymously, either by partnering with a proxy server or by integrating search with an onion-router proxy like Tor.[256] Alternatively, Google might have pursued a data minimization approach by managing internal access to users' IP addresses for uses beyond search quality and anti-fraud issues. Third, Google might have stepped up its transparency efforts with respect to its search practices. While Google's disclosures met or exceeded industry standards,[257] it neither explained the potential privacy harms of monitoring and tracking search queries nor stated explicitly whether it combined search query data with any other information it collected from Gmail and other services requiring account registration.[258] Finally, Google might have facilitated multiple online accounts early on, thereby allowing users to segment their lives and adjust their public personas in accordance with the social insights of Altman and Goffman.[259]

Chloe Albanesius, *Yahoo to Keep Your Search Data for 18 Months, Not Three*, PCMAG (Apr. 18, 2011), http://www.pcmag.com/article2/0,2817,2383711,00.asp (describing Yahoo!'s attempt at a three-month retention period before it was forced to retreat to 18 months due to search quality issues); Michael Zimmer, *Microsoft to Delete IP Addresses from Bing Search Logs after 6 Months*, MICHAELZIMMER.ORG (Jan. 19, 2010), http://www.michaelzimmer.org/2010/01/19/microsoft-to-delete-ip-addresses-from-bing-search-logs-after-6-months.

256. An onion router repeatedly encrypts and forwards requests through a chain of proxies and, like a layer of an onion, removes a layer of the encryption to determine the next destination. The end effect is that the traffic cannot be traced back to its original source, unlike common network connections. *See* TOR, https://www.torproject.org (last visited Mar. 8, 2013); *see also supra* note 108 and accompanying text.

257. For example, in 2007, Google launched an innovative series of short videos to explain basic privacy concepts including search privacy. *See* Peter Fleischer, *Google Search Privacy: Plain and Simple*, GOOGLE OFFICIAL BLOG (Aug. 8, 2007), http://googleblog.blogspot.com/2007/08/google-search-privacy-plain-and-simple.html.

258. In 2010, Google launched a tool for reporting on government requests for user data. Dorothy Chou, *Transparency Report: Government Requests on the Rise*, GOOGLE OFFICIAL BLOG (Nov. 13, 2012), http://googleblog.blogspot.com/2012/11/transparency-report-government-requests.html; *see Google Transparency Report*, GOOGLE, http://www.google.com/transparencyreport/ (last visited July 23, 2012).

259. *See supra* Section II.B.3.b; *supra* note 192. Privacy advocates maintain that allowing users multiple accounts and identities allows users to engage in more natural online interactions and to adjust their privacy requirements as needed. *See, e.g.*, danah boyd, *Why Youth (Heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life*, *in* YOUTH, IDENTITY, AND DIGITAL MEDIA 119, 131–34 (David Buckingham ed., 2008) (describing teenagers' use of multiple accounts on MySpace, each tailored for different audiences); *Anonymity*, ELEC. FRONTIER FOUND., https://www.eff.org/issues/anonymity (last visited Apr. 11, 2013) (describing the importance and need for protection of anonymous speech on

3. Google Street View

Street View presents a more complex privacy scenario than either Gmail or Search. Launched initially in the United States in May 2007, Street View is an adjunct to Google Maps.[260] It displays panoramic images of many cities, which are photographed from cars equipped with specially adapted digital cameras and antennas.[261] Advocates and regulators objected early on to Google's collection and display of identifiable faces and license plates in conjunction with buildings or street scenes that might associate individuals with embarrassing or sensitive activities or locations (e.g., sunbathing in the nude or leaving a strip club).[262]

At first, Google defended its actions in the United States by arguing that all of the images were taken on public streets, where expectations of privacy are minimal.[263] Over time, Google improved its procedures for removing objectionable images and adopted digital "pixelation" technology (i.e., facial blurring) on a worldwide basis.[264] And yet Street View continued to be closely scrutinized in many jurisdictions where privacy laws prohibited the publication of images of people without their explicit consent, or local norms treated residential streets as part of one's private space.[265] Although privacy officials in several countries opened Street View investigations and ordered Google to stop taking pictures until these were completed, Google seemed on a path to resolve most of these matters.[266] Then, in late April 2010, Google revealed that its Street View cars had been inadvertently collecting "payload data" from Wi-Fi networks ("payload data" refers to information

the Internet). While consumers have regularly used multiple online identities to segment their lives, Google has recently made it easier for consumers to do so. *See Manage Multiple Users on Chrome*, CHROME HELP, https://support.google.com/chrome/bin/answer.py?hl=en&answer=2364824 (last visited July 23, 2012).

260. *See* Josh Lowensohn, *Google Launches Street View, Mapplets*, CNET (May 29, 2007), http://news.cnet.com/8301-17939_109-9723263-2.html.

261. SIVA VAIDHYANATHAN, THE GOOGLIZATION OF EVERYTHING (AND WHY WE SHOULD WORRY) 98 (2011).

262. *See* Elinor Mills, *Cameras Everywhere, Even in Online Maps*, CNET (May 30, 2007), http://news.cnet.com/Cameras-everywhere,-even-in-online-maps/2100-1038_3-6187556.html.

263. Peter Fleischer, *Street View and Privacy*, GOOGLE MAPS BLOG (Sept. 24, 2007), http://google-latlong.blogspot.com/2007/09/street-view-and-privacy.html ("There's an important public policy debate in every country around what privacy means in public spaces. That balance will vary from country to country, and Street View will respect it.").

264. *See* VAIDHYANATHAN, *supra* note 261, at 98–107.

265. *Id.* at 102–03 (noting that Google was forced to reshoot its Street View photos in Japan with cameras mounted lower to avoid peering over hedges and fences).

266. *See id.* at 111 ("The vast majority of those who use Google find Street View more beneficial . . . than harmful. . . . For every person who complains about Street View, millions more find it useful.").

sent over unprotected networks that includes locational data, passwords, email address, and contents of communications).[267] Public censure, private lawsuits, and dozens of new investigations quickly followed.[268] Google responded by intensifying its efforts to address these new (and old) concerns but with mixed results.[269] Additionally, the Federal Communications Commission ("FCC") fined Google for obstructing its inquiry into the company's collection of Wi-Fi payload data.[270] According to new details that emerged upon publication of a full version of the FCC report, data collection "was neither a mistake nor the work of a rogue engineer, as the company long maintained, but a program that supervisors knew about."[271] As a result, a number of regulators are considering whether to reopen their investigations.[272] Meanwhile, in Switzerland, an appeals court issued a mixed ruling, concluding that a ninety-nine percent accuracy rate in Google's blurring technology was acceptable, yet still upholding several conditions demanded by the privacy commissioner.[273]

Although many commentators take it for granted that it was highly invasive for Street View to publish images of people at specific geographical

267. *See* Alan Eustace, *Wifi Data Collection: An Update*, GOOGLE OFFICIAL BLOG (May 14, 2010) (updated June 9, 2010), http://googleblog.blogspot.com/2010/05/wifi-data-collection-update.html.

268. *See* Joshua Keating, *Google's Most Controversial Feature*, FOREIGN POLICY (Aug. 10, 2010), http://blog.foreignpolicy.com/posts/2010/08/10/googles_most_controversial_feature (noting that "nearly half of the 60 legal or criminal investigations being faced by Google are related to Street View").

269. In a 2009 agreement with Germany, Google agreed to let property owners opt out of Street View before it was activated. *See* Kevin J. O'Brien, *Many Germans Opt Out of Google's Street View*, N.Y. TIMES, Oct. 15, 2010, http://www.nytimes.com/2010/10/16/technology/16streetview.html. In contrast, Google failed to reach an agreement with Switzerland—which insisted that Google's pixelation technology achieve a 100 percent success rate—forcing Google to litigate. *See* Kevin J. O'Brien & David Streitfeld, *Swiss Court Allows Google Street View*, N.Y. TIMES, June 9, 2012, at B2.

270. *See* David Streitfeld, *Google Engineer Told Others of Data Collection, Full Version of F.C.C. Report Reveals*, N.Y. TIMES, Apr. 29, 2012, at A22.

271. *Id.* A Google privacy official later stated that the documents Google released in connection with the investigation "show some isolated references to payload collection early in the project that could have been seen as red flags by the recipients with the benefit of hindsight. But in context, at the time, the red flags were missed or not understood." *See* Letter from Peter Fleischer, Global Privacy Counsel, Google, to Steve Eckersley, Head of Enforcement, UK ICO (June 18, 2012), *available at* http://www.telegraph.co.uk/technology/google/9339113/Google-snooping-investigation-response-in-full.html.

272. *See* Kevin J. O'Brien, *Rethinking an Inquiry of Google*, N.Y. TIMES, May 3, 2012, at B1.

273. *See* O'Brien & Streitfeld, *supra* note 269 (explaining that the conditions require Google "to lower the height of its Street View cameras so they would not peer over garden walls and hedges, to completely blur out sensitive facilities like women's shelters, prisons, retirement homes and schools, and to advise communities in advance of scheduled tapings").

locations, the social norms governing public images differ cross-culturally and remain somewhat unsettled.[274] This is especially true in the United States, where people became accustomed to Street View with little difficulty.[275] Google acknowledged these cultural differences when it released Street View overseas with additional privacy protections (blurring faces and license plates) and made local adjustments in Japan[276] and Germany.[277] Despite these efforts, many users and foreign governments strongly objected to Google's failure to provide advance notice or obtain explicit consent prior to recording and distributing personal images via Street View.[278] Granted, Google provided an ex post mechanism for removing objectionable images from Street View but did not provide tools for residents of city streets to signal, ex ante, that they did not want Google to photograph them or their residences.[279] Google might object that an ex ante mechanism was impractical at the massive scale of Street View, but obviously this begs the question of whether the service violates norms of appropriateness and what should be done about it.[280] Additionally, Google might have included blurring technology in the initial U.S. roll out.[281] Instead, it assumed that the attitudes of American city dwellers would perfectly mirror U.S. legal doctrine, which offers weaker protection of public streets than of the interior of the home.[282] When Google later added blurring technology to Street View, however, it did so on a worldwide basis, something it might have done from the outset. In sum, Street View combines design successes—such as digital pixelation and

274. For a discussion of the disruptive nature of technologies that capture visual information and enable visual recognition and analysis on a massive scale, see generally Ryan Shaw, Recognition Markets and Visual Privacy (Nov. 2006) (unpublished paper), www.law.berkeley.edu/files/bclt_unblinking_shaw.pdf.

275. *See* VAIDHYANATHAN, *supra* note 261, at 99 ("Over time, as no horror stories emerged, American Google users became accustomed to the new function [of Street View] . . . .").

276. *See* Chris Matyszczyk, *Google Street View Has to Reshoot in Japan*, CNET (May 13, 2009), http://news.cnet.com/8301-17852_3-10240459-71.html.

277. *See* Frederic Lardinois, *Several Hundred Thousand Germans Opt out of Google Street View*, READWRITEWEB (Sept. 20, 2010), http://readwrite.com/2010/09/20/hundreds_of_thousands_of_germans_opt_out_of_google.

278. *See* VAIDHYANATHAN, *supra* note 261, at 100–07.

279. *See id.* at 102–05 (describing how residents of Kiel, Germany, put stickers on their front doors demanding that Google not photograph their homes, and how residents of Broughton, England formed a human chain to prevent the "Googlemobile" from entering their streets).

280. *See* NISSENBAUM, PRIVACY IN CONTEXT, *supra* note 182, at 192–93; *see also* VAIDHYANATHAN, *supra* note 261, at 102–03.

281. Pixelation may be thought of as a form of data minimization. *See supra* note 137 and accompanying text.

282. *See* Fleischer, *supra* note 263.

opt-out—with design failures—such as delayed introduction of privacy-protective features and a still unexplained breakdown of its privacy process resulting in the Wi-Fi payload data scandal.[283]

*****

In 2009, the year preceding the launch of Buzz, Google released Latitude, a location-tracking service that shares a user's position on a map with her friends and included a large number of privacy-protective features.[284] It also announced several major privacy initiatives such as the Data Liberation Front, which sought to ensure that data from any Google property was easily exportable for use with other applications and services;[285] a "Data Dashboard," which provided users with a single location to control and view their settings for every services they subscribed to or otherwise utilized;[286] and, just two weeks before launching Buzz, a new set of privacy principles.[287]

4. Buzz and Google+

On February 9, 2010, Google launched Buzz, with great hopes for competing directly with Facebook in the SNS space.[288] Towards that goal,

283. For the details of collection of Wi-Fi payload data, see Notice of Apparent Liability for Forfeiture, Google, Inc., 27 FCC Rcd. 4012 (Apr. 13, 2012), *available at* http://hraunfoss.fcc.gov/edocs_public/attachmatch/DA-12-592A1.pdf. While Google maintained that it only collected fragments of payload data, the FCC was unable to determine what use, if any, Google made of certain data due to its inability to compel testimony from key witnesses. Additionally, in March 2013, Google reached a settlement with the attorneys general of thirty-eight states and the District of Columbia to resolve state claims concerning the Wi-Fi payload data controversy. *See* Alexei Oreskovic, *Google Pays $7 Million to Settle 38-State WiFi Investigation*, REUTERS (Mar. 12, 2013), http://www.reuters.com/article/2013/03/12/us-google-wifi-fine-idUSBRE92B0VX20130312.

284. *See* Michael Zimmer, *With Latitude, Google Actually Got It (Mostly) Right*, MICHAELZIMMER.ORG (Feb. 6, 2009), http://michaelzimmer.org/2009/02/06/with-latitude-google-actually-got-it-mostly-right.

285. *See* DATA LIBERATION, http://www.dataliberation.org (last visited Apr. 11, 2012).

286. *About the Dashboard*, GOOGLE, http://support.google.com/accounts/bin/answer.py?hl=en&answer=162744 (last visited Mar. 18, 2013).

287. *See* Alan Eustace, *Google's Privacy Principles*, THE OFFICIAL GOOGLE BLOG (Jan. 27, 2010), http://googleblog.blogspot.com/2010/01/googles-privacy-principles.html. These anodyne principles obligate Google to:

- Use information to provide [its] users with valuable products and services.
- Develop products that reflect strong privacy standards and practices.
- Make the collection of personal information transparent.
- Give users meaningful choices to protect their privacy.
- Be a responsible steward of the information [it holds].

Id.

288. Todd Jackson, *Introducing Google Buzz*, THE OFFICIAL GOOGLE BLOG (Feb. 9, 2010), http://googleblog.blogspot.com/2010/02/introducing-google-buzz.html.

Buzz included a feature that, "without prior notice or the opportunity to consent, Gmail users were, in many instances, automatically set up with 'followers' (people following the user)."[289] In addition, after enrolling in Buzz, Gmail users were automatically set up to "follow" other users.[290] Moreover, Google made this information publicly accessible to anyone viewing a user's profile.[291] This decision to jump-start the Buzz social network by exploiting existing Gmail contact lists backfired, turning Buzz into a "danger zone" for investigative reporters, human rights activists, abuse victims, or anyone whose most frequent contacts were—and needed to remain—confidential.[292] Google immediately created a war room and sought to resolve problems without delay; two days later, it adjusted Buzz's user interface by making it easier to opt-out of disclosing the lists of followers and people one follows, although the disclosure option was still pre-selected.[293] In a blog post announcing further changes, Google sought to justify its decision to implement "auto-following" by noting, "we wanted to make the getting started experience as quick and easy as possible."[294] But in response to customer concerns, Google introduced a new "auto-suggest" feature, which allowed users to review and approve follower suggestions based on their most frequent contacts.[295]

These changes failed to satisfy the Electronic Privacy Information Center ("EPIC"), which soon filed a complaint with the FTC.[296] In a blog post, the Electronic Frontier Foundation ("EFF") blamed Google's privacy problems on its attempt "to overcome its market disadvantage in competing with

289. *See* Complaint at *2–5, Google, Inc., F.T.C. No. 102-3136, 2011 WL 5089551 (Mar. 30, 2011), *available at* http://www.ftc.gov/os/caselist/1023136/110330googlebuzz cmpt.pdf.

290. *See id.* These lists of "followers" and people being "followed" were based on the individuals to or with whom Gmail users most frequently emailed and/or chatted. *Id.*

291. *See id.* at *3.

292. *See* James Grimmelmann, *Privacy as Product Safety*, 19 WIDENER L.J. 793, 823–26 (2010) (quoting Nicholas Carlson). For more user complaints about Google Buzz, see EPIC's complaint filed with the FTC. Complaint, Request for Investigation, Injunction, and Other Relief ¶¶ 25–30, 37–40, Google, Inc., EPIC (Feb. 16, 2009), *available at* http://epic.org/privacy/ftc/googlebuzz/GoogleBuzz_Complaint.pdf [hereinafter EPIC Buzz Complaint].

293. *See* Nicholas Carlson, *How Google Went into "Code Red" and Saved Google Buzz*, BUSINESS INSIDER (Feb. 16, 2010), http://articles.businessinsider.com/2010-02-16/tech/30071123_1_google-buzz-google-employees-googleplex. Google's blog posts chronicled the implementation of the changes. *See* Todd Jackson, *Millions of Buzz Users, and Improvements Based on Your Feedback*, OFFICIAL GMAIL BLOG (Feb. 11, 2010), http://gmailblog.blogspot.com/2010/02/millions-of-buzz-users-and-improvements.html.

294. Jackson, *supra* note 288.

295. *Id.*

296. *See* EPIC Buzz Complaint, *supra* note 292.

Twitter and Facebook by making a secondary use of [users'] information."[297] An op-ed posted the next day by Leslie Harris of the Center for Democracy and Technology ("CDT") called Buzz "a textbook example of how to violate the principles of Privacy by Design."[298] Google was also hit with a class-action lawsuit, which it eventually settled for $8.5 million.[299]

Buzz raised multiple privacy concerns that brought about its untimely demise. Buzz violated several FIPs and related privacy engineering requirements, including inadequate and misleading notice and lack of informed consent, and these deficiencies eventually forced Google to settle both a class action lawsuit and an FTC complaint.[300] Buzz also disregarded several design guidelines, including all five pitfalls of Lederer et al.[301] and many of the design guidelines of Lipford et al. as well.[302] Google might have done things very differently with Buzz. First, it might have more clearly disclosed that users' frequent Gmail contacts would be made public by default. Second, it might have released the service not with an auto-following feature, but rather with the auto-suggest feature that it was forced to hurriedly develop under pressure. Finally, it might have provided easier and more effective options for users to exit the new service. In short, it might have made Buzz more configurable (per Feigenbaum et al.)[303] from the outset.

Because Buzz was such a spectacular defeat for an otherwise successful and savvy company, it is worth pausing for a moment to ask a slightly different question: why did Google get Buzz so wrong? danah boyd suggests two reasons: first, Google launched Buzz as a "public-facing service inside a service that people understand as extremely private."[304] But this disrupted social expectations, or as Nissenbaum would say, violated contextual integrity.[305] Second, "Google assumed that people would opt-out of Buzz if

297. Kurt Opsahl, *Google Buzz Privacy Update*, ELEC. FRONTIER FOUND. (Feb. 16, 2010), https://www.eff.org/deeplinks/2010/02/google-buzz-privacy-update.

298. Leslie Harris, *Buzz or Bust*, THE HUFFINGTON POST (Feb. 17, 2010), www.huffington post.com/leslie-harris/buzz-or-bust_b_466133.html.

299. Nick Saint, *Google Settles Over Buzz, Will Establish $8.5 Million Fund to Promote Privacy Education*, BUSINESS INSIDER (Nov. 2, 2010), http://www.businessinsider.com/google-settles-over-buzz-will-establish-85-million-fund-to-promote-privacy-education-2010-11.

300. *See id.*; Google Settlement, *supra* note 25.

301. *See supra* note 200 and accompanying text.

302. *See supra* note 230 and accompanying text.

303. *See supra* note 114 and accompanying text.

304. danah boyd, Remarks at SXSW, Making Sense of Privacy and Publicity (Mar. 13, 2010), http://www.danah.org/papers/talks/2010/SXSW2010.html.

305. *See* Nissenbaum, *Privacy as Contextual Integrity*, *supra* note 182.

they did not want to participate."[306] But this premise was flawed, as many unsuspecting users jumped into Buzz without understanding its information flows, became confused, and found it hard to exit, which only intensified their anxiety.[307]

*****

During the remainder of 2010 and through the first quarter of 2011, Google also made a number of significant policy announcements regarding its commitment to end-user privacy. To begin with, Google apologized for Buzz.[308] Eight months later, it named Alma Whitten as its director of privacy, with responsibility for ensuring that Google "build[s] effective privacy controls into [its] products and internal practices."[309] At the same time, Google committed to educating new employees on its privacy principles with an enhanced curriculum for engineers, product managers, and legal teams, and announced a new compliance process,

> in which every engineering project leader will be required to maintain a privacy design document for each initiative they are working on. This document will record how user data is handled and will be reviewed regularly by managers, as well as by an independent internal audit team.[310]

About five months later, Google agreed to a consent decree regarding Buzz, which "bar[red] the company from future privacy misrepresentations, require[d] it to implement a comprehensive privacy program, and call[ed] for regular, independent privacy audits for the next 20 years."[311] That same day, Whitten posted a blog response reaffirming Google's commitment to privacy and apologizing again for Buzz's privacy concerns.[312] The saga ended when

306. *See* boyd, *supra* note 304.

307. *Id.* (noting that she gave Google "the benefit of the doubt on this one because a more insidious framing would be to say that they wanted to force people into opting-in because this makes the service more viral and more monetizable.").

308. *See* Miguel Helft, *Anger Leads to Apology from Google About Buzz*, N.Y. TIMES, Feb. 14, 2010, at B3.

309. Alan Eustace, *Creating Stronger Privacy Controls Inside Google*, GOOGLE OFFICIAL BLOG (Oct. 22, 2010), http://googleblog.blogspot.com/2010/10/creating-stronger-privacy-controls.html.

310. *Id.*

311. Press Release, Fed. Trade Comm'n, FTC Charges Deceptive Privacy Practices in Google's Rollout of Its Buzz Social Network (Mar. 30, 2011), http://www.ftc.gov/opa/2011/03/google.shtm.

312. Alma Whitten, *An Update on Buzz*, GOOGLE OFFICIAL BLOG (Mar. 30, 2011), http://googleblog.blogspot.com/2011/03/update-on-buzz.html.

Google officially announced on October 14, 2011, that it was discontinuing Buzz.[313]

In releasing Google+ in the summer of 2011, Google sought to take a major step to recover from the Buzz debacle.[314] Built from the ground up around notions of privacy and user control, Google+ was the first SNS to rely explicitly on the idea of dividing users into "circles" (e.g., family, friends, co-workers) and to organize controls for both individual posts and one's profile information with these customized groups in mind.[315] Google's new approach was widely applauded and even cited for exemplifying privacy by design.[316] Although Google+ encountered a few privacy-related objections, they were fairly minor and did not tarnish Google's newly restored reputation.[317] Indeed, it seemed as if Google had learned its lessons from Buzz and was now making a concerted effort not only to move into the social space in a responsible way that respected users' privacy, but also to outdo its competitors in providing innovative privacy tools.[318] But was Google's new, privacy-centric SNS the first fruit of Whitten's new focus on building effective privacy controls into Google's products and internal practices[319] or just a very successful instance of Google turning lemons into lemonade?

5. Google's New Privacy Policy

On January 24, 2012, Google announced that it would soon combine almost sixty different privacy policies into one document covering virtually

313. Bradley Horowitz, *A Fall Sweep*, THE OFFICIAL GOOGLE BLOG (Oct. 14, 2011), http://googleblog.blogspot.com/2011/10/fall-sweep.html.

314. *See* Bradley Horowitz, *Buzzkiller*, GOOGLE+ (Oct. 14, 2011), https://plus.google.com/+BradleyHorowitz/posts/WjNHiWiZtYR (explaining that Google had learned from the Buzz's failures and would use the experience to improve).

315. *See* Adams, *supra* note 172.

316. Kashmir Hill, *Why 'Privacy by Design' Is the New Corporate Hotness*, FORBES (July 28, 2011), http://www.forbes.com/sites/kashmirhill/2011/07/28/why-privacy-by-design-is-the-new-corporate-hotness ("After flunking Privacy 101 with Buzz, . . . Google has designed a social network with privacy as its building block.").

317. Users initially criticized Google+ for requiring them to publicize their gender and refusing to allow anonymous names or pseudonyms. *See* Susana Polo, *Google+ Won't Let You Keep Your Gender Private, and Why That's Interesting*, THE MARY SUE (July 8, 2011), http://www.themarysue.com/google-plus-gender-private; Bradley Horowitz, *Toward a More Inclusive Naming Policy for Google+*, GOOGLE+ (June 11, 2012), https://plus.google.com/u/0/113116318008017777871/posts/SM5RjubbMmV.

318. *See generally Know Your Google Security and Privacy Tools*, GOOGLE, http://www.google.com/goodtoknow/online-safety/security-tools (last visited Feb. 27, 2012).

319. *See* Whitten, *supra* note 312.

all of its online services.[320] To ensure that users had sufficient notice of these changes prior to the March 1, 2012 implementation, the company emailed hundreds of millions of users and prominently displayed a notice on its homepage.[321] Google justified the new policy as a response to regulatory demands for shorter, simpler privacy terms and explained that the main change would be for registered users, that is, consumers with Google Accounts.[322]

Commentators and politicians alike expressed alarm over the proposed changes, mainly due to Google's decision to combine information across diverse services (some of which were previously separate).[323] They worried that these changes would create a single, large repository of user data without an easy way for the average user to opt out.[324] EPIC claimed that the proposed changes violated the Buzz consent decree and unsuccessfully sought to compel the FTC to enforce the order against Google.[325] Then, in rapid succession, Members of Congress,[326] state Attorneys General,[327] and

320. *See* Cecilia Kang, *Google Announces Privacy Changes Across Products; Users Can't Opt Out*, WASH. POST (Jan. 24, 2012), http://articles.washingtonpost.com/2012-01-24/business/35440035_1_google-web-sites-privacy-policies; Alma Whitten, *Google's New Privacy Policy*, GOOGLE OFFICIAL BLOG (Feb. 29, 2012), http://googleblog.blogspot.com/2012/02/googles-new-privacy-policy.html.

321. *See* Eyder Peralta, *Google's New Privacy Policy Will Allow Tracking Across Services*, NPR (Jan. 25, 2012), http://www.npr.org/blogs/thetwo-way/2012/01/25/145830858/googles-new-privacy-policy-will-allow-tracking-across-services.

322. Alma Whitten, *Updating our Privacy Policies and Terms of Service*, GOOGLE OFFICIAL BLOG (Jan. 24, 2012), http://googleblog.blogspot.com/2012/01/updating-our-privacy-policies-and-terms.html ("[Google+'s] new Privacy Policy makes clear that, if you're signed in, we may combine information you've provided from one service with information from other services. In short, we'll treat you as a single user across all our products, which will mean a simpler, more intuitive Google experience.").

323. *See* Hayley Tsukayama, *Google Faces Backlash over Privacy Changes*, WASH. POST (Jan. 25, 2012), http://articles.washingtonpost.com/2012-01-25/business/35439034_1_google-account-google-services-rachel-whetstone.

324. *See* Samantha Grossman, G*oogle's New Privacy Policy: Five Ways to Minimize Your Online Exposure*, TIME (Mar. 1, 2012), http://techland.time.com/2012/03/01/googles-new-privacy-policy-six-tips-for-minimizing-your-online-exposure.

325. Brenda Sasso, *Judge Dismisses Lawsuit Over Google Privacy Changes*, THE HILL (Feb. 24, 2012), http://thehill.com/blogs/hillicon-valley/technology/212509-judge-dismisses-lawsuit-over-google-privacy-changes.

326. Declan McCullagh, *Politicians Aim Some Pointed Privacy Questions at Google*, CNET (Jan. 26, 2012), http://news.cnet.com/8301-31921_3-57367059-281/politicians-aim-some-pointed-privacy-questions-at-google.

327. Letter from Nat'l Assoc. of Attys. Gen. (NAAG) to Larry Page, CEO, Google (Feb. 22, 2012) [hereinafter NAAG Letter], *available at* http://www.naag.org/assets/files/pdf/signons/20120222.Google%20Privacy%20Policy%20Final.pdf.

E.U. privacy officials[328] fired off letters to Google expressing concern and seeking additional information as well as a delay in the proposed changes. In its lengthy reply to Congress, Google reiterated that its intentions were to simplify its privacy policy and enhance user services, and that any newly combined information would not be provided to third parties.[329] Despite a preliminary assessment by the CNIL that the proposed changes violated the European privacy law,[330] they took effect as scheduled on March 1, 2012.[331]

Google's revision of its privacy policy reinforced its preexisting policy of combining data from services requiring users to sign in (e.g., Gmail but now also Web History, YouTube, and Google+) with data from many other services (including Search). While Google did an admirable job of notifying users of the pending change, critics objected to the all-or-nothing nature of the choice.[332] Far worse, multiple regulators accused Google of misleading consumers by changing its existing privacy policy without consent, failing to provide an adequate opt-out mechanism, and failing to adequately disclose "exactly which data is combined between which services for which purposes."[333] It is too soon to say what Google might have done differently until all the facts have emerged, but a few preliminary observations are in order. First, Google might have used a multilayered notice to simplify its privacy policy, given that regulators already endorsed this approach.[334] Second, Google might have permitted users to opt out of data sharing, although this would have been inconsistent both with its own model of what it means to give all users "a simpler, more intuitive Google experience"[335] and with the competitive reasons driving its shift to an integrated privacy policy.[336] Third, Google might have been more "forthcoming" in responding

328. *See* Letter from Commission Nationale de l'Informatique et des Libertés ("CNIL") to Larry Page, CEO, Google Inc. (Feb. 27, 2012) [hereinafter CNIL Letter], *available at* http://www.cnil.fr/fileadmin/documents/en/Courrier_Google_CE121115_27-02-2012.pdf.

329. Letter from Pablo Chavez, Dir. of Pub. Policy, Google Inc., to Members of Cong. Regarding Privacy Policy (Jan. 30, 2012) [hereinafter Chavez Letter], *available at* https://docs.google.com/file/d/0BwxyRPFduTN2NTZhNDlkZDgtMmM3MC00Yjc0LTg4YTMtYTM3NDkxZTE2OWRi/edit?hl=en_US.

330. Eric Pfanner, *France Says Google Plan Violates Law*, N.Y. TIMES, Feb 29, 2012, at B9.

331. *See* Sam Grobart, *Google's New Privacy Policy: What to Do*, N.Y. TIMES: GADGETWISE BLOG (Mar. 1, 2012), http://gadgetwise.blogs.nytimes.com/2012/03/01/googles-new-privacy-policy-what-to-do.

332. *See* NAAG Letter, *supra* note 327.

333. *See* CNIL Letter, *supra* note 328.

334. *Id.* The Microsoft Privacy Guidelines also recommend layered notices. *See Microsoft Privacy Guidelines*, *supra* note 141.

335. *See* Chavez Letter, *supra* note 329.

336. *See* Complaint, DeMars v. Google, No. CV12-01382, 2012 WL 4811194 at 5 (N.D. Cal. Mar. 20, 2012), *dismissed sub nom. In re* Google, Inc. Privacy Policy Litig., 2012 WL

to government inquiries.[337] Whether regulators penalize Google or eventually force it to modify any of these decision remains to be seen. What already seems clear is that Google's decisions were driven neither by the privacy by design principles discussed in Section II.B nor by its new commitment to stronger privacy controls as Whitten previously announced but rather largely by business considerations.

B. FACEBOOK

Facebook is a free, ad-supported SNS with just over 1 billion active users.[338] On May 7, 2012, the company completed an initial public offering with an estimated market value of almost $100 billion, based on approximately $4 billion in annual revenues, almost all of which derives from its online advertising business.[339] During its eight years in business, Facebook has suffered numerous privacy controversies, partly as a result of how the service works: users of Facebook create online profiles, which contain a great deal of personal and sensitive information including their name, their interests, the names of their friends, photos and videos they upload, and content they add to their friends' profiles by sending comments and sharing photos.[340] Users may also "tag" their friends' images (i.e., identify them by name) without prior consent from those friends[341] and install games and other applications developed by third parties that obtain access to the profile information of both the users and their friends.[342] In short, Facebook, by its

6738343 (2012). DeMars claimed that "Google previously targeted its advertising using bits and pieces of anonymous information garnered from each, discrete Google service," as compared with Facebook's more "holistic view of each consumer." *Id.* ¶ 18. He concluded that "Google's new privacy policy is nothing more than Google's effort to garner a larger market share of advertising revenue by offering targeted advertising capabilities that compete with or surpass those offered by social networks, such as Facebook." *Id.* ¶ 19.

337. *See* Brendan Sasso, *Google Isn't Being 'Forthcoming' with Congress on Privacy*, THE HILL (Feb. 2, 2012), http://thehill.com/blogs/hillicon-valley/technology/208385-google-not-forthcoming-during-congressional-questioning ("I don't think their answers to us were very forthcoming necessarily in what this really means for the safety of our families and our children." (quoting Representative Mary Bono Mack)).

338. *See Key Facts*, FACEBOOK, http://newsroom.fb.com/Key-Facts (last visited Mar. 8, 2013).

339. *See* Shayndi Raice, Anupreeta Das & John Letzing, *Facebook Targets $96 Billion Value*, WALL ST. J. ONLINE (May 3, 2012), http://online.wsj.com/article/SB10001424052702304746604577382210530114498.html.

340. *See* Samuel W. Lessin, *Tell Your Story with Timeline*, THE FACEBOOK BLOG (Sept. 22, 2011), http://www.facebook.com/blog/blog.php?post=10150289612087131.

341. *See What Is Tagging and How Does It Work?*, FACEBOOK, https://www.facebook.com/help/124970597582337/?q=tagging (last visited Feb. 27, 2013).

342. *See* Angwin & Singer-Vine, *supra* note 16 ("[D]on't be surprised if details about your religious, political and even sexual preferences start popping up in unexpected places.").

very nature, raises fundamental privacy challenges because it enables users to disclose unprecedented volumes of highly personal information, not only to friends and friends of friends, but, depending on one's privacy settings, to very large and unfamiliar audiences as well. We review four major Facebook features: News Feed, Beacon, Facebook Apps, and Photo Sharing; as well as a related controversy over ongoing revisions to the company's privacy polices and practices.

1. News Feed

Facebook's first major privacy incident occurred in 2006 with the launch of News Feed, a new feature that created a stream of headlines sent to all users based on the activities of their friends throughout the day including newly uploaded pictures, changes in relationships, and so on.[343] News Feed automatically enrolled all Facebook users on an opt-out basis and the feature lacked any controls over what information was shared or with which friends.[344] Users reacted with alarm over the unintended consequences of Facebook broadcasting their activities to their entire list of friends.[345] Within days, Facebook CEO Mark Zuckerberg released an open letter apologizing to users for "[messing] this one up" by failing to build in privacy controls from the outset, which Facebook promptly corrected by introducing new controls.[346] Interestingly, the controversy soon faded as users adjusted to this sudden shift from a "pull" model of sharing updates to a "push" model which, given enough time, they came to appreciate and even depend on.[347]

What went wrong with News Feed is easily seen. As Grimmelmann points out, News Feed amounted to a "privacy lurch"—that is, "an overnight change that instantly made highly salient what had previously been practically obscure."[348] Similarly, boyd—relying on Altman's work—compares

343. Ruchi Sanghvi, *Facebook Gets a Facelift*, THE FACEBOOK BLOG (Sept. 5, 2006), http://blog.facebook.com/blog.php?post=2207967130.

344. *See* Mark Zuckerberg, *An Open Letter from Mark Zuckerberg*, THE FACEBOOK BLOG (Sept. 8, 2006), http://blog.facebook.com/blog.php?post=2208562130.

345. *Id.*; *see also* Mark Zuckerberg, *Calm Down. Breathe. We Hear You.*, THE FACEBOOK BLOG (Sept. 5, 2006), http://blog.facebook.com/blog.php?post=2208197130.

346. Zuckerberg, *supra* note 344.

347. These models are drawn from logistic and supply chain management. *See* Janet Hunt, *Push System vs. Pull System Inventory Control*, CHRON, http://smallbusiness.chron.com/push-system-vs-pull-system-inventory-control-12650.html (last visited Mar. 8, 2013). In a "pull" model, users request information explicitly to be viewed. In a "push" model, information is automatically fed to a user interface that is updated without a user's explicit request, similar to a "ticker" feed in the newsroom, or a news broadcast, where information is flowing to the user in a stream rather than on demand.

348. *See* Grimmelmann, *supra* note 59, at 1201.

Facebook users to partygoers who felt "protected by the acoustics" of the loud music at a party as they exchanged intimacies with a friend only to find themselves exposed in mid-sentence when the music abruptly stopped.[349] Finally, Hull et al. describe the sudden switch to News Feed as "do[ing] violence to users' norms of distribution."[350] What might Facebook have done differently? At the very least, and consistent with the design guidelines described above, it might have given users more granular controls over information sharing as well as more time to adjust to this new model.[351]

2. Beacon

A year later, Facebook released Beacon, an addition to their developing ad platform.[352] Beacon provided targeted ads based on items a user purchased or browsed on the websites of some forty-four partner sites and shared this information with a user's friends via the News Feed.[353] Although early versions of Beacon apparently included a global opt-out capability, Facebook removed this feature prior to release in favor of more limited privacy controls.[354] Moreover, even if a Facebook user decided not to share

349. danah boyd, *Facebook's "Privacy Trainwreck": Exposure, Invasion and Drama*, APOPHENIA BLOG (Sept. 8, 2006), http://www.danah.org/papers/FacebookAndPrivacy.html.

350. *See* Gordon Hull et al., *Contextual Gaps: Privacy Issues on Facebook*, 13 ETHICS & INFO. TECH. 289, 297 (2010) (following boyd in chastising News Feed for taking somewhat obscure snippets of social action and making them highly visible); *see also id.*

351. *See supra* notes 200–02, 230 and accompanying text. Does the eventual adaptation of users to News Feed support the view that privacy concerns should not be allowed to hamper the rapid deployment of newer and better technologies, especially where a company believes that "overnight changes" are vital to its success in attracting new customers and achieving network effects? This is a difficult question requiring careful analysis of competing values. *See, e.g.*, NISSENBAUM, PRIVACY IN CONTEXT, *supra* note 182, at 229–30. Nissenbaum argues that the adjustments people make to new technologies such as News Feed "will be tempered by explicit and implicit respect for those entrenched context-relative informational norms that have been finely calibrated to support goals, purposes, and values of the context of social life and kinship, such as trust, accommodation, unconditional regard and loyalty." *Id.* Thus, adaptive patterns may include people seeking out alternative SNSs that are more sensitive to the informational norms of the relevant contexts or adjusting their behavior by finding "workarounds that mimic the constraints of the informational norms they seek." *Id.*

352. Press Release, Facebook, Leading Websites Offer Facebook Beacon for Social Distribution (Nov. 6, 2007), http://newsroom.fb.com/News/234/Leading-Websites-Offer-Facebook-Beacon-for-Social-Distribution.

353. Om Malik, *Is Facebook Beacon a Privacy Nightmare?*, GIGAOM (Nov. 6, 2007) http://gigaom.com/2007/11/06/facebook-beacon-privacy-issues ("Fandango users could now publish information about the movies they saw [on Facebook].").

354. *See* Michael Arrington, *Ok Here's at Least Part of What Facebook Is Announcing on Tuesday: Project Beacon*, TECHCRUNCH (Nov. 2, 2007), http://techcrunch.com/2007/11/02/ok-heres-at-least-part-of-what-Facebook-is-announcing-on-tuesday (describing the "number of privacy options" likely to be available to Facebook users, prior to Beacon's release).

such information with a friend, Facebook still received it.[355] Although commentators quickly labeled Beacon "a privacy disaster waiting to happen,"[356] Facebook decided to ride out the controversy, hoping that consumers might still "fall in love" with Beacon once they understood it better.[357] Instead, Facebook users revolted, voicing concerns over the risk of embarrassment or the ruining of a surprise if activity at a partner website was shared with the wrong friend or at the wrong time.[358] As the controversy heated up, Facebook tweaked Beacon's privacy notice and eventually converted Beacon to an opt-in model, with a global opt-out feature that turned it off entirely.[359] But the damage was already done: Facebook discontinued Beacon in 2009 but not before settling a class action lawsuit for $9.5 million.[360]

All of the earlier observations concerning News Feed apply, *mutatis mutandis*, to Beacon. As both Grimmelmann and Nissenbaum correctly observe, Facebook's attempt to fix Beacon by making it easier to opt out was doomed to fail as it should have been opt-in from the start.[361] Of course, good design practices might have made a difference here, but only if Facebook stepped up to the plate before releasing the new feature. This is a case where a company's preference for innovation over privacy had predictably unfortunate results.[362] Firms already have weak incentives to invest in privacy by design, and giving a pass to any firm for innovative products would only further tip the scales in favor of business goals to the detriment of sound privacy practices.

3. Facebook Apps

In 2007, Facebook launched the Facebook Platform, a set of Application programming interfaces ("APIs") and tools enabling developers to create

355. *See* Malik, *supra* note 353.

356. *Id.*

357. Louise Story & Brad Stone, *Facebook Retreats on Online Tracking*, N.Y. TIMES, Nov. 30, 2007, http://www.nytimes.com/2007/11/30/technology/30face.html?_r=0.

358. *See* Jim Tobin, *The Problem with Facebook's "Beacon,"* WEBPRONEWS (Nov. 27, 2007), http://www.webpronews.com/backlash-against-facebooks-beacon-2007-11.

359. *See* Mark Zuckerberg, *Thoughts on Beacon*, THE FACEBOOK BLOG (Dec. 5, 2007), http://blog.facebook.com/blog.php?post=7584397130.

360. Jon Brodkin, *Facebook Halts Beacon, Gives $9.5M to Settle Lawsuit*, PCWORLD (Dec. 8, 2009), http://www.pcworld.com/article/184029/facebook_halts_beacon_gives_9_5_million_to_settle_lawsuit.html.

361. *See* Grimmelmann, *supra* note 59, at 1201–02; NISSENBAUM, PRIVACY IN CONTEXT, *supra* note 182, at 223.

362. *See* Rubinstein, *supra* note 1, at 1436–40 (discussing the economic reasons why firms underinvest in privacy and security).

hundreds of thousands of third-party applications ("apps") for Facebook users.[363] Popular apps include games, instant messaging, and a forum for social activists to share their ideas.[364] Once approved by Facebook, apps may retrieve or post information to member profiles and request information about users and their friends.[365] Users are required to grant access privileges to apps as a condition of installing them.[366] However, most applications were given access to far more private information than they needed.[367] Moreover, many users lacked understanding of what data they were sharing when they installed an app, either because they hurried through the installation process and ignored notices or relied on the fact that applications ran within the boundary of Facebook, wrongly inferring that their data would remain within the Facebook network.[368] These issues led Canadian privacy regulators to investigate such complaints. They found that Facebook lacked adequate safeguards effectively restricting outside developers from accessing a user's profile information,[369] and called for technological measures restricting access to the information that was actually required to run a specific application.[370]

Facebook responded by restricting third-party app access to only the public parts of a user's profile unless the user granted additional permission.[371] It then announced a new permissions model for third-party applications,[372] which eventually satisfied the Canadian regulators.[373] A year later, Facebook took additional steps to address privacy issues in third-party apps by releasing a new dashboard allowing users to see exactly how and

363. *See* Jonathan Strickland, *How Facebook Works*, HOWSTUFFWORKS.COM, http://computer.howstuffworks.com/internet/social-networking/networks/facebook3.htm (last visited Mar. 25, 2013).

364. *See* Jennifer King et al., *Privacy: Is There an App for That?*, *in* PROCEEDINGS OF THE SYMPOSIUM ON USABLE PRIVACY AND SECURITY (SOUPS) art. 12 (2011), *available at* http://cups.cs.cmu.edu/soups/2011/proceedings/a12_King.pdf.

365. *See* Strickland, *supra* note 363.

366. *See* Lipford et al., *supra* note 219, at 987.

367. *See* Adrienne Felt & David Evans, *Privacy Protection for Social Networking APIs*, U. VA., http://www.cs.virginia.edu/felt/privacy (last visited Feb. 27, 2013) (noting that over ninety percent of the top 150 apps received more privileges than they actually needed).

368. *See* Lipford et al., *supra* note 219.

369. *See* DENHAM, *supra* note 70, at 37–47.

370. *Id.*

371. Emily Steel & Geoffrey A. Fowler, *Facebook in Privacy Breach*, WALL ST. J. (Oct. 17, 2010), http://online.wsj.com/article/SB10001424052702304772804575558484075236968.html.

372. InsideFacebook.com, *Facebook Announces Significant Changes to the Way Applications Can Access User Data*, FACEBOOK (Aug. 27, 2009), http://www.facebook.com/note.php?note_id=123271997739.

373. *Backgrounder: Facebook Investigation Follow-up Complete*, OFFICE OF THE PRIVACY COMMISSIONER OF CANADA (Sept. 22, 2010), http://www.priv.gc.ca/media/nr-c/2010/bg_100922_e.cfm.

when their data has been accessed through the Facebook Platform, and giving users the option to remove unwanted apps, games, or sites, or to revoke persistent permissions.[374]

Despite laboring to address these longstanding privacy issues, Facebook encountered new problems with third-party apps. For example, a Wall Street Journal investigation revealed that many Facebook apps were not only providing data to advertisers but also linking it directly to users' names and their friends' names.[375] Then, in November 2011, Facebook agreed to a consent decree with the FTC based on an eight-count complaint including allegations concerning the consent model of Facebook Apps; it was ordered not to misrepresent "the extent to which it makes or has made covered information accessible to third parties."[376] In December 2011, the Data Protection Commissioner of Ireland completed a very extensive audit of Facebook and asked it, inter alia, to create a system to allow users to control how their data is shared with third-party apps.[377] More recently, a new Wall Street Journal investigation found that even though apps must ask permission before accessing a user's personal details, "a user's friends aren't notified if information about them is used by a friend's app. An examination of the apps' activities also suggests that Facebook occasionally isn't enforcing its own rules on data privacy."[378]

Facebook Apps is more complex than News Feed or Beacon and raises multiple issues. To begin with, Facebook introduced its apps platform with permissive defaults for developers and overly broad access to profile information, while limiting users to an all-or-nothing choice over what information they had to share as a condition of installing the app. Moreover, Facebook Apps violated norms of distribution by forcing users to share their own and their friends' information in unexpected ways with unknown third parties who were not vetted by Facebook and remained largely invisible to ordinary users, who were in no position to conduct their own evaluations.[379]

Facebook might have done several things differently and still have succeeded in launching a successful apps platform. First, it might have followed a data minimization approach by restricting what information apps

374. Josh Constine, *How to Use Facebook's Application Settings Dashboard*, INSIDE FACEBOOK (Oct. 7, 2010), http://www.insidefacebook.com/2010/10/07/how-to-application-settings-dashboard.

375. Steel & Fowler, *supra* note 371.

376. Facebook Settlement, *supra* note 25.

377. *See* IRISH AUDIT, *supra* note 70.

378. Angwin & Singer-Vine, *supra* note 16.

379. *See* Lipford et al., *supra* note 219.

could access from the outset.[380] Second, it could have shipped the API with a permissions model and a Dashboard, rather than waiting several years to implement these features—and only then under regulatory pressure. Finally, it might have designed better user interfaces with the goal of disclosing and emphasizing the information flows that occur when users install various apps.[381]

4. Photo Sharing

Facebook allows users to share photos with their friends in multiple ways. Users can upload photos to an album, post photos directly to their profile, or post directly to someone else's profile.[382] Once a photo has been posted, users may tag it, which creates a link between the tagged photo and a person, page, or place, thereby revealing additional information about the identity and associations of the people depicted in the photo.[383] Users may tag themselves or their friends, who will be notified of the tag.[384] Tagging people also alters the potential audience who can view a photo.[385] Users can remove the tag from the photo, which removes the explicit reference to the user (by eliminating the link to the user's profile), but the photo remains on Facebook, accessible from any friends' profiles to which it is cross-linked.[386]

As Facebook tagging has taken off, so has the desire of individuals to retain control over unflattering images.[387] Individuals are especially concerned about the unintended results of tagged photos, which may cause embarrassment or humiliation if family, employers, school officials, or law enforcement officials see photos meant for different eyes.[388] These tagging

380. *See supra* notes 115–16 and accompanying text.

381. *See supra* notes 200, 230 and accompanying text; *see also* ANDREW BESMER ET AL., *SOCIAL APPLICATIONS*: EXPLORING A MORE SECURE FRAMEWORK 5 (2009), *available at* http://www.andrewbesmer.com/wordpress/wp-content/uploads/2009/08/socialapplications.pdf (describing an interface prototype for Facebook Apps that "provides a more accurate mental model" of sharing and "serves to catch the user's attention").

382. *See Uploading Photos & Profile Pictures*, FACEBOOK HELP CENTER https://www.facebook.com/help/photos/upload-photos-and-profile-pictures (last visited July 23, 2012).

383. *See What Is Tagging and How Does It Work?*, FACEBOOK, https://www.facebook.com/help/124970597582337/?q=tagging&sid=0GzpxRPeunNeIBXiI (last visited Feb. 27, 2013) ("[I]f you or a friend tags someone in your post and the post is set to Friends or more, the post could be visible to the audience you selected plus friends of the tagged person.").

384. *See How Tagging Works*, *supra* note 215.

385. *Id.*

386. *Id.*

387. *See* Lisa Guernsey, *Picture Your Name Here*, N.Y. TIMES, July 27, 2008, http://www.nytimes.com/2008/07/27/education/edlife/27facebook-innovation.html.

388. *See supra* note 221 and accompanying text; *see also* Besmer & Lipford, *supra* note 218.

disputes are exacerbated by the fact that the tagging process often involves three distinct individuals—the photographer, the tagger, and the tagged subject—who may disagree over the propriety of tagging a given photo.[389] These issues will likely become even more prevalent given Facebook's creation of the Photo Tag Suggest feature, which uses facial recognition technology to help users tag even more photos.[390] Users can opt out of this feature and provide direct feedback about any items that friends post or share.[391]

After the rollout of Photo Tag Suggest, Facebook announced changes in August 2011 to enhance users' control over who could see photos, tags, and other content.[392] The main change was moving the privacy controls from a settings page to an inline control adjacent to the affected photos.[393] Each photo or album now has a drop down menu that allows a user to control exactly who can access it.[394] Facebook also added a new Profile Tag Review feature that allowed users to approve or reject any photo in which they were tagged before it became visible on their profile.[395] Finally, Facebook changed the way the options for removing tags or content on Facebook are presented to users.[396] They now have options to remove a photo from their profile, remove the tag itself, send a message to the owner or tagger, or request that the content be taken down.[397] The Irish regulators raised some initial concerns about photo tagging but were generally satisfied by these new controls.[398]

Photo Sharing introduces a new set of issues involving two kinds of peer-produced privacy violations. The first arises due to the "shrinking perceived audience" problem, in which users indiscriminately disclose potentially embarrassing photos because they forget just how many people can view them notwithstanding their intentions to share them with a much smaller

389. *See* Grimmelmann, *supra* note 59, at 1137, 1172.

390. *See* Justin Mitchell, *Making Photo Tagging Easier*, THE FACEBOOK BLOG (June 30, 2011), https://www.facebook.com/blog/blog.php?post=467145887130.

391. *Id.*

392. *See* Chris Cox, *Making It Easier to Share with Who You Want*, THE FACEBOOK BLOG (Aug. 23, 2011), https://blog.facebook.com/blog.php?post=10150251867797131.

393. *Id.*

394. *Id.*

395. *Id.*

396. *Id.*

397. *Id.*

398. IRISH AUDIT, *supra* note 70. Section 3.12 of the audit suggests that users be given the option to prevent themselves from being tagged and Facebook has agreed to look into this option. *Id.* § 3.12.

audience.[399] The second implicates the social fallout from tagging disputes, where the photographer, the tagger, and the subject disagree over whether the photo should be untagged, made private, or even removed.[400] As Grimmelmann notes, Facebook is the catalyst of these privacy violations, not the perpetrator.[401]

Might Facebook have taken steps to assist users in avoiding or limiting these peer-produced privacy harms? Yes. First, it might have done much more to avoid the "five pitfalls for designers" identified by Lederer et al.—for example, by ensuring that users understood the potential and actual information flows when they posted photos and making it easier for them to configure the relevant privacy settings as part of their ordinary use of the photo-posting feature.[402] Second, it might have developed innovative privacy tools along the lines of Restrict Others when it released new features such as photo tagging.[403] Granted, Facebook did just that in August 2011 with Photo Tag Suggest, but this was already late in the game and in response to regulatory pressure.[404]

5. Changes in Privacy Settings and Policies

Over the years, Facebook has modified both its privacy settings and policies many times. Here we focus on the period from late June 2009 to December 2011. On June 24, 2009, Facebook launched a beta version of a "publisher privacy control" that allowed users to decide who can see their published content (status updates, photos, etc.) on a per-post basis using a standardized drop-down menu.[405] A week later, Facebook moved to simplify its privacy settings by putting them all on the same page and creating a transition tool.[406] These changes were at least partly motivated by Canada's far-ranging investigation of Facebook's privacy practices and policies.[407] One of the issues that Facebook resolved related to default privacy settings.[408]

399. *See* Hull et al., *supra* note 350, at 227.

400. *See* Grimmelmann, *supra* note 59, at 1172.

401. *Id.* at 1164.

402. *See supra* note 200 and accompanying text.

403. *See supra* note 223 and accompanying text.

404. *See* IRISH AUDIT, *supra* note 70, § 3.12.

405. *See supra* notes 390–97 and accompanying text.

406. Chris Kelly, *Improving Sharing Through Control, Simplicity and Connection*, THE FACEBOOK BLOG (July 1, 2009), http://blog.facebook.com/blog.php?post=101470352130 (stating that "the compounding effect of more and more settings has made controlling privacy on Facebook too complicated" and noting that the transition tool was designed to respect users' previous decisions to limit access to information).

407. *See* DENHAM, *supra* note 70.

408. *See id.*

Although the Commissioner's Office was especially concerned with the default settings for photo sharing (specifically, that "Everyone"—all Internet users—could view the photos) and for public search listings (pre-checked to make name, networks, thumbnail picture, and friends available to search engines for indexing), it concluded that Facebook's plans to introduce a privacy wizard and implement a per-object privacy tool resolved its concerns.[409]

As a result of the Canadian investigation, Facebook modified its privacy policy and settings in August[410] and again in late October.[411] Privacy advocates praised Facebook's efforts to simplify privacy settings and liked the transition tool, at least in principle.[412] At the same time, they took issue with several changes, most notably Facebook's expansion of profile information classified as publicly available: from name and network, to profile picture, current city, friends list, gender, and fan pages.[413] Although Facebook soon backtracked on making friends lists publicly available,[414] EPIC filed a complaint with the FTC urging it to open an investigation into Facebook's revised privacy settings,[415] while Canadian privacy regulators opened a new investigation that was not resolved until September 2010.[416]

The next major chapter in this saga occurred in Spring 2010. In April, Facebook made a significant change to how it classified and disclosed users' profiles by requiring all users to designate personal information as publically available "Links," "Pages," or "Connections"; if they declined, Facebook would delete this previously restricted information from their profiles.[417] At

409. *See id.* ¶¶ 88–95.

410. InsideFacebook.com, *supra* note 372 (adopting a "permissions model" for application developers, improving explanations of collection of date of birth and of account deactivation versus deletion, and explaining privacy settings during signup).

411. Elliot Schrage, *Improving Transparency Around Privacy*, THE FACEBOOK BLOG (Oct. 29, 2009), http://blog.facebook.com/blog.php?post=167389372130.

412. *See* Nicole Ozer, *Facebook Privacy in Transition—But Where Is It Heading?*, ACLU OF N. CAL. (Dec. 9, 2009), http://www.aclunc.org/issues/technology/blog/facebook_privacy_in_transition_-_but_where_is_it_heading.shtml.

413. *Id.*

414. Caroline McCarthy, *Facebook Backtracks on Public Friend Lists*, CNET (Dec. 11, 2009), http://news.cnet.com/8301-13577_3-10413835-36.html.

415. Brad Stone, *Privacy Group Files Complaint on Facebook Changes*, N.Y. TIMES: BITS BLOG (Dec. 17, 2009), http://bits.blogs.nytimes.com/2009/12/17/privacy-group-files-complaint-on-facebook-privacy-changes.

416. DENHAM, *supra* note 70.

417. This profile data included a user's friends list, music preferences, affiliated organizations, employment information, educational institutions, film preferences, reading preferences, and other information. Facebook did not permit users to opt-out of linking their profiles to publicly available "Links," "Pages," or "Connections"; rather, it stated, "if

the same time, Facebook announced two new features: social plug-ins (which added "like" and "recommend" buttons to third-party websites without clearly indicating to users when their profile information might be shared with these websites), and "instant personalization" (which allowed a few select partners to personalize their web pages by using personal information that Facebook disclosed without a user's explicit consent).[418] These changes were immediately and widely criticized by privacy advocates, bloggers, and Members of Congress, and led EPIC to file a second complaint with the FTC.[419] The bad press continued into the month of May with the *New York Times* publishing, in graphic detail, the complexity of Facebook privacy settings,[420] and the *Wall Street Journal* exposing a serious privacy loophole.[421]

Responding to the growing controversy, Facebook announced, in late May, a complete overhaul of its privacy settings.[422] The new controls, which were based on extensive consultations with consumers and critics alike, promised to give users "control over how their information is shared" and to avoid sharing personal information "with people or services users don't want."[423] This was followed three months later by several more improvements in its privacy controls addressing many of the issues previously identified in complaints filed with the FTC. The major changes

you don't link to any pages, these sections on your profile will be empty. By linking your profile to pages, you will be making these connections public." *See* Complaint, Request for Investigation, Injunction, and Other Relief ¶ 53, Facebook, Inc., EPIC, F.T.C. No. 092-3184 (May 5, 2010), *available at* http://epic.org/privacy/facebook/EPIC_FTC_FB_Complaint.pdf [hereinafter EPIC Facebook Complaint]. For Facebook's explanation of these new features, see Alex Li, *Connecting to Everything You Care About*, THE FACEBOOK BLOG (Apr. 19, 2010), http://blog.facebook.com/blog.php?post=382978412130.

418. *See* Austin Haugen, *Answers to Your Questions on Personalized Web Tools*, THE FACEBOOK BLOG (Apr. 26, 2010), http://blog.facebook.com/blog.php?post=384733792130.

419. EPIC Facebook Complaint, *supra* note 417.

420. Guilbert Gates, *Facebook Privacy: A Bewildering Tangle of Options*, N.Y. TIMES, May 12, 2010, http://www.nytimes.com/interactive/2010/05/12/business/facebook-privacy.html (noting that managing privacy on Facebook means navigating "through 50 settings with more than 170 options").

421. Emily Steel & Jessica E. Vascellaro, *Facebook, MySpace Confront Privacy Loophole*, WALL ST. J. ONLINE, May 21, 2010, http://online.wsj.com/article/SB10001424052748704513104575256701215465596.html?mod=WSJ_hps_LEFTWhatsNews (describing how Facebook and others gave online ad firms data that could be used to look up individual profiles).

422. *See* Mark Zuckerberg, *From Facebook, Answering Privacy Concerns with New Settings*, WASH. POST, May 24, 2010, http://www.washingtonpost.com/wp-dyn/content/article/2010/05/23/AR2010052303828.html.

423. *Id.*

included new inline profile and posting controls, profile and content tag reviews, and the ability to remove tags or content from Facebook.[424]

In parallel with these changes, Facebook continued to press the boundaries of privacy through the remainder of 2011. In September 2011, Facebook announced several key design changes as well as new opportunities for advertisers.[425] The first was a new user interface known as "Timeline," which included all of a user's former posts, apps, and Facebook-related information organized into a graphic timeline of the user's life.[426] The second was the concept of "Frictionless Sharing," a means for users to share their interactions with websites and advertiser's products automatically with their friends via News Feed.[427] The third, what Facebook dubbed "Open Graph," was a platform that expanded on the notion of frictionless sharing by allowing apps to insert interactions into a user's News Feed.[428] Open Graph also allowed apps to post ads via News Feed.[429] Within days, privacy advocates were asking the FTC to ban several of these new features.[430] They voiced concerns about the automatic sharing of news articles and other information if users choose to enable "social readers," and about Facebook's use of the "Like" button, which continued to track users even after they logged out of Facebook.[431]

At the end of November, Facebook settled with the FTC.[432] In the aftermath of the settlement, Zuckerberg publicly conceded that although Facebook had made mistakes in the past, it was now committed to becoming

424. *See Facbook to Allow Users to Pre-Approve Photo Tags*, BILLBOARD BIZ (Aug. 24, 2011), http://www.billboard.com/biz/articles/news/1173330/facebook-to-allow-users-to-pre-approve-photo-tags.

425. *See* Daniel Terdiman, *What Facebook Announced at F8 Today*, CNET (Sept. 22, 2011), http://news.cnet.com/8301-1023_3-20110181-93/what-facebook-announced-at-f8-today.

426. *See* Samuel W. Lessin, *Tell Your Story with Timeline*, THE FACEBOOK BLOG (Sept. 22, 2011), https://www.facebook.com/blog/blog.php?post=10150289612087131.

427. *See* Mathew Ingram, *Why Facebook's Frictionless Sharing Is the Future*, GIGAOM (Oct. 3, 2011) http://www.businessweek.com/technology/why-facebooks-frictionless-sharing-is-the-future-10032011.html.

428. *See* Terdiman, *supra* note 425.

429. *Open Graph Concepts*, FACEBOOK DEVELOPERS, https://developers.facebook.com/docs/opengraph (last visited Mar. 23, 2012). For example, an auto company could create an app for users to comment on test drives and post this information to their News Feed.

430. *See* Declan McCullagh, *Groups Ask Feds to Ban Facebook's 'Frictionless Sharing,'* CNET (Sept. 29, 2011), http://news.cnet.com/8301-31921_3-20113457-281/groups-ask-feds-to-ban-facebooks-frictionless-sharing.

431. *See id.*

432. *See* Facebook Settlement, *supra* note 25.

a leader in transparency and user control.[433] According to his blog post, Facebook would begin to formalize privacy reviews by making them part of the company's design and development process.[434]

European regulators were also concerned with Facebook's privacy practices. On December 12, 2011 the Irish Data Protection Commissioner released a 150-page audit report, by far the most extensive government audit of a major Internet firm to date.[435] The report describes numerous changes in policies and practices that Facebook had agreed to, including a new mechanism for users to convey an informed choice for how their information is used and shared on the site and in relation to Third Party Apps, as well as increased transparency and controls for the use of personal data for advertising purposes.[436] A few days later, however, Facebook announced that it would post archived user information on Timeline without user consent.[437] With this feature scheduled to go live on December 22, 2011, users had just one week to clean up their entire history of Facebook activities.[438] This was particularly troubling given Facebook's later announcement that Timeline would ultimately become mandatory for all Facebook users.[439]

As the year ended, EPIC filed comments with the FTC regarding the November consent decree in which it elaborated on its concerns with Timeline, not only labeling it a privacy risk but pointing out that security experts deemed it a "treasure trove" of personal information that easily could

433. *See* Mark Zuckerberg, *Our Commitment to the Facebook Community*, THE FACEBOOK BLOG (Nov. 29, 2011, 9:39 AM), http://blog.Facebook.com/blog.php?post=10150378701937131.

434. *Id.*

435. *See* IRISH AUDIT, *supra* note 70.

436. *Id.*

437. Kristin Burnham, *Facebook's New Timeline: Important Privacy Settings to Adjust Now*, CIO (Dec. 21, 2011), http://www.cio.com/article/690742/Facebook_s_New_Timeline_Important_Privacy_Settings_to_Adjust_Now.

438. *Id.*

439. Paul McDonald, *Timeline: Now Available Worldwide*, THE OFFICIAL FACEBOOK BLOG (Dec. 15, 2011) (updated Jan. 24, 2012) (noting that "[o]ver the next few weeks, everyone will get timeline"). Facebook started the migration with "Pages," which automatically switched over to Timeline on March 29, 2012. *See* Josh Constine, *Don't Dread Today's Mandatory Switch to Timeline, Studies Show It's Good for 95% of Facebook Pages*, TECHCRUNCH (Mar. 29, 2012), http://techcrunch.com/2012/03/29/mandatory-switch-to-timeline. Facebook initiated the mandatory transition for users later in August 2012. *See* Mike Flacy, *Facebook Finally Starts Forcing Timeline on All Users*, DIGITAL TRENDS (Aug. 2, 2012), http://www.digitaltrends.com/social-media/facebook-finally-starts-forcing-timeline-out-to-users.

be used to compromise a user's identity.[440] Users also complained that Timeline revealed too much information, essentially opening up their entire history to anyone they had ever added as a friend.[441] Facebook responded with a blog entry describing several new, privacy-enhancing measures including a seven-day review period before a user's Timeline went live, an activity log, a more easily accessible "view as" feature, the ability to easily control who could view posts (including an "only me" feature), and the ability to limit the audience for past posts.[442]

Despite years of negative press, user revolts, the exacting scrutiny of privacy advocates, foreign and domestic investigations, audits, settlements, and other concessions, Facebook users still migrated to Timeline as scheduled.[443] Moreover, in anticipation of going public, the company continued to experiment with new ways to increase ad revenues by targeting users based not only on their profile information and on-site social activities, but also on their purchasing plans as expressed by their so-called "in-app" activity.[444] In short, privacy incidents seem to have had limited impact on the company's rapid and relentless pace of product development.

News Feed and Beacon were discrete events that flared up quickly, drew an immediate company response, and then died down or led to the new feature's modification or demise. Along similar lines, Facebook Apps and Photo Sharing, even if more protracted, eventually led to design modifications and/or new privacy settings. However, the controversies surrounding Facebook's frequent changes in privacy policies and settings exhibit a far more complex pattern. Over time, advocacy groups filed

440. EPIC, Comments to the FTC at 27; Facebook, Inc., F.T.C. No. 092-3184 (Dec. 27, 2011), http://epic.org/privacy/facebook/Facebook-FTC-Settlement-Comments-FINAL.pdf.

441. *See* Anthony Bond, *Facebook's Controversial 'Timeline' Feature Is Supported by Just One in Ten Users*, MAIL ONLINE (Jan. 30, 2012), http://www.dailymail.co.uk/sciencetech/article-2093811/Facebooks-controversial-timeline-feature-supported-just-users.html.

442. *See Controlling What You Share on Timeline*, FACEBOOK PRIVACY (Dec. 20, 2011), https://www.facebook.com/notes/facebook-and-privacy/controlling-what-you-share-on-timeline/271872722862617.

443. *See supra* note 439.

444. *See* Josh Constine, *Facebook's Revenue Growth Strategy: Ad Targeting by In-App Behavior*, TECHCRUNCH (Feb. 1, 2012), http://techcrunch.com/2012/02/01/action-spec-ad-targeting/ (describing how "Facebook has been quietly rolling out the beta of 'Open Graph action spec targeting' which allows advertisers to target users by what they listen to, where they travel, what they buy, and other in-app activity"); *see also* Tanzina Vega, *Substantial Growth in Ads Is on the Way to Facebook*, N.Y. TIMES, Mar. 1, 2012, at B2 (noting that "Facebook is moving all marketers' pages to its new Timeline format that allows advertisers to have more dynamic pages for their own brands" and that "anything posted on an advertiser's own page—status updates, photos and videos—can be made into an ad that can be pushed out to users' newsfeeds and mobile feeds").

complaints with regulators based on a diverse set of accumulated privacy concerns. Many months later, as regulators released their findings, Facebook implemented or announced changes in the relevant practices. But this activity occurred in parallel with a steady flow of fresh or newly designed features; these features often supported, but sometimes undermined, agreed-upon compliance measures and spawned another round of complaints, regulatory demands, and yet another cycle of adjustment.

One might argue that Facebook ought to have slowed both its rapid pace of innovation and its incessant tinkering with privacy settings. The former does not fly, but might the latter? At the very least, Facebook might have avoided coupling privacy revisions at the behest of regulators with sudden changes to how it classified profile information (i.e., as "publicly available"). Second, in making changes in response to regulatory concerns, it might have ensured that any transition tools or privacy wizards it offered were neutral, not self-serving, and given users a full range of privacy-protective options. Third, Facebook might have continued down the road taken in May 2010, when it engaged in consultations with consumers and privacy advocates before overhauling its privacy settings. Indeed, Facebook has taken steps to address privacy issues by adding design staff with a background in HCI, as well as policy professionals with deep privacy expertise.[445]

C. SUMMARY

In sum, the preceding counterfactual analyses suggest that all ten privacy incidents might have been avoided by the application of the engineering and usability principles and related design practices discussed in this Article. This is important for two reasons. First, it strongly supports the claim that privacy by design (when so understood) effectively protects consumer privacy. Second, it also suggests that Part II offers a reasonably comprehensive account of privacy engineering and design practices, at least as measured by these ten incidents. Specifically, notice and informed consent were applicable to all of the incidents except Search; data avoidance and minimization were applicable to Gmail, Search, Street View, Buzz, and Facebook Apps; and data retention limits were applicable to Search. In addition, avoiding design pitfalls (per Lederer et al.)[446] and following design guidelines (per Hull et al.)[447] would have improved Buzz, News Feed, Beacon, Facebook Apps, and

445. *See* Whitney, *supra* note 86 (discussing Facebook's hiring of Chris Weeldreyer, product design manager from Apple); Zuckerberg, *supra* note 433 (announcing the appointment of Erin Egan as Chief Privacy Officer, Policy, and Michael Richter as Chief Privacy Officer, Products).

446. *See supra* note 200 and accompanying text.

447. *See* Hull et al., *supra* note 350.

Photo Tagging, possibly averting all of the privacy incidents involving social networking. We suspect that these and other principles and practices described in Section II.B would be relevant to a broader set of privacy incidents.

The ten Google and Facebook privacy incidents also suggest other interesting patterns. Not every incident involved an unmitigated failure—both Gmail and Street View involved a mix of design successes and failures. Several incidents involved norm violations (News Feed and Beacon) or unsettled norms (Gmail and Street View). Quite a few incidents—Street View, Buzz, News Feed, Beacon, Facebook Apps, and Photo Tagging—were characterized by company delay in adding privacy features, revealing a "ship now and ask privacy questions later" mentality.[448] Both Google and Facebook ran into privacy difficulties when they allowed business necessities to override privacy concerns or forced users into all or nothing choices, specifically in Gmail, Search, Buzz, Google's new privacy policy, Facebook Apps, and several new Facebook features rolled out at the F8 developers' conference.[449] In all of these business-driven cases, the stated rationale of both firms was either opaque or self-serving. Almost all of the Google and Facebook privacy incidents resulted from multiple causes or flaws. Interestingly, only one incident—the Street View Wi-Fi data collection—was attributable to an apparent break down in internal review processes. In short, these patterns seem to confirm that all of these incidents were largely avoidable.

IV. LESSONS LEARNED

Having analyzed what went wrong and what Google and Facebook might have done differently in ten privacy incidents, what have we learned? What lessons does this counterfactual analysis hold for regulators that are now placing bets on privacy by design?

The first lesson is that companies and regulators should avail themselves of the rich body of research related to privacy engineering and usability design as described in Section II.B. Too often, regulators recommend that companies "build in" privacy or "design and implement" reasonable privacy controls, without explaining what they mean.[450] As designers motivate their

448. Rapid, iterative software development tends to neglect security and privacy requirements, but this is no excuse given the availability of relevant guidance adapted to the fast pace of Agile development methods. *See supra* note 78.

449. *See* Terdiman, *supra* note 425.

450. *See, e.g.*, FTC FINAL REPORT, *supra* note 2, at 2 ("The concept of privacy by design includes limitations on data collection and retention, as well as reasonable security and data

own work by means of principles and examples, it would be very helpful for regulators to provide more detailed principles and specific examples as well. We hope that Section II.B begins the process of defining what privacy by design means in engineering and design terms.

The second lesson is that usability is just as important as engineering principles and practices. As we have seen, usability and user experiences are especially relevant to the privacy issues that arise whenever people voluntarily share personal information via social networks such as Buzz, Google+, and Facebook. We believe that the best way to preserve the social dynamics of privacy is by following design guidelines as summarized above.

The third lesson is that more work needs to be done on refining and elaborating design principles—both in privacy engineering and usability design. This implies that U.S. and European regulators need to increase their efforts to understand and develop these principles by convening working sessions with companies, academics, user groups, and design professionals; identifying and codifying best practices; funding more research in privacy engineering and usability studies; and encouraging ongoing efforts to define international privacy standards.[451]

The fourth and final lesson is that regulators must do more than merely recommend the adoption and implementation of privacy by design. Recommending—or even requiring—privacy by design seems insufficient given the fact that, throughout the period of time involving the ten privacy incidents, Google and Facebook already were committed to embedding privacy into their development processes. And yet these privacy incidents still occurred. It is not at all clear that anything would change if these companies now recommitted—voluntarily or under a regulatory mandate—to adopting privacy by design. Something more is needed.

Recall that Gmail, Search, and Street View are all well-engineered services and that, in advance of their release, Google gave considerable thought to their privacy implications.[452] Buzz, of course, is different. Was it rushed to market for competitive reasons without a proper internal privacy review? Perhaps. And yet it seems unlikely that a major product release like Buzz

accuracy."); Google Settlement, *supra* note 25, at 5 (requiring Google to establish a program including "the design and implementation of reasonable privacy controls and procedures to address the risks identified through the privacy risk assessment"). In neither case does the FTC explore this notion in greater depth. *Id.*

451. *See, e.g.*, INTERNATIONAL STANDARDS, ISO/IEC 29100 INFORMATION TECHNOLOGY-SECURITY TECHNIQUES-PRIVACY FRAMEWORK (Dec. 15, 2011).

452. For purposes of this argument, we emphasize the notice and consent aspects of Street View, including opt-out and visual blurring mechanisms, and not the disputed aspects of Google's collection of Wi-Fi payload data.

would escape internal review by Google's privacy experts or that no one realized the privacy implications of the auto-enroll feature. It seems more plausible to suppose that—much like the internal debates at Microsoft over how proposed privacy features in IE 8 might affect business goals such as enabling a desirable ad platform for business partners[453]—there were internal divisions at Google over whether a more privacy-friendly version of Buzz would hinder business imperatives such as quickly catching up with Facebook and Twitter.[454] As for Google's newly integrated privacy policy, it strikes the authors as ludicrous to think that Google failed to conduct an internal privacy review before announcing, much less implementing, such major policy changes. To the contrary, the foregoing analysis suggests that these changes were carefully planned and very well executed, notwithstanding the negative reactions they garnered from regulators and the public. Indeed, the Buzz settlement legally obligated Google to implement a comprehensive privacy program and to all appearances it has done so.[455] So what is happening here? We believe that Google (like many of its peers) understands privacy requirements in a flexible manner that nicely accommodates its own business interests. We believe that the five privacy incidents we examined in Section III.A demonstrate that Google's corporate policy permits it to "balance" privacy requirements against core business goal like increasing advertising revenues.[456] Furthermore, this balancing process is almost completely hidden from outside observers.

Along similar lines, Facebook, despite its many privacy woes, has long prided itself on offering users extensive control over how they share information and who has access to it. In a nutshell, this is what Facebook seems to mean by privacy—it is a recurrent theme in Facebook's public statements about privacy, dating back at least to September 2005 when it hired its first Chief Privacy Officer.[457] Of course, Facebook offered very weak controls in rolling out a few early features like News Feed[458] and Beacon.[459] But in announcing new privacy settings for News Feed and later

453. *See supra* note 32 and accompanying text.

454. *See supra* note 297 and accompanying text.

455. *See* Google Settlement, *supra* note 25.

456. *See* Fleischer, *supra* note 251 (citing "balance" as a factor in Search privacy); Fleischer, *supra* note 263 (citing "balance" as a factor in Street View privacy).

457. *See Making the Internet Safe for Kids: The Role of ISP's and Social Networking Sites: Hearings Before the Subcomm. on Oversight and Investigations of the H. Comm. of Energy and Commerce*, 109th Cong. 214, 215 (2006) (written statement of Chris Kelly, Chief Privacy Officer, Facebook) ("[W]e put power in our users' hands to make choices about how they reveal information.").

458. *See supra* Section III.B.1.

459. *See supra* Section III.B.2.

products, Zuckerberg and other company officials consistently described what they were doing in terms of developing new privacy controls.[460] Even after Facebook settled with the FTC, at which point it was legally obligated to implement a comprehensive privacy program, Zuckerberg insisted that giving people "complete control over who they share with at all times" has been "the core of Facebook since day one."[461] And while Zuckerberg conceded that the company had to "improve and formalize the way we do privacy review as part of our ongoing product development process," he continued to emphasize the "more than 20 new tools and resources designed to give you more control over your Facebook experience."[462] In short, Facebook, just like Google, has its own preferred and idiosyncratic way of defining privacy. For Facebook, privacy means giving users multiple controls and settings over profile and other information sharing on a feature-by-feature basis, which may be redesigned from time to time when the sheer number and complexity of these controls becomes overwhelming. Like Google, however, Facebook always reserves the right to weigh the need for any privacy controls against business objectives such as maximizing advertising revenues, and it too reaches these decisions behind closed doors.[463]

---

460. *See* Press Release, Facebook, Facebook Launches Additional Privacy Controls for News Feed and Mini-Feed (Sept. 8, 2006), http://www.marketwire.com/press-release/facebook-launches-additional-privacy-controls-news-feed-mini-feed-facebook-responds-775181.htm ("[The features] put control of who sees what information . . . directly into the hands of our users, just as they requested." (quoting Mark Zuckerberg, founder and CEO)). Chris Kelly used similar language when he testified before Congress for a second time two years later. *See Privacy Implications of Online Advertising: Hearing Before the S. Comm. on Commerce, Sci., and Transp.*, 110th Cong. 40, 41 (2008) (statement of Chris Kelly, Chief Privacy Officer, Facebook) ("Facebook follows two core principles: First, you should have control over your personal information . . . . Two, you should have access to the information that others want to share."). Elliot Schrage did the same in announcing Facebook's August 2009 response to the recommendations of the Canadian privacy regulators. *See Facebook Announces Privacy Improvements in Response to Recommendations by Canadian Privacy Commissioner*, FACEBOOK NEWSROOM (Aug. 27, 2009), http://newsroom.fb.com/News/194/Facebook-Announces-Privacy-Improvements-in-Response-to-Recommendations-by-Canadian-Privacy-Commissioner ("Our productive and constructive dialogue with the Commissioner's office has given us an opportunity to improve our policies and practices in a way that will provide even greater transparency and control for Facebook users." (quoting Elliot Schrage, Vice-President of Global Communications and Public Policy)). Zuckerberg most recently echoed this sentiment in a May 2010 op-ed announcing Facebook's plans to redesign its privacy controls. *See* Zuckerberg, *supra* note 422 ("If we give people control over what they share, they will want to share more.").

461. Zuckerberg, *supra* note 433.

462. *Id.*

463. *See* Joseph Bonneau & Sören Preibusch, *The Privacy Jungle: On the Market for Data Protection in Social Networks*, WEIS '09: PROCEEDINGS OF THE EIGHTH WORKSHOP ON THE

Given these behaviors of Google and Facebook, the fourth lesson, then, is that regulators wishing to embrace privacy by design must grapple with the inherent tensions between business models that seek to monetize personal data, and engineering and usability principles which, if properly implemented, tend to inhibit the collection and use of such data, and the balancing that companies undertake as part of their internal business processes. It follows that if regulators want privacy by design to be an effective means of improving consumer privacy, they must take at least two additional steps.

To begin with, regulators must ensure that when companies balance privacy design requirements against business objectives, they adhere to the well-established engineering and usability principles discussed throughout this Article. Because business and privacy demands often conflict, companies would benefit from regulatory clarity. Without well-defined guidelines about what it means to implement privacy by design, business considerations will always prevail over privacy: internal privacy champions will never have enough weight on their side to win the close calls.[464] In contrast, if regulators developed a reasonableness standard for designing privacy into products and services, companies would both know what was expected of them and take design requirements more seriously in achieving an appropriate balance.[465]

ECONOMICS OF INFORMATION SECURITY 1 (2009) (arguing that the "economically rational choice for a site operator is to make privacy control available to evade criticism from privacy fundamentalists, while obfuscating the privacy control interface and privacy policy to maximise sign-up numbers and encourage data sharing from the pragmatic majority of users"). Thus Bonneau and Preibusch claim that Facebook deploys "overly-complicated privacy settings with open defaults . . . [to] reduc[e] privacy complaints while still minimising salience." *Id.* at 31.

464. *See* Bamberger & Mulligan, *supra* note 5, at 274, 277.

465. The Proposed E.U. Regulation would require data controllers to "implement *appropriate* technical and organisational measures" for safeguarding personal data. *Proposed E.U. Regulation*, *supra* note 2, art. 23(1) (emphasis added). Similarly, In the United States, section 103 of the proposed Commercial Privacy Bill of Rights Act would have required:

> Each covered entity shall . . . implement a comprehensive information privacy program by—
>
> (1) incorporating necessary development processes and practices throughout the product life cycle that are designed to safeguard the personally identifiable information that is covered information of individuals based on—
>
> (A) the reasonable expectations of such individuals regarding privacy; and
>
> (B) the relevant threats that need to be guarded against in meeting those expectations . . . .

Commercial Privacy Bill of Rights Act of 2011, S. 799, 112th Cong. § 103 (2011).

Additionally, regulators should consider developing oversight mechanisms that would allow them to assess whether companies that claimed the mantle of privacy by design are adhering to the engineering and usability principles identified in this Article and related works. For example, they might require companies to maintain privacy design documents and, if appropriate, disclose them in the event of an investigation or lawsuit. Of course, disclosure is no magic bullet. Requiring disclosure after the fact may have less effect on the way that companies make privacy decisions than on how they discuss and document them.[466] It worth noting, however, that firms, U.S. regulators, and European regulators have already begun experimenting with maintaining privacy-related documentation,[467] which might be easily extended to cover "design documents"[468] as well.

V. CONCLUSION

Privacy regulators in both the United States and Europe are placing great faith in privacy by design as they set out to reform existing privacy regimes

466. We thank Tal Zarsky for sharing this observation. For discussion of effective transparency policies, see generally ARCHON FUNG ET AL., FULL DISCLOSURE: THE PERILS AND PROMISE OF TRANSPARENCY (2007). For additional discussion of regulatory approaches to privacy by design, see Rubinstein, *supra* note 1, at 1444–53; Ira S. Rubinstein, *Privacy and Regulatory Innovation: Moving Beyond Voluntary Codes*, 6 I/S: J.L. & POL'Y INFO. SOC'Y 355 (2011).

467. As previously noted, this is already the case for Google. *See supra* note 310 and accompanying text (describing Google's voluntary pledge to maintain privacy design documents for internal purposes). In the United States, the FTC consent decrees with both Google and Facebook obligate them to develop comprehensive privacy programs and to conduct third-party audits certifying that these programs satisfy the requirements of the FTC order, while maintaining pertinent records, which extends to "all materials relied upon . . . including but not limited to all plans, reports, studies, reviews, audits, audit trails, policies, training materials, and assessments." *See* Google Settlement, *supra* note 25, at 6; Facebook Settlement, *supra* note 25, at 7–8. In Europe, Article 25 of the Proposed E.U. Regulation introduces the obligation for controllers and processors to maintain documentation of the processing operations under their responsibility. *Proposed E.U. Regulation*, *supra* note 2, art. 25. If read in conjunction with Article 23 (data protection by design and default), this documentation requirement arguably covers "design documents." *Id.*, art. 23.

468. For UX specialists, "design documents" address alternative designs considerations in the form of mockups, wireframes, presentations, etc. *See Design Documents in Programming Methodology*, EXFORSYS INC. (Sept. 17, 2006), http://www.exforsys.com/tutorials/programming-concepts/design-documents-in-programming-methodology.html ("[T]he design document gives in a nutshell the main idea and structure of the product that would be developed by developers."). For example, an early mockup might have a button clicked on instead of off. More broadly, design documents in engineering might include information about every version of code stored in a code repository including comments, code changes, authors, date and time.

and make them more protective of consumers. This Article's goal has been to show what privacy by design might entail by undertaking a counterfactual analysis of ten privacy incidents. These incidents included five from Google—Gmail, Search, Street View, Buzz, and recent changes to its privacy policy; and five from Facebook—News Feed, Beacon, Facebook Apps, Photo Sharing, and recent changes to its privacy policies and settings. Using engineering and usability principles and practices derived from the research literature and described in Section II.B, we determined that each of these incidents might have been avoided if Google and Facebook had followed these principles and practices. Moreover, we described in specific detail what the two companies might have done differently in each of the ten cases.

This Article also explored the strengths and weaknesses of FIPs as the basis for privacy engineering and repeatedly emphasized the need to complement a FIPs-based engineering approach with engineering and usability principles and an extension of such principles to address the "social dynamics" of privacy. It explored the connections between privacy and existing design processes, such as UX design, which focus on usability. It also provided a more detailed look at privacy design pitfalls and guidelines inspired by the work of Altman and Nissenbaum. Sections III.A and III.B offered ten case studies and counterfactual analyses, which found that privacy engineering and usable privacy design were highly relevant to evaluating and overcoming a range of privacy problems including emergent issues affecting social networking services. The Article closed with a few modest lessons for regulators, which should be heeded if privacy by design is to achieve its promise of improving consumer privacy.