

35:4 BERKELEY TECHNOLOGY LAW JOURNAL

2020

Pages

911

to

1456

Berkeley Technology Law Journal
Volume 35, Number 4
Annual Review

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.

Copyright © 2020 Regents of the University of California.
All Rights Reserved.

Berkeley Technology Law Journal
University of California
School of Law
3 Law Building
Berkeley, California 94720-7200
editor@btlj.org
<https://www.btlj.org>



BERKELEY TECHNOLOGY LAW JOURNAL

VOLUME 35

NUMBER 4: ANNUAL REVIEW

2020

TABLE OF CONTENTS

ARTICLES

FOREWORD	911
<i>Julea Lipitz & Miranda Rutherford</i>	
A FRAMEWORK FOR PATENT EXHAUSTION OF SELF-REPLICATING TECHNOLOGIES	917
<i>Liz Freeman Rosenzweig</i>	
CAPRICIOUS PATENT BOUNDARIES: THE “RELIBERALIZATION” OF POST- ISSUANCE PATENT CLAIM AMENDMENTS	961
<i>Julien Crockett</i>	
RE-CLASSIFYING GOVERNMENTAL PETITIONERS AS “PERSONS” IN AIA REVIEW PROCEEDINGS	1003
<i>Artin Au-Yeung</i>	
FOURTH ESTATE AS A VEHICLE TO IMPEL REGISTRATION AND LIMIT ORPHAN WORKS?	1041
<i>Kevin Han Yang</i>	
KEEPING THE DMCA AWAY FROM FUNCTIONAL USE	1067
<i>Madison Bower</i>	
GOING BEYOND THE MUSIC MODERNIZATION ACT: CREATION IN THE DIGITAL ERA	1093
<i>Shreya M. Santhanam</i>	
RIMINI STREET V. ORACLE AND THE PROBLEM OF HIGH TRANSACTIONAL COSTS IN COPYRIGHT LITIGATION	1123
<i>Alistair McIntyre</i>	
DO ANDROIDS DREAM OF COPYRIGHT?: EXAMINING AI COPYRIGHT OWNERSHIP	1151
<i>Gia Jung</i>	
THE BREWING BATTLE: COPYRIGHT VS. LINKING	1179
<i>Marta Rocha</i>	

FLAUNTING THE SCARLET LETTER: CONSUMER REGULATION OF TRADEMARK MORALITY AFTER <i>LANCU V. BRUNETTI</i>	1213
<i>Kyung-Lee Kelly Go</i>	
THE GDPR: A RETROSPECTIVE AND PROSPECTIVE LOOK AT THE FIRST TWO YEARS.....	1245
<i>Erin Hilliard</i>	
EXPLAINING OPAQUE AI DECISIONS, LEGALLY.....	1291
<i>Walter A. Mostony</i>	
WHAT WE DON'T KNOW THEY KNOW: WHAT TO DO ABOUT INFERENCES IN EUROPEAN AND CALIFORNIA DATA PROTECTION LAW.....	1331
<i>Allan E. Holder</i>	
DEFINING THE PRIVACIES OF LIFE: LOWER COURT TRENDS IN THE WAKE OF <i>CARPENTER</i>	1365
<i>Tiffany Chen</i>	
BIPA AND ARTICLE III STANDING: ARE NOTICE AND CONSENT MORE THAN “BARE PROCEDURAL” RIGHTS?.....	1391
<i>Carmen Sobczak</i>	
“ILLINOIS BRICK-BREAKER” FOR SALE IN THE APP STORE: <i>APPLE V.</i> <i>PEPPER</i> AND THE NEED FOR A NEW ANTITRUST STANDING DOCTRINE.....	1427
<i>Haley Johnson</i>	

SUBSCRIBER INFORMATION

The *Berkeley Technology Law Journal* (ISSN1086-3818), a continuation of the *High Technology Law Journal* effective Volume 11, is edited by the students of the University of California, Berkeley, School of Law 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., 20th ed. 2015). Please cite this issue of the *Berkeley Technology Law Journal* as 35 BERKELEY TECH. L.J. ____ (2020).

BTLJ ONLINE

The full text and abstracts of many previously published *Berkeley Technology Law Journal* articles can be found at <https://www.btlj.org>. Our site also contains a cumulative index; general information about the *Journal*; the BTLJ Blog, a collection of short comments and updates about new developments in law and technology written by BTLJ members; and *BTLJ Commentaries*, an exclusively online publication for pieces that are especially time-sensitive and shorter than typical law review articles.

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 Scholastica 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 Scholastica submission website can be found at <https://btlj.scholasticahq.com/for-authors>.

Citations. All citations should conform to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (Columbia Law Review Ass'n et al. eds., 21st ed. 2020).

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

FENWICK & WEST LLP

ORRICK, HERRINGTON &
SUTCLIFFE LLP

WHITE & CASE LLP

Benefactors

BAKER BOTTS LLP

MORRISON & FOERSTER LLP

COOLEY LLP

PAUL HASTINGS LLP

COVINGTON & BURLING LLP

POLSINELLI LLP

FISH & RICHARDSON P.C.

SIDLEY AUSTIN LLP

JONES DAY

WEIL, GOTSHAL & MANGES LLP

KIRKLAND & ELLIS LLP

WILMER CUTLER PICKERING HALE
AND DORR LLP

LATHAM & WATKINS LLP

WILSON SONSINI GOODRICH &
ROSATI

MCDERMOTT WILL & EMERY

WINSTON & STRAWN LLP

Corporate, Government, Individual, and Foundation Sponsors

ATLASSIAN

KILBURN STRODE LLP

CORNERSTONE RESEARCH

LITNOMICS

DARTS IP

MARKS & CLERK LAW LLP

DORSEY & WHITNEY

MICROSOFT CORPORATION

FUTURE OF PRIVACY FORUM

NERA ECONOMIC CONSULTING

ROBERT GLUSHKO

PALANTIR

GOOGLE, INC.

PHARMACEUTICAL RESEARCH AND
MANUFACTURERS OF AMERICA

H. WILLIAM HARLAN

PWC

HICKMAN PALERMO BECKER
BINGHAM LLC

QUALCOMM

INTEL

VIA LICENSING CORP

INVENTIONSHARE

VINYL

WESTERN DIGITAL

Members

ANJIE LAW FIRM	KILPATRICK TOWNSEND & STOCKTON LLP
BAKER & MCKENZIE LLP	KNOBBE MARTENS OLSON & BEAR LLP
BEIJING EAST IP	MORGAN, LEWIS & BOCKIUS LLP
CROWELL & MORING	ROBINS KAPLAN LLP
DESMARAIS LL	ROPES & GRAY LLP
DURIE TANGRI LLP	SIMPSON THACHER & BARTLETT LLP
GREENBERG TRAURIG	TENSEGRITY LAW GROUP LLP
GTC LAW GROUP LLP & AFFILIATES	TROUTMAN SANDERS LLP
HAYNES AND BOONE, LLP	VAN PELT, YI & JAMES LLP
HOGAN LOVELLS, LLP	WANHUIDA INTELLECTUAL PROPERTY
IRELL & MANELLA LLP	WEAVER AUSTIN VILLENEUVE & SAMPSON LLP
KEKER VAN NEST & PETERS LLP	WILKIE FARR & GALLAGHER LLP
WOMBLE BOND DICKINSON LLP	

BOARD OF EDITORS

2019–2020

Executive Board

Editor-in-Chief
CHELSEA ANDRE

Senior Articles Editors
LESLIE DIAZ
CRISTINA MORA
COURTNEY REED

Senior Executive Editor
SAVANNAH CARNES

Managing Editor
AISLINN SMALLING

Senior Production Editor
MEGAN MCKNELLY

Senior Scholarship Editors
DANIEL CHASE

Senior Annual Review Editors
JULEA LIPIZ
MIRANDA RUTHERFORD

Senior Online Content Editor
CONCORD CHEUNG

Editorial Board

Production Editors
ANGELA GRIGGS
JANELLE LAMB
EMILY ROBERTS
HAILEY YOOK

Symposium Editor
ARMBIEN SABILLO

Online Content Editor
GINETTA SAGAN

Technical Editors
MADISON BOWER
MIN JUNG “MJ” HAN
ZACK JACOBS
RACHEL WILSON

Annual Review Editors
KRISTINA KRASNIKOVA
KEVIN YANG

Submissions Editors
CHRISTINA CROWLEY
DAVID FANG
MEHTAB KHAN

Notes & Comments Editors
HARRISON GERON
ALLAA MAGEID

Podcast Editor
ALLAN HOLDER

Web & Technology Editor
KARNIK HAJJAR

LLM Editor
IGOR SILVA

Alumni Relations Editor
NICK CALCATERRA

External Relations Editor
ASHLEIGH LUSSENDEN

Member Relations Editor
MICHELLE ZIPERSTEIN

Articles Editors
MUHTADI CHOUDHURY
MATTHEW CHUNG
SHWETA DUGGAL
JASON FRANCIS

Articles Editors
KELLY GO
EMMA LEE
WALTER MOSTOWY

JOSH SEDGWICK
CARMEN SOBCZAK
MARTA STUDNICKA
MEI XUAN

MEMBERSHIP

Vol. 35 No. 4

Associate Editors

ELIZABETH FU	NOAHLANI	JENNY QUANG
LOC HO	LITWINSELLA	ANDY ZACHRICH

Members

LIAM AARTASH	CALVIN HANNAGAN	MAXIMIN ORSERO
SHAHAD ALFAWAZ	ALEXANDRA HARVEY	JOSHUA PARZIVAND
BADER ALSHABANAT	ELIZABETH HECKMANN	NINA POUGET
TAIT ANDERSON	SOONYOUNG HEO	GAYATRI RAGHUNANDAN
CHRISTOPHER BARCLAY	JENNIFER HEWITT	EMILY ROBERTS
JOHN BATOHA	THOMAS HORN	FABIOLA ROSSY SEPTYA
MAYA BAUMER	JEFFREY JACOBSEN	CHRISTINA
MARGERITE BLASÉ	TOM JAMES	EILEEN SANFORD
RASMUS BLOM	CARTER JANSEN	SHEETAL SARAN
CONNOR BOEHM	ANJANAYE JARIWALA	YEMAJ SHEIK
VERONICA BOGNOT	GIA JUNG	ZIYU SHI
JONATHAN CHACON	PHILIP KATZ	DAKOTA SNEED
SUSIE CHEN	IAN KELLY	TAYLOR TAM
KEVIN J. CHEN	YEJI KIM	THERESA TAN
JENNIFER CHUNG	GRACE HO JUNG KIM	AMREEN TANEJA
HENDRIK COPPOOLSE	JOSEPH KROON	RACHEL TERRELL-PERICA
NATALIE CRAWFORD	TZU-I LEE	ELHITA THAMPURAN
KATHARNE CURRAULT	YI SHYUAN LEE	RACHEL THOMPSON
JAMESON DAVIS	GASPARE LODERER	MICKEALA TU
CHENXI DUAN	ANNE LUQUETTE	BLAINE VALENCIA
EVAN ENZER	MARGARET LYNCH	ARPAWAN WAEN VEJJAJIVA
ANUJ EZEKIEL	CHRISTIAN MCFALL	DEREK WEST
JOSHUA FRANK	GRACE MCFEE	CRISTINA WHITE
LIZ FREEMAN ROSENZWEIG	MEET MEHTA	MELODY WONG
ADITI GHATLIA	MEHREEN MIR	BILLY WU
BEN GOLDFEIN	ERIN MOORE	YEXI XU
EDWIN JESUS GONZALEZ	JACKSON MORAWSKI	MEI XUAN
JAKE GORHAM	DEBBIE MOSLEY	JOSHUA YOO
ISHA GULATI	GODHULLI NANDA	CHENZHAO YU
SALONI GUPTA	DAN NOEL	MICHELLE ZHOU

BTLJ ADVISORY BOARD

JIM DEMPSEY
*Executive Director of the
Berkeley Center for Law & Technology*
U.C. Berkeley School of Law

ROBERT C. BERRING, JR.
Walter Perry Johnson Professor of Law, Emeritus
U.C. Berkeley School of Law

MATTHEW D. POWERS
Tensegrity Law Group, LLP

JESSE H. CHOPER
Earl Warren Professor of Public Law, Emeritus
U.C. Berkeley School of Law

PAMELA SAMUELSON
*Richard M. Sherman Distinguished Professor of
Law & Information and Faculty Director of
Berkeley Center for Law & Technology*
U.C. Berkeley School of Law

REGIS MCKENNA
Chairman and CEO
Regis McKenna, Inc.

LIONEL S. SOBEL
*Professor of Law, Emeritus and Director of the
International Entertainment & Media Law
Summer Program in London*
Southwestern University School of Law

PETER S. MENELL
*Koret Professor of Law and Faculty
Director of the Berkeley Center
for Law & Technology*
U.C. Berkeley School of Law

LARRY W. SONSINI
Wilson Sonsini Goodrich & Rosati

ROBERT P. MERGES
*Wilson Sonsini Goodrich & Rosati Professor of
Law and Faculty
Director of the Berkeley Center
for Law & Technology*
U.C. Berkeley School of Law

MICHAEL STERN
Cooley LLP

DEIRDRE K. MULLIGAN
*Assistant Professor and Faculty Director of the
Berkeley Center for Law & Technology*
U.C. Berkeley School of Information

MICHAEL TRAYNOR
Cobalt LLP

JAMES POOLEY
James Pooley, PLC

THOMAS F. VILLENEUVE
Gunderson Dettmer Stough Villeneuve
Franklin & Hachigian LLP

BERKELEY CENTER FOR LAW & TECHNOLOGY 2019–2020

Executive Director

JIM DEMPSEY

Faculty Directors

KENNETH A. BAMBERGER	ORIN KERR	PAUL SCHWARTZ
CATHERINE CRUMP	PETER S. MENELL	ERIK STALLMAN
CATHERINE FISK	ROBERT P. MERGES	JENNIFER M. URBAN
CHRIS HOOFNAGLE	DEIRDRE K. MULLIGAN	MOLLY S. VAN HOUWELING
SONIA KATYAL	ANDREA ROTH	REBECCA WEXLER
	PAMELA SAMUELSON	

Fellow

KATHRYN HASHIMOTO	CHRISTINA KONNINGISOR
-------------------	-----------------------

Staff

MARK COHEN	RICHARD RISK
NATHALIE COLETTA	MATTHEW RAY
JANN DUDLEY	IRYS SCHENKER

FOREWORD

Julea Lipiz[†] & Miranda Rutherford^{††}

The Annual Review is a yearly publication of the *Berkeley Technology Law Journal* that provides a summary of many of the year's major developments at the intersection of law and technology. Our aim is to provide a valuable resource for judges, policymakers, practitioners, students, and scholars. Each Note provides a primer on a particular area of law, a development in that area of law, and commentary on that development.

The sixteen Notes in this Issue continue a tradition of covering a wide range of topics. The Notes address developments in patent, copyright, trademark, privacy, and antitrust law.

I. PATENT LAW

The first Note¹ in this Section examines how patent exhaustion applies to self-replicating technologies, with a focus on plants. The Court narrowly addressed this question in the 2013 case, *Bowman v. Monsanto*,² but explicitly left open whether and how exhaustion applies to self-replicating technologies more broadly. The Note uses *Bowman*'s theory and discussion of exhaustion, from the recent *Impression Products v. Lexmark International*³ decision, to create an exhaustion framework for self-replicating technology. This framework efficiently addresses the questions left unanswered in *Bowman* and can be used for future self-replicating exhaustion questions.

The second Note⁴ in this Section investigates whether “re-liberalizing” post-issuance patent claim amendments is beneficial for the patent system, focusing on America Invents Act trials. The Note probes the history of the claim amendments, analyzing the justifications given in three recent events that typify this “re-liberalization” trend—the Federal Circuit’s decision in *Aqua*

© 2020 Julea Lipiz & Miranda Rutherford.

† Senior Student Production Editor, *Berkeley Technology Law Journal*; J.D., 2020, University of California, Berkeley, School of Law.

†† Senior Student Production Editor, *Berkeley Technology Law Journal*; J.D., 2020, University of California, Berkeley, School of Law.

1. Liz Freeman Rosenzweig, Note, *A Framework for Patent Exhaustion of Self-Replicating Technologies*, 35 *BERKELEY TECH. L.J.* 917 (2020).

2. 569 U.S. 278 (2013).

3. 137 S. Ct. 1523 (2017).

4. Julien Crockett, *Capricious Patent Boundaries: The “Reliberalization” of Post-Issuance Patent Claim Amendments*, 35 *BERKELEY TECH. L.J.* 961 (2020).

Products v. Matal,⁵ the Patent and Trademark Office’s embrace of the *Phillips* standard, and the Patent and Trademark Office’s Motion to Amend Pilot Program—and collecting data regarding the impact of this “re-liberalization” on various actors within the patent system. The Note concludes by arguing that the “re-liberalization” of post-issuance claim amendments within America Invents Act trials hurts the patent system.

The third and final Note⁶ in this Section discusses the Supreme Court’s recent decision in *Return Mail v. United States Postal Service*.⁷ The Note ultimately concludes that the Court’s decision was incorrect, because it incorrectly neglected to apply the PTO’s administrative interpretations of reexamination to AIA review proceedings. The Note also suggests how the PTO could rectify this decision via further administrative rules.

II. COPYRIGHT LAW

The first Note⁸ in this Section analyzes the consequences of the Court’s holding in *Fourth Estate v. Wall-Street.com*,⁹ which finds that plaintiffs must file applications for copyright registration and receive a decision from the Copyright Office, before filing suit. It argues that this harsher approach, of requiring registration, serves as a motivator for registration. This Note concludes that this stricter standard can be combined with academic proposals that call for a required recording of copyright transfer with the Copyright Office, to help mitigate the orphan works problem.

The second Note¹⁰ in this Section examines anticircumvention provisions within the DMCA and finds that a key piece of copyright law, established in *Baker v. Selden*,¹¹ has been forgotten. *Baker*¹² finds that copyright protections do not extend to functional aspects of a work and this Note argues that courts should inquire whether an accused infringer was accessing functionality, as opposed to copying expressive works. As a more feasible solution, the Note argues that a statutory exception should be added to anticircumvention provisions, as to clarify the law and protect non-infringing uses.

5. 872 F.3d 1290 (Fed. Cir. 2017).

6. Artin Au-Yeung, Note, *Re-Classifying Governmental Petitioners as “Persons in AIA Review Proceedings*, 35 BERKELEY TECH. L.J. 1003 (2020).

7. 1389 S. Ct. 1853 (2019).

8. Kevin Han Yang, Note, *Fourth Estate As a Vehicle to Impel Registration and Limit Orphan Works*, 35 BERKELEY TECH. L.J. 1041 (2020).

9. 139 S. Ct. 881 (2019).

10. Madison Bower, Note, *Keeping the DMCA Away From Functional Use*, 35 BERKELEY TECH. L.J. 1067 (2020).

11. 101 U.S. 99 (1879).

12. *Id.*

The third Note¹³ in this Section analyzes the Music Modernization Act (MMA)¹⁴ and finds that the act realigns music copyright for those within the music industry, but fails to address a gap in music copyright for those not in the music industry. Video and music sharing platforms act as arbiters of online content and, due to online enforcement realities, readily take down works with copyrighted work, even when the work constitutes fair use. This Note argues that non-commercial mashups and covers constitute fair use, and there should be a statutory presumption of fair use for these types of work, which will allow them to remain on internet platforms and further the goals of copyright law.

The fourth Note¹⁵ in this Section examines *Rimini Street v. Oracle*.¹⁶ The Note looks at American fee shifting and judicial interpretations leading to *Rimini*,¹⁷ and analyzes how cost-shifting affected copyright litigation pre-*Rimini*. Using this background, the Note concludes that *Rimini*¹⁸ was correctly decided, but that legislation should address the heavy transactional costs remaining, so that copyright law can better achieve its goals.

The fifth Note¹⁹ in this Section analyzes the issue of copyright in AI-created works. As artificial intelligence (AI) becomes more sophisticated and extends into the generation of artistic works like music, poetry, novels, paintings, and sculpture, the issue of whether that work can be copyrighted, and by whom, has yet to be answered. This Note argues that AI works can surmount the traditional hurdles of copyrightability and should be protected by a unique regime that incorporates accountability and administrability. By mapping out legal precedent, underlying policies, and recent scholarly debate around copyrightability, this Note reveals and advocates for a continuing willingness to adapt the law to technological advances.

The sixth and final Note²⁰ in this Section discusses copyright protections in embedded content online. Ultimately, the Note concludes that the negative consequences of copyright protection for linked content outweigh its positive

13. Shreya M. Santhanam, Note, *Going Beyond the Music Modernization Act: Creation in the Digital Era*, 35 BERKELEY TECH. L.J. 1093 (2020).

14. Orrin G. Hatch-Bob Goodlatte Music Modernization Act, Pub. L. No. 115-264, 132 Stat. 3676 (2018).

15. Alistair McIntyre, Note, *Rimini Street v. Oracle and the Problem of High Transactional Costs in Copyright Litigation*, 35 BERKELEY TECH. L.J. 1123 (2020).

16. 139 S. Ct. 873 (2019).

17. *Id.*

18. *Id.*

19. Gia Jung, Note, *Do Androids Dream of Copyright?: Examining AI Copyright Ownership*, 35 BERKELEY TECH. L.J. 1151 (2020).

20. Marta Rocha, Note, *The Brewing Battle: Copyright vs. Linking*, 35 BERKELEY TECH. L.J. 1179 (2020).

effects, and proposes a licensing scheme to compensate creators without chilling linking.

III. TRADEMARK LAW

The Note²¹ in this Section examines the Court's holding in *Iancu v. Brunetti*,²² which found the "immoral or scandalous" provision of the Lanham Act to be unconstitutional. It argues that the Court reached the correct conclusion and that consumers, not Congress, should be entrusted with regulating future trademark morality, since consumers have the power to determine which words can be used as trademarks in the marketplace.

IV. PRIVACY

The first Note²³ in this Section provides a quantitative summary of the European Union's General Data Protection Regulation (GDPR) and provides an exhaustive Internet search of all GDPR fines and sanctions imposed and pending by EU and EEA nations from May 25, 2018 (the day the GDPR became enforceable), through March 31, 2020. A data set of publicly available enforcements was compiled for all twenty-seven E.U. countries, the United Kingdom, and the three EEA countries—Lichtenstein, Iceland, and Norway, which are also subject to the GDPR. This Note uses descriptive statistics (mean, median, mode, range, and linear regression) to analyze the data and present results, including noteworthy findings, patterns, and trends of enforcement since implementation.

The second Note²⁴ in this Section examines Article 22 of the GDPR and its "right to contest" artificial intelligence (AI) automated decision-making. This Note provides a technical overview of "opaque AI," including the technological tools currently available to help explain AI behavior and what it might mean to explain an AI. This understanding is used to examine Article 22's "right to contest" and argues that it grants a right to an ex post explanation of AI decisions. The Note also looks forward and analyzes due process issues and recommends explanations that are most likely to fulfill the obligations listed in Article 22.

21. Kyung-Lee Kelly Go, Note, *Flaunting the Scarlett Letter: Consumer Regulation of Trademark Morality After Iancu v. Brunetti*, 35 BERKELEY TECH. L.J. 1213 (2020).

22. 139 S. Ct. 2294 (2019).

23. Erin Hilliard, Note, *The GDPR Two Years In: A Retrospective and Prospective Look*, 35 BERKELEY TECH. L.J. 1245 (2020).

24. Walter A. Mostowy, Note, *Explaining Opaque AI Decisions: How to Satisfy the GDPR's Right to an Ex Post Explanation*, 35 BERKELEY TECH. L.J. 1291 (2020).

The third Note²⁵ of this Section analyzes data inferences, which are information derived or predicted from pre-existing data. Inferences are generated and used by businesses and other data controllers in great amounts, and they have become the type of data that drives most of the digital economy. This Note concludes that the paradigmatic data protection statutes in the European Union and the United States—the General Data Protection Regulation and the California Consumer Privacy Act, respectively—leave concerning gaps in the protection of this type of data. Consequentially, the Note argues that inferences deserve the same protection as the most sensitive types of data receive in both statutes, as data inferences were among the primary reasons both statutes were enacted.

The fourth Note²⁶ of this Section surveys lower court decisions in the wake of *Carpenter v. United States*.²⁷ The Note identifies that lower courts have declined to expand *Carpenter*'s holding past cell site location information. However, the Note suggests that lower courts may need to consider broadening *Carpenter* as new technologies present even more sophisticated opportunities for law enforcement surveillance.

The fifth and final Note²⁸ of this Section analyzes the standing requirement under Illinois's Biometric Information Privacy Act (BIPA).²⁹ Courts have disagreed as to whether plaintiffs can sue for a mere procedural violation of BIPA, without an additional showing of harm. The Note argues that BIPA provides a concrete interest in controlling one's own private information, and as such, additional concrete harm should not be necessary for standing.

V. ANTITRUST

The Note³⁰ in this Section analyzes the Court's decision in *Apple Inc. v. Pepper*³¹ and argues that the "indirect purchaser rule" from *Illinois Brick*³² is not a suitable tool for analyzing antitrust standing in two-sided marketplaces. The

25. Allan E. Holder, Note, *What We Don't Know They Know: What to Do About Inferences in European and California Data Protection Law*, 35 BERKELEY TECH. L.J. 1331 (2020).

26. Tiffany Chen, Note, *Defining the Privacies of Life: Lower Court Trends in the Wake of Carpenter*, 35 BERKELEY TECH. L.J. 1365 (2020).

27. *Carpenter v. United States*, 138 S. Ct. 2206 (2018).

28. Carmen Sobczak, Note, *BIPA and Article III Standing: Are Notice and Consent More Than "Bare Procedural" Rights?*, 35 BERKELEY TECH. L.J. 1391(2020).

29. 740 ILL. COMP. STAT. 14/1 et seq.

30. Haley Johnson, Note, "Illinois Brick-Breaker" For Sale in the App Store: *Apple v. Pepper and the Need for a New Antitrust Standing Doctrine*, 35 BERKELEY TECH. L.J. 1427 (2020).

31. 139 S. Ct. 1514 (2019).

32. 431 U.S. 720 (1977).

Note proposes an alternative test that would be better suited to two-sided marketplaces and applies this test to *Apple*.³³

33. *Apple*, 139 S. Ct. at 1514.

A FRAMEWORK FOR PATENT EXHAUSTION OF SELF-REPLICATING TECHNOLOGIES

Liz Freeman Rosenzweig[†]

I. INTRODUCTION

A patent is a “right to exclude others from making, using, offering for sale, [] selling,” or importing the patented invention.¹ But that right has limits, one of which is the doctrine of patent exhaustion. Exhaustion extinguishes the patent owner’s rights of exclusion over the use, sale, and import of objects that have been the subject of a patentee-authorized sale.² The effect: the buyer can use, sell, and import the *particular object* that they bought without being liable for infringement.

There are two main levers that drive the operation of patent exhaustion: the *objects* over which exhaustion occurs and the *rights* that exhaustion affects. The key to the *objects* lever is that exhaustion applies only to tangible objects that embody the patented invention, not to the underlying intangible knowledge. The effect of this distinction is that exhaustion “restricts a patentee’s rights *only as to the ‘particular article’ sold . . .*”³ The key to the *rights* lever is that exhaustion applies to *using*, but not *making*. The effect of this distinction is that the application of exhaustion “leaves untouched the patentee’s ability to prevent a buyer from *making new copies* of the patented item.”⁴ The effect of both levers in concert is a balance between the rights of someone who purchases a tangible object against the rights of someone who

DOI: <https://doi.org/10.15779/Z38SF2MC66>

© 2020 Elizabeth S. Freeman Rosenzweig.

[†] Ph.D., Stanford University, Department of Biology, 2017; J.D., University of California, Berkeley, School of Law, 2020. Special thanks are due to Talha Syed, Kenneth Bamberger, Barbara McClung, Julea Lipiz, Erin Delaney, and Artin Au-Yeung for providing commentary, feedback, and editorial assistance. Dedicated to the memory of Nicholas E. Calcaterra, Ph.D., J.D., dear friend, classmate, patent attorney, and early reader of this work.

1. 35 U.S.C. § 154(a)(1) (2018).

2. *See* *Impression Prods. v. Lexmark Int’l, Inc.*, 137 S. Ct. 1523, 1526 (2017) (stating that exhaustion occurs “[w]hen a patentee sells an item,” making “that product ‘[] no longer within the limits of the patent monopoly’ and instead [] the ‘private, individual property’ of the purchaser”) (original alterations excluded) (quoting *Bloomer v. McQuewan*, 55 U.S. 539, 549–550 (1853)).

3. *Bowman v. Monsanto Co.*, 569 U.S. 278, 284 (2013) (emphases added) (citation omitted).

4. *Id.*

owns a patent covering the intangible knowledge embodied in that tangible object.

But the application of these two distinctions—between the tangible and the intangible, and between using and making—appears to break down in the context of a crucial segment of the modern invention economy: self-reproducing technologies.⁵ Such technologies, as the name suggests, are inherently able to make more copies of themselves. They therefore present a complex twist on the application of the doctrine of patent exhaustion. When an item has the ability to reproduce itself, the line between “using” (which is excused by exhaustion) and “making” (which is not), is quickly blurred.

The concept of self-reproducing technology sounds futuristic, conjuring images of “nanorobots or organic computers,”⁶ or even construction modules in outer space.⁷ But self-reproducing technologies are more common than that list implies. Many modern technologies—from vaccines,⁸ cell cultures,⁹ and recombinant protein production systems,¹⁰ to even “‘living’ cement”¹¹—contain DNA or RNA, enabling self-reproduction.¹² Yet the quintessential

5. This Note uses the terms “self-replicating” and “self-reproducing” interchangeably.

6. Jeremy N. Sheff, *Self-Replicating Technologies*, 16 STAN. TECH. L. REV. 229, 230 (2013).

7. See Tom Kalil, *Bootstrapping a Solar System Civilization*, OBAMA WHITE HOUSE BLOG ARCHIVES (Oct. 14, 2014), <https://obamawhitehouse.archives.gov/blog/2014/10/14/bootstrapping-solar-system-civilization>.

8. See, e.g., Karl Ljungberg & Peter Liljeström, *Self-Replicating Alphavirus RNA Vaccines*, 14 EXPERT REV. VACCINES 177–194 (Oct. 1, 2014), <https://doi.org/10.1586/14760584.2015.965690> (describing self-replicating vaccines).

9. See, e.g., ThermoFisher Scientific, *Introduction to Cell Culture*, TECH. REFERENCE LIBR., <https://www.thermofisher.com/us/en/home/references/gibco-cell-culture-basics/introduction-to-cell-culture.html> (last visited Dec. 13, 2019).

10. For example, the vast majority of rennet (a protein used to make cheese) and insulin (for diabetes treatment) used today are recombinantly produced in industrial microbial bioreactors. Valentin Waschulin & Liz Specht, *Cellular Agriculture: An Extension of Common Production Methods for Food*, GOOD FOOD INST. 1, 5, 10 (Mar. 6, 2018), <https://www.gfi.org/images/uploads/2018/03/Cellular-Agriculture-for-Animal-Protein.pdf>.

11. See Robert F. Service, *From ‘Living’ Cement to Medicine-Delivering Biofilms*, *Biologists Remake the Material World*, SCIENCE (Feb. 18, 2020, 1:50 PM), <https://www.sciencemag.org/news/2020/02/living-cement-medicine-delivering-biofilms-biologists-remake-material-world#>.

12. Not everything that contains DNA or RNA is necessarily capable of self-reproduction, however. For example, engineers recently 3D-printed a plastic bunny containing DNA that encodes instructions for 3D printing more bunnies, but these bunnies do not themselves read and carry out the instructions to reproduce *on their own*. Loyal Liverpool, *3D-Printed Bunny Contains DNA Instructions to Make a Copy of Itself*, NEW SCI. (Dec. 9, 2019), <https://www.newscientist.com/article/2226644-3d-printed-bunny-contains-dna-instructions-to-make-a-copy-of-itself/#ixzz680sF3YdD>.

self-reproducing technology is even more familiar and provides the basis for a large segment of the economy: the seed.

Despite these technologies' growing importance, there is currently no clear guidance for applying exhaustion to self-reproducing technologies. The Supreme Court has most recently addressed patent exhaustion in two cases: *Bowman v. Monsanto Co.*¹³ and *Impression Prods. v. Lexmark Int'l, Inc.*¹⁴ *Bowman* addressed self-reproducing technologies specifically but issued a narrow holding, while *Lexmark* did not specifically address self-reproducing technologies but issued a broader holding. These two cases speak to different parts of the problem and have not yet been integrated against the background of a clear understanding of the tort elements of patent law. This Note does just that, providing the first comprehensive integration of modern exhaustion caselaw with self-reproducing technologies.

In *Bowman*, the Court held that a farmer's intentional replanting of patented seeds was not protected by exhaustion.¹⁵ But *Bowman* explicitly left open two questions about how the doctrine of exhaustion might apply to self-reproducing technologies more generally.¹⁶ Four years later, the Court in *Lexmark* held that exhaustion automatically applies in authorized sales of patented items, regardless of the presence of post-sale restrictions, and regardless of where the sale takes place.¹⁷ But *Lexmark* did not address or even reference *Bowman's* questions, which today remain apparently unanswered.¹⁸

In this Note, I argue that we already have all the tools to answer *Bowman's* questions: the analysis in *Bowman* implicitly suggests answers to its own questions, and the broad framework of *Lexmark* makes those answers clear. Specifically, *Bowman* deconstructs the core using-versus-making issue, and *Lexmark* puts exhaustion to work on that issue by broadly structuring exhaustion's theory and policy rationale. Put together against a firmly-rooted tort-based understanding of patent infringement, these two cases build a concise and robust framework—which I have termed “Bowmark”—for applying exhaustion to self-reproducing technologies in situations beyond *Bowman's* limited holding. This framework proves to be workable against *Bowman's* open questions and reveals that, to the extent that there are still gray

13. *Bowman v. Monsanto Co.*, 569 U.S. 278 (2013).

14. *Impression Prods. v. Lexmark Int'l, Inc.*, 137 S. Ct. 1523 (2017).

15. *Bowman*, 569 U.S. at 280.

16. *Id.* at 289.

17. *Lexmark*, 137 S. Ct. at 1535.

18. See James B. Kobak, Jr., *Lexmark*, *The Overruling of Mallinckrodt and The Future of Restraints on Alienation for Patented Goods*, 99 J. PAT. & TRADEMARK OFF. SOC'Y 609, 622 (2017).

areas regarding the role of patent law in the situations explicitly unaddressed by *Bowman*, those gray areas do not involve or invoke exhaustion.

The remainder of this Note is divided into five Parts. Part II summarizes the exhaustion doctrine. Part III walks through special aspects of protecting intellectual property (IP) in self-reproducing technology, focusing on plants. Part IV delves into the contexts and reasoning of *Bowman* and *Lexmark*. Part V unpacks *Bowman*'s questions using the *Bowmark* framework, working through how exhaustion could apply to self-reproducing technologies whose reproduction is (1) "outside the purchaser's control"¹⁹ or (2) is "a necessary but incidental step in using the item for another purpose."²⁰ Part V ultimately concludes that the exhaustion doctrine does not need to be altered for the sake of self-replicating technologies.

II. WHAT EXHAUSTION IS—AND WHAT IT ISN'T

Before dissecting *Bowman* and *Lexmark*, we must build a thorough foundation of what exhaustion is and is not. Exhaustion discourse is laden with archaic terms like "alienation of chattels,"²¹ which distract from the fact that exhaustion is simply a defense to patent infringement.²² Thus, to understand the proper application of exhaustion, infringement must first be clear.

A. EXHAUSTION FIRST REQUIRES AN INFRINGEMENT ANALYSIS

It is easy to forget that patent infringement is and "has always been a tort . . ." ²³ Indeed, courts have often been confused or "undisciplined"²⁴ about the tort dimensions of patent law, so it is worth explicitly walking through the standards for infringement.

19. *Bowman*, 569 U.S. at 289.

20. *Id.*

21. *See, e.g., Lexmark*, 137 S. Ct. at 1526 (quoting *Kirtsaeng v. John Wiley & Sons, Inc.*, 568 U.S. 519, 538 (2013)).

22. *See id.* at 1530. Note that "[w]hile the exhaustion doctrine has been developed in the context of utility patents," it also applies in the context of plant patents. Leah Chan Grinvald & Ofer Tur-Sinai, *Intellectual Property Law and the Right to Repair*, 88 FORDHAM L. REV. 63, 112 n.272 (Oct. 2019). *See, e.g., Int'l Fruit Genetics, L.L.C. v. Orcharddepot.com*, No. 4:17-CV-02905-JSW, slip op. at 4 (N.D. Cal. Feb. 12, 2018) (order denying motion to dismiss) (stating that exhaustion did not apply to the defendant's infringement of grapevines covered by plant patents because the plants at issue were licensed, not sold).

23. Saurabh Vishnubhakat, *Commil v. Cisco and the Tort of Patent Infringement*, WRITTEN DESCRIPTION (Dec. 9, 2014, 12:46 AM), <https://writtendescription.blogspot.com/2014/12/commil-v-cisco-and-tort.html>; *see also* PETER S. MENELL, MARK A. LEMLEY & ROBERT P. MERGES, *INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE*: 2017 368 (2017).

24. Vishnubhakat, *Commil v. Cisco*, *supra* note 23.

Infringement can be direct or indirect, but indirect infringement still requires “at least one act of direct infringement.”²⁵ Direct patent infringement is often called a “strict liability” offense²⁶ in that “innocent” infringement is still infringement,²⁷ and there is no requirement that the infringer intended or even knew of their own infringement.²⁸ That being said, there is still a minimum standard: the direct infringer must *act* in some way.²⁹ And that act must implicate the intangible knowledge of the patent.³⁰ For direct infringement, the “act” is making, using, selling, offering to sell, or importing a patented invention without authority from the patent holder.³¹

It is useful at this point to keep in mind Syed’s articulation of what a patent actually represents: patent rights, he argues, are ultimately tied to the *intangible knowledge* of tangible things, not to the tangible things themselves.³² Patents, after all, are a form of *intellectual* property. Tangible things are mere chattels embodying patented knowledge. Infringing a patent requires acting on that knowledge in an unauthorized manner.³³

25. Menell, Lemley & Merges, *supra* note 23, at 368.

26. Saurabh Vishnubhakat, *An Intentional Tort Theory of Patents*, 68 FLA. L. REV. 571, 573 (2016), <https://ssrn.com/abstract=2492200>.

27. See 5 Donald S. Chisum, *Chisum on Patents* § 16.02(2) (2019).

28. Vishnubhakat, *An International Tort Theory of Patents*, *supra* note 26, at 573.

29. Furthermore, the act must be manifested externally and committed voluntarily. The necessity in tort law for such an act as a prerequisite for liability is well-established. See Restatement (Second) of Torts § 2 (Am. Law Inst. 1965). Its continued necessity specifically within the context of intellectual property law is also well-established. See *ABC, Inc. v. Aereo, Inc.*, 573 U.S. 431, 453 (2014) (Scalia, J., dissenting).

30. See Talha Syed, *Dephysicalizing Patent Eligibility*, 9 (draft on file with the author) (“And similarly for patents: just as the underlying object of copyright’s tangibly-fixed expressive form remains an intangible form of expression, so the object of a tangibly-embodied knowledge space remains an intangible space of knowledge.”).

31. 35 U.S.C. § 271(a) (2018). Any one of those acts, if “performed by or attributable to a single entity,” constitutes direct infringement. Acts by another can be attributed to a single entity for direct infringement either vicariously or through joint enterprise. Vicarious liability requires that the entity “directs or controls” the other’s actions. *Akamai Techs., Inc. v. Limelight Networks, Inc.*, 797 F.3d 1020, 1022 (Fed. Cir. 2015) (citations omitted). Thus, there is no direct infringement (and no infringement at all) unless a single entity completes an entire infringing act themselves—either personally, vicariously, or jointly. See Menell, Lemley & Merges, *supra* note 23, at 368.

32. See Syed, *supra* note 30, at 4; see also *Impression Prods. v. Lexmark Int’l, Inc.*, 137 S. Ct. 1523, 1537 (2017) (distinguishing “the *article* and the *invention which it embodies*”) (emphases added) (citations omitted). Note that “intangible knowledge of tangible things” is *not* the same as abstract ideas or laws of nature, which the Court has held to not be patent eligible subject matter under 35 U.S.C. § 101. See Syed, *supra* note 30, at 5–6.

33. *E.g.*, unauthorized use, sale, import, or making. See Syed, *supra* note 30, at 10–11.

Once there is direct infringement, there may also be indirect infringement.³⁴ Unlike direct infringement, indirect infringement requires a heightened mental state regarding the infringement.³⁵ Indirect infringement includes induced infringement³⁶ and contributory infringement.³⁷

Induced infringement “involves behavior that omits any direct making, using, or selling of the patented invention, but that nevertheless amounts to an attempt to appropriate the value of the invention.”³⁸ It can encompass a broad range of activities and “is often described as activity that ‘aids and abets’ infringement.”³⁹ Liability for induced infringement “requires knowledge that the induced acts constitute patent infringement,”⁴⁰ or willful blindness thereto.⁴¹

Contributory infringement essentially involves providing components of patented inventions,⁴² while knowing “that the combination for which [the] component was especially designed was both patented and infringing,”⁴³ even if the components themselves do not infringe.⁴⁴

B. EXHAUSTION

Once there is infringement, exhaustion comes into play at the boundary of the tangible and the intangible. The “tangible” simply refers to a physical item that embodies the patent—the “chattel,” in common law parlance⁴⁵—and the “intangible” refers to the underlying *knowledge* that is the actual object of the patent.⁴⁶ Exhaustion applies to the *tangible*,⁴⁷ and covers the intangible

34. See Menell, Lemley, and Merges, *supra* note 23, at 385.

35. See *id.*

36. 35 U.S.C. § 271(b) (2018).

37. 35 U.S.C. § 271(c) (2018).

38. Menell, Lemley & Merges, *supra* note 23, at 385.

39. *Id.*

40. Glob.-Tech Appliances, Inc. v. SEB S.A., 563 U.S. 754, 766 (2011).

41. *Id.* at 768.

42. 35 U.S.C. § 271(c) (2018).

43. Aro Mfg. Co. v. Convertible Top Replacement Co., 377 U.S. 476, 488 (1964).

44. The component must also “constitut[e] a material part of the [patented] invention . . . and not [be] a staple article . . . of commerce suitable for substantial noninfringing use” 35 U.S.C. § 271(c) (2018); see also Menell, Lemley & Merges, *supra* note 23, at 395–96.

45. In terms of seeds, the tangible object would be an individual seed.

46. See Syed, *supra* note 30, at 4; see also United States v. Univis Lens Co., 316 U.S. 241, 251 (1942) (comfortably distinguishing between an “article and the invention which it embodies”). For seeds, the “intangible” would be the knowledge of any special traits encoded in the seed.

47. See Bowman v. Monsanto Co., 569 U.S. 278, 284 (2013) (stating that “the doctrine restricts a patentee’s rights *only as to the ‘particular article’ sold*”) (emphasis added) (internal quotation and citation omitted).

knowledge only to the extent that the intangible knowledge is incidental to using, making, selling, or importing the tangible.⁴⁸ Exhaustion “marks the point where patent rights yield to the common law principle against restraints on alienation”⁴⁹ of chattels. Applying exhaustion plucks the individual purchased item out of the patent’s monopoly,⁵⁰ allowing that tangible chattel to freely “flow[] through the market.”⁵¹ It can then be used or sold in ways that implicate the patented knowledge without the specter of infringement.

Exhaustion has traditionally not applied when the underlying infringing act is “making.”⁵² That may seem like an odd departure, but it makes sense in light of the policy against restraints on alienation of chattels:⁵³ One cannot reasonably be “restrained” from alienating oneself from a chattel that does not yet exist and may not necessarily ever exist.⁵⁴ More importantly, it makes sense insofar as the patented knowledge is essential to “making.”

III. SPECIAL ASPECTS OF PROTECTING INTELLECTUAL PROPERTY IN SELF-REPRODUCING TECHNOLOGIES

Seeds are the quintessential self-reproducing technology. When someone “uses” a seed to grow a plant, that plant will go on to provide the user exactly with which they started out: new seeds.⁵⁵ Seeds are a major driver of the economy. The commercial seed market in the United States was estimated to

48. But the patent monopoly will eventually expire.

49. *Impression Prods. v. Lexmark Int’l, Inc.*, 137 S. Ct. 1523, 1531 (2017).

50. *See id.* at 1531.

51. *Id.* at 1538.

52. *See Bowman*, 569 U.S. at 284 (first citing *Mitchell v. Hawley*, 83 U.S. 544, 548 (1873); then citing *Wilbur-Ellis Co. v. Kuther*, 377 U.S. 422, 424 (1964)).

53. *See Lexmark*, 137 S. Ct. at 1526.

54. *See* 3 DAVID A. THOMAS, THOMPSON ON REAL PROPERTY, Thomas Editions § 28.08 (2019) (stating the rule against perpetuities: “No interest is good unless it must vest, if at all, not later than twenty-one years after some life in being at the creation of the interest”) (citations omitted). Not all seeds necessarily germinate (“vest”). *See* Steve Knox, *Soybean Seed Germination Concerns*, U. NEB.—LINCOLN INST. AGRIC. AND NAT. RESOURCES: CROPWATCH (Feb. 13, 2019), <https://cropwatch.unl.edu/2019/soybean-seed-germination-concerns>.

55. Whether or not the next generation of seeds is genetically identical to the starting seeds will depend on the genetics of the system. For example, self-pollinating inbred plants will produce progeny that are “genetically identical to each other and to the inbred parent,” University of Nebraska—Lincoln Plant & Soil Sciences eLibrary, *Inbreeding, Hybrid Vigor, and Hybrid Corn*, CORN BREEDING: LESSONS FROM THE PAST, <http://passel.unl.edu/pages/informationmodule.php?idinformationmodule=1075412493&topicorder=9&maxto=12> (last visited Dec. 11, 2019), but plants grown from seeds produced from hybrids will not match their parents. *See* Home & Garden Information Center, *Fruits Not True to Type – Vegetables*, U. MD. EXTENSION, <https://extension.umd.edu/hgic/topics/fruits-not-true-type-vegetables> (last visited Dec. 11, 2019).

be \$5.7 billion in 1997,⁵⁶ and in 2017, \$132.8 billion of the gross domestic product came from farms.⁵⁷ Globally, the commercial seed market has been estimated to be worth roughly \$50 billion.⁵⁸ As such, this Note will focus on seeds as the dominant example of self-reproducing technology, but essentially the same analyses can be applied to other self-reproducing technologies.

A. EXHAUSTION AND SELF-REPRODUCING TECHNOLOGIES

The concept of exhaustion can quickly become confusing when applied to self-reproducing technologies. This is because exhaustion traditionally excuses a purchaser from infringement for *using* a patented item, but not for *making new copies* of that item.⁵⁹ But self-reproducing technologies, by their very nature, flow seamlessly through cycles of “using” and “making.”

So how can it be that exhaustion allows “use” but not “making”? Traditionally, the reason “use” is exhausted while “making” is not is because the intangible knowledge is not “incidental” to making: it is crucial. But is that also true when an item replicates *itself*? Some have argued⁶⁰ that, for self-reproducing technologies, “using” *is* “making,” and thus “making” should thus also be excused by exhaustion. On the other hand, one might just as easily flip that argument on its head and say that, for the same reason, exhaustion applies to *neither* use nor making of self-reproducing technologies. In that case, a legitimate purchaser would be left only with the ability to sell or obtain a license on the item. Both of these options are frustrating. Can one reasonably draw a boundary line somewhere between “use” and “making” in this context? These are exactly the questions that were swirling in the seed intellectual property space⁶¹ as the Supreme Court took up *Bowman*.

56. Jorge Fernandez-Cornejo, United States Department of Agriculture, *The Seed Industry in U.S. Agriculture: An Exploration of Data and Information on Crop Seed Markets, Regulation, Industry Structure, and Research and Development*, ECON. RES. SERV.: AGRIC. INFO. BULL. NO. 786 1, 7 (Feb. 2004).

57. U.S. Department of Agriculture (USDA), *What is Agriculture's Share of the Overall U.S. Economy?*, ECON. RES. SERV.: DATA PRODS. (Apr. 16, 2019), <https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=58270>. Note that this number includes both animal and plant farming. But even animal farming is an indirect output of plant crops, because livestock depend on crops.

58. See Sylvie Bonny, *Corporate Concentration and Technological Change in the Global Seed Industry*, 9 SUSTAINABILITY no. 1632 1, 4–5 (Sept. 14, 2017).

59. See *Bowman v. Monsanto Co.*, 569 U.S. 278, 287 (2013).

60. *Bowman* (unsuccessfully) attempted this argument. See *id.* at 287.

61. See, e.g., Jeremy N. Sheff, *Self-Replicating Technologies*, 16 STAN. TECH. L. REV. 229, 238 (2013); Michael R. Ward, Rachel Krevans & Matthew Chivvis, *Patent Exhaustion & Self-Replicating Technologies*, GENET. ENG. BIOTECHN. N. (Aug. 1, 2012), <https://www.genengnews.com/magazine/186/patent-exhaustion-self-replicating-technologies/>.

But before applying the exhaustion doctrine to seeds, it is important to first understand the special forms of IP protection available for plants in the United States, and how the agricultural market typically handles those rights.

B. AN OVERVIEW OF PLANT IP PROTECTION IN THE UNITED STATES

U.S. federal law provides three structures for plant IP protection: plant patents, Plant Variety Protection (PVP) certificates, and utility patents.⁶² Plant patents were the first formal form of IP available for living organisms,⁶³ and are available only for asexually reproducing⁶⁴ plants, with the exclusion of tubers.⁶⁵ Plant patents “have very limited coverage and less stringent requirements than § 101 utility patents.”⁶⁶

In 1970, the United States passed the Plant Variety Protection Act (PVPA).⁶⁷ The U.S. Department of Agriculture (USDA) issues PVP certificates under the PVPA,⁶⁸ providing IP protection for new, distinct, uniform, and stable plant varieties.⁶⁹ PVP applications are simpler than patent applications. But in some respects, they provide weaker protection than

62. USDA, *Plant Variety Protection*, AGRIC. MARKETING SERV., <https://www.ams.usda.gov/services/plant-variety-protection> (last visited Dec. 11, 2019). Plants may also be trade secrets, and variety names may be trademarked. See, e.g., Michael R. Ward & Elizabeth Freeman Rosenzweig, *Planting the Seeds of Change*, Dec. 2018/Jan. 2019 INTELL. PROP. MAG. 43, 43–44, <https://www.intellectualpropertymagazine.com/patent/planting-the-seeds-of-change-132851.htm>. Those forms of IP are beyond the scope of this Note.

63. Petra Moser & Paul W. Rhode, *Did Plant Patents Create the American Rose?*, THE RATE AND DIRECTION OF INVENTIVE ACTIVITY REVISITED, 413, 413 (Josh Lerner & Scott Stern, eds., 2012). Plant patents came into being after the passage of the Plant Patent Act (PPA) of 1930, codified as 35 U.S.C. §§ 161–164 (2018).

64. As Justice Thomas succinctly described in *J.E.M. Ag Supply, Inc. v. Pioneer Hi-Bred Int'l, Inc.*, “[a]sexual reproduction occurs by grafting, budding, or the like, and produces an offspring with a genetic combination identical to that of the single parent—essentially a clone.” 534 U.S. 124, 132 (2001).

65. 35 U.S.C. § 161 (2018). Plant patents are administered by the U.S. Patent and Trademark Office (USPTO). USPTO, *General Information About 35 U.S.C. 161 Plant Patents*, TYPES OF PAT. APPLICATIONS, <https://www.uspto.gov/patents-getting-started/patent-basics/types-patent-applications/general-information-about-35-usc-161#> (last visited Dec. 11, 2019). Plant patents provide “the right to exclude others from asexually reproducing the plant, and from using, offering for sale, or selling the plant so reproduced, or any of its parts, throughout the United States, or from importing the plant so reproduced, or any parts thereof, into the United States.” 35 U.S.C. § 163 (2018). Examples of plants for which plant patents are granted include roses and fruit trees. See Moser & Rhode, *supra* note 63, at 413.

66. *J.E.M. Ag Supply*, 534 U.S. at 133. Plant patents may contain only one claim, see 35 U.S.C. § 162 (2018), and have a “relaxed” written description requirement compared to utility patents. *J.E.M. Ag Supply*, 534 U.S. at 133.

67. Codified as 7 U.S.C. §§ 2321–2583 (2018).

68. See 7 U.S.C. §§ 2321–2372 (2018).

69. The PVPA covers sexually- and asexually-produced plants. 7 U.S.C. § 2402(a) (2018).

patents, because the PVPa contains crop and research exemptions.⁷⁰ These exemptions permit farmers to save and replant seed⁷¹ and permit breeders and researchers to use and reproduce protected varieties for “breeding or other bona fide research”⁷² without infringing the PVP.

The U.S. Patent and Trademark Office (USPTO) began granting *utility* patents on plants in 1985, after the USPTO Board of Patent Appeals and Interferences held in *Ex parte Hibberd*⁷³ “that plants were . . . within the subject matter of [35 U.S.C.] § 101.”⁷⁴ The Supreme Court endorsed this practice in *J.E.M. Ag Supply, Inc. v. Pioneer Hi-Bred Int’l, Inc.*, holding that the existence of plant patents and PVPs does not preclude proper granting of utility patents on plants.⁷⁵ Currently, all three types exist in harmony.⁷⁶

C. A BRIEF PRIMER OF SOME UNIQUE FEATURES OF THE AGRICULTURAL SEED MARKET

The agricultural seed market as we know it today owes its prominence to the development of hybrid seed technology for corn in the early 1900s.⁷⁷ Before then, most farmers planted seed saved from previous harvests.⁷⁸ Hybrid corn offered a dramatic increase in yield over so-called open-pollinated varieties⁷⁹—but the genetics of the breeding programs required to produce hybrid seeds are so complex that the average farmer cannot produce them themselves, and is thus required to purchase new seed for every crop.⁸⁰ Hybrid seeds thus represent a de-coupling of “seed as ‘seed’ from seed as ‘grain,’”⁸¹

70. Jim Chen, *The Parable of the Seeds: Interpreting the Plant Variety Protection Act in Furtherance of Innovation Policy*, 81 NOTRE DAME L. REV. 105, 125–26 (2006).

71. See 7 U.S.C. § 2543 (2018).

72. 7 U.S.C. § 2544 (2018).

73. 227 U.S.P.Q. (BNA) 443, 444 (1985).

74. *J.E.M. Ag Supply, Inc. v. Pioneer Hi-Bred Int’l, Inc.*, 534 U.S. 124, 131 (2001) (summarizing *Ex parte Hibberd*, 227 U.S.P.Q. at 444).

75. *J.E.M. Ag Supply*, 534 U.S. at 145.

76. See USDA, *supra* note 62.

77. Fernandez-Cornejo, *supra* note 56, at 25.

78. *Id.*

79. See Jack Ralph Kloppenburg, *First the Seed: The Political Economy of Plant Biotechnology* 91 (2004).

80. See *id.* at 99. Hybrid plants do produce seed, but the quirks of hybrid genetics are such that saved seed would produce irregular and lower-yielding plants; “although hybrid seed is not biologically sterile . . . it is in effect ‘economically sterile.’” *Id.* at 97 (citations omitted).

81. *Id.* at 93. That is, the commodity “grains” are no longer equivalent to the seeds required to produce the next season’s crop.

officially making seeds a commodity, whose production and use are necessarily separated.⁸² Notably, this took place before the advent of plant patents.⁸³

Hybrids quickly found wild success,⁸⁴ leading to a handful of large firms dominating seed production for major field crops.⁸⁵ Between the spread of hybrid technology and the advent of modern biotechnology and recombinant DNA technology, these seed companies continue to improve the productivity of their seeds.⁸⁶ Correspondingly, farmers are purchasing (rather than saving) higher and higher proportions of the seed they use.⁸⁷

Many factors contribute to the market price companies set for their seed, but a large portion is driven by high research and development costs.⁸⁸ Even though the growth of the seed market was initially propelled by the inability to replant hybrid seeds,⁸⁹ many proprietary seeds today are not hybrid and are fully biologically capable of being replanted.⁹⁰ In order to recoup their costs and maintain a profit,⁹¹ therefore, major seed companies typically require

82. *See id.*

83. *See id.* at 97–105 (tracking the development of hybrid corn from 1905 to 1924).

84. By 1965, more than ninety-five percent by acre of corn in the United States was grown from hybrid seed. Fernandez-Cornejo, *supra* note 56, at 25.

85. *Id.* at 28. “Major field crops” are: corn, soybeans, cotton, and wheat. *Id.*

86. *See* Julie Babinard, *A Short History of Agricultural Biotechnology*, in GENETICALLY MODIFIED ORGANISMS IN AGRICULTURE 272–73 (Elsevier, 2001).

87. For example, acres of soybean and cotton from purchased seed in the United States increased from fifty-five to eighty-one percent and fifty to seventy-eight percent, respectively, between 1982 and 1997. Fernandez-Cornejo, *supra* note 56, at 7. By 1997, seed purchases accounted for four percent of total farm expenditures. *Id.*

88. *Id.* at 29.

89. *See* Kloppenburg, *supra* note 79, at 94.

90. Hybrid seeds are still available today and are the predominant mode of production for many crops, including nearly all corn, sugar beet, and sorghum. *See id.* at 125. But not all crops have been amenable to hybrid technology; wheat, for example, has proven particularly challenging, and most commercial wheat today is not hybrid. *See* Pushpendra Kumar Gupta, Harindra Singh Balyan, Vijay Gahlaut, Gautam Saripalli, Bijendra Pal, Bhoja Raj Basnet, and Arun Kumar Joshi, *Hybrid Wheat: Past, Present and Future*, 132 THEOR. APPL. GENET. 2463, 2463 (2019); Steve Mercer, *A Welcome Look at Hybrid Wheat Research*, U.S. WHEAT ASSOCIATES (May 17, 2018), <https://www.uswheat.org/wheatletter/a-welcome-look-at-hybrid-wheat-research/>. As far as replanting of non-hybrid seed is concerned, note that so-called “Terminator” seeds, containing Genetic Use Restriction Technology (GURT) that causes second-generation seeds to be sterile, have never been commercialized. *See* Genetic Literacy Project, *What’s the Controversy Over ‘Terminator’ Seeds?*, GMO FAQs, <https://gmo.geneticliteracyproject.org/FAQ/whats-controversy-gmos-terminator-seeds/> (last visited Dec. 11, 2019).

91. “Seed companies involved in research and development reinvest approximately 10[%] of sales into new innovations”; because new varieties take approximately eight years and between about \$1 million to \$7 million to develop “depending on the crop, and the cost of commercializing a new trait [is] estimated to be \$136 million over 13 years, companies have

farmers to consent to a specialized licensing agreement that does not permit saving and replanting.⁹² These licenses are often called “bag tag” or “seed wrap” licenses because they are frequently displayed on seed bags.⁹³

IV. *LEXMARK AND BOWMAN*

The Supreme Court has issued four decisions this century on patent exhaustion or the related concept of copyright first sale: first in 2008 for patents in *Quanta Computer, Inc. v. LG Electronics, Inc.*,⁹⁴ then in 2013 for copyright in *Kirtsaeng v. John Wiley & Sons, Inc.*,⁹⁵ then just shy of two months later for patents again in *Bowman*,⁹⁶ and most recently in 2017 for patents in *Lexmark*.⁹⁷

a financial responsibility to their shareholders and customers.” James Weatherly, *Can I Save That Seed?*, SEEDWORLD (Nov. 28, 2018), <https://seedworld.com/can-i-save-that-seed/>.

92. See Mark D. Janis, *Intellectual Property Issues in Plant Breeding and Plant Biotechnology*, ARTICLES BY MAURER FACULTY 2560 (2002), available at <http://www.repository.law.indiana.edu/facpub/2560>. For example, Monsanto/Bayer refers to its limited use seed license it provides as a “Technology Stewardship Agreement.” See Bayer, *2019 U.S. Tech. Use Guide and Insect Resistance Mgmt. Overview*, TECH. USE GUIDE AND INSECT RESISTANT MGMT. OVERVIEW: DOWNLOADS 1, 62, <http://tug.monsanto.com/> (last visited Dec. 11, 2019). These detail rules for how the seeds are to be used, including that the grower must: only acquire Monsanto seed from authorized companies; pay royalties and technology fees; use the seed in only “a single planting” and not save it for-replanting; not transfer the seed for planting or export; not plant seed for production without a license; and not plant seed for breeding or research purposes. *Id.* The Agreement at issue in *Bowman* was similar, see *Bowman v. Monsanto Co.*, 569 U.S. 278, 281 (2013), and listed the relevant patents. See *Monsanto Co. v. Bowman*, 657 F.3d 1341, 1349 (Fed. Cir. 2011). Monsanto had also told *Bowman* that, per the Agreement, “[p]lanting of seed that is covered by a patent would be making the patented invention and using the patented invention,” and thereby infringe. *Id.* (alteration in original) (citation omitted).

93. See Janis, *supra* note 92. Interestingly, there has been a recent push for “open source seed” license. See Johannes Kotschi & Bernd Horneburg, *The Open Source Seed Licence: A novel approach to safeguarding access to plant germplasm*, 16 PLOS BIOL. e3000023 (Oct. 23, 2018), <https://doi.org/10.1371/journal.pbio.3000023>. Currently, a small handful of seed products are available under such licenses in Europe, but it remains to be seen how widespread this practice will become. See *id.*; see also Jack Kloppenburg, *Re-Purposing the Master’s Tools: The Open Source Seed Initiative and the Struggle for Seed Sovereignty*, Food Sovereignty: A Critical Dialogue Conference Paper #56 (2013), <https://dces.wisc.edu/wp-content/uploads/sites/128/2013/08/2013-Repurposing.pdf>.

94. *Quanta Computer, Inc. v. LG Elecs., Inc.*, 553 U.S. 617 (2008).

95. See *Kirtsaeng v. John Wiley & Sons, Inc.*, 568 U.S. 519 (2013).

96. See *Bowman v. Monsanto Co.*, 569 U.S. 278, 284 (2013).

97. See *Impression Prods. v. Lexmark Int’l, Inc.*, 137 S. Ct. 1523, 1526 (2017).

A. A BRIEF SUMMARY OF EXHAUSTION JURISPRUDENCE PRE-*BOWMAN*

The Supreme Court did not touch exhaustion for sixty-six years, between *Univis*⁹⁸ in 1942 and *Quanta*⁹⁹ in 2008. In *Univis*, the Court held that exhaustion prevents a patentee from using an infringement suit to “control the resale price of patented articles which he has sold,”¹⁰⁰ because exhaustion “applies when the item sufficiently embodies the patent . . . such that its only and intended use is to be finished under the terms of the patent.”¹⁰¹ In *Quanta*, the Court held that exhaustion can apply to method claims,¹⁰² and that sales authorized in a licensing agreement invoke exhaustion.¹⁰³

In the meantime, the Federal Circuit built its own scheme for exhaustion and single-use restrictions (later to be rejected by the Supreme Court¹⁰⁴) in *Mallinckrodt*.¹⁰⁵ In that case, the accused infringement consisted of disobeying a single-use restriction that accompanied the sale of a patented item; the court held that “[u]se in violation of a valid restriction may be remedied under the patent law,”¹⁰⁶ meaning that patentees “could circumvent exhaustion”¹⁰⁷

In 2013, the Supreme Court again took up the concept of exhaustion in *Kirtsaeng*,¹⁰⁸ which, even though the case is about copyright first sale and not patent exhaustion, delves into their shared common law origins.¹⁰⁹ In *Kirtsaeng*, the “Court held that the first sale doctrine applies to copies of works made and sold abroad.”¹¹⁰

B. *BOWMAN V. MONSANTO*

Eight weeks after issuing *Kirtsaeng*,¹¹¹ the Court issued its decision in *Bowman*, again wading into the topic of exhaustion, albeit narrowly.¹¹² Justice

98. *United States v. Univis Lens Co.*, 316 U.S. 241 (1942).

99. *Quanta*, 553 U.S. 617 (2008).

100. *Univis*, 316 U.S. at 250.

101. *Quanta*, 553 U.S. at 628.

102. *Id.* at 629–30.

103. *Id.* at 637.

104. *See generally* Kobak, Jr., *supra* note 18.

105. *Mallinckrodt, Inc. v. Medipart, Inc.*, 976 F.2d 700 (Fed. Cir. 1992).

106. *Id.* at 701.

107. Kobak, Jr., *supra* note 18, at 609.

108. *Kirtsaeng v. John Wiley & Sons, Inc.*, 568 U.S. 519 (2013).

109. *See* *Impression Prods. v. Lexmark Int’l, Inc.*, 137 S. Ct. 1523, 1526 (2017).

110. *Id.* at 1527.

111. *Kirtsaeng*, 568 U.S. 519 (decided Mar. 19, 2013).

112. *Bowman v. Monsanto Co.*, 569 U.S. 278 (decided May 13, 2013).

Kagan delivered the unanimous *Bowman* opinion,¹¹³ which relies on *Univis*¹¹⁴ and *Quanta*,¹¹⁵ but cites neither *Mallinckrodt* nor *Kirtsaeng*.

Previously, when *Bowman* had reached the Federal Circuit, it was the next case in a narrow line of cases about disputes between Monsanto and farmers over Monsanto's patented "Roundup Ready®"¹¹⁶ herbicide resistant seeds,¹¹⁷ and, "in each case, the Federal Circuit found for Monsanto and against the farmer,"¹¹⁸ but the Supreme Court had not yet weighed in.

The facts of *Bowman* are similar to those of *McFarling* and *Scruggs*;¹¹⁹ here, Indiana farmer Vernon Bowman replanted "Roundup Ready®" soybeans for eight seasons¹²⁰ before Monsanto sued him for patent infringement.¹²¹ The license permitted Bowman "to plant the purchased seeds in one (and only one) season."¹²² After that season, the license permitted him to "consume the

113. *Id.* at 279.

114. *See id.* at 283–5.

115. *See id.* at 283.

116. "Roundup Ready" is Monsanto's (now Bayer's) trademark for its genetically engineered seeds that are resistant to the herbicide glyphosate. *See* Bayer, *History of Excellence*, ROUNDUP READY PLUS, <http://www.roundupreadyplus.com/platform/history> (last visited Dec. 11, 2019).

117. *See* *Monsanto Co. v. Scruggs*, 459 F.3d 1328 (Fed. Cir. 2006); *Monsanto Co. v. McFarling*, 363 F.3d 1336 (Fed. Cir. 2004) [hereinafter *McFarling II*]; *Monsanto Co. v. McFarling*, 302 F.3d 1291 (Fed. Cir. 2002) [hereinafter *McFarling I*].

118. Jeremy N. Sheff, *Self-Replicating Technologies*, 16 STAN. TECH. L. REV. 229, 231 (2013). In the *McFarling* cases, Monsanto sued a farmer for patent infringement for replanting saved seed in violation of the licensing agreement. *McFarling II*, 363 F.3d at 1339. The court noted that only the first-generation of seeds was licensed, and that creating the saved seed constituted making a patented good, which exhaustion does not cover. *Id.* at 1343. Furthermore, exhaustion could not apply to the second-generation seeds because the patentee had never sold them. *McFarling I*, 302 F.3d at 1299. In *Scruggs*, farmers purchased Monsanto seeds from a Monsanto-licensed seed seller, planted them without signing a licensing agreement, and then saved and replanted subsequent generations. 459 F.3d at 1333. Monsanto sued them for patent infringement, and the court found that the farmers lacked both actual and implied licenses. *Id.* at 1336. The court also held that exhaustion did not apply, because "the new seeds . . . had never been sold." *Id.* The court then took a stance on self-reproducing technologies more broadly, stating that "[t]he fact that a patented technology can replicate itself does not give a purchaser the right to use replicated copies of the technology," and that, if exhaustion applied "to subsequent generations of self-replicating technology[, it] would eviscerate the rights of the patent holder." *Id.*

119. *See supra* note 118 (summarizing the *McFarling* and *Scruggs* cases).

120. *Bowman v. Monsanto Co.*, 569 U.S. 278, 282 (2013).

121. The patents at issue cover particular, genetically-engineered genes and DNA sequences, plant cells comprising those genes or sequences, plants comprising those cells, seeds produced by those plants, and a method for using glyphosate on those seeds or plants. *Monsanto Co. v. Bowman*, 657 F.3d 1341, 1343–44 (Fed. Cir. 2011).

122. *Bowman*, 569 U.S. at 281.

resulting crop or sell it as a commodity,” but not to save seeds for replanting.¹²³ Bowman, who planted twice a year, followed these restrictions every year for his first harvest, but not for his second.¹²⁴ For his second harvests, he purchased commodity soybeans¹²⁵ from a grain elevator, planted them, applied Roundup® weed killer to select for “Roundup Ready®” plants, and then saved and replanted that seed.¹²⁶

Bowman argued that patent exhaustion absolved him of infringement.¹²⁷ The district court, the Federal Circuit, and the Supreme Court, however, did not agree.¹²⁸ The Federal Circuit held that exhaustion did not permit Bowman to “‘replicate’ Monsanto’s patented technology by planting it in the ground to create *newly infringing* genetic material, seeds, and plants.”¹²⁹

The Supreme Court affirmed, reasoning that growing more seed from purchased seed is properly categorized as *making a new copy of a patented item*, rather than simply *using a purchased item*.¹³⁰ The Court emphasized that exhaustion “restricts a patentee’s rights *only as to the particular article sold*; it leaves untouched the patentee’s ability to prevent a buyer from *making new copies . . .*”¹³¹ Thus, exhaustion would allow Bowman to “resell the patented soybeans he purchased[,] . . . consume the beans himself or feed them to his animals”¹³² without infringing, but not replant saved seed.¹³³

The Court noted that if it were to have held for Bowman, Monsanto would be deprived of a fair reward for its invention.¹³⁴ The Court also explained that a holding for Bowman would have collapsed the heightened protections available through patents compared to PVP certificates, which would contradict the structure the Court specifically endorsed in *J.E.M. Ag Supply*.¹³⁵ Finally, the Court rejected Bowman’s so-called “blame-the-bean” defense,¹³⁶

123. Nor did it permit him to “supply them to anyone else for that purpose.” *Id.* at 281.

124. *Id.* at 281–82.

125. Which are intended only for consumption, not planting.

126. *Bowman*, 569 U.S. at 281–82.

127. This was because “the soybeans . . . were the subject of a prior authorized sale (from local farmers to the grain elevator).” *Id.* at 283.

128. *Id.*

129. *Monsanto Co. v. Bowman*, 657 F.3d 1341, 1348 (Fed. Cir. 2011) (emphasis added).

130. *See Bowman*, 569 U.S. at 287.

131. *Id.* at 284 (emphases added) (internal quotation marks and citations omitted).

132. *Id.* at 284.

133. *Id.* at 284–85.

134. *Id.* at 285–86.

135. *Id.* at 286.

136. Bowman argued that because “soybeans naturally self-replicate[,] . . . it was the planted soybean, not Bowman himself, that made replicas of Monsanto’s patented invention.” *Id.* at 288 (internal quotations and citation removed).

emphasizing that “Bowman was not a passive observer of his soybeans’ multiplication,”¹³⁷ but rather undertook a long list of active steps to “harvest crops from Roundup Ready seeds without paying the usual premium.”¹³⁸

Critically, however, the Court was clear that the *Bowman* holding was narrow, “addressing the situation before us, rather than every one involving a self-replicating product.”¹³⁹ Exhaustion, the Court reiterated, “provides no haven for [] conduct”¹⁴⁰ intended “solely to make and market replicas”¹⁴¹ But it explicitly left open two questions about self-reproducing technologies, declining to answer “whether or how the doctrine of patent exhaustion would apply” in other cases in which (1) “the article’s self-replication might occur outside the purchaser’s control” or (2) the article’s self-replication “might be a necessary but incidental step in using the item for another purpose.”¹⁴²

Bowman thus clarified that, at least for the facts before it, exhaustion only applies to the first iteration of use of any particular self-replicating product sold. When Bowman created and then used subsequent generations of seed, he was operating outside the realm of exhaustion: his actions to “make” these new generations reached beyond the tangible object he purchased, and necessarily implicated the knowledge from Monsanto’s patent about the Roundup Ready® trait.¹⁴³ In a way, exhaustion itself exhausts after one generation of replication, such that acts that freshly implicate the patented knowledge constitute new, unexhausted infringement.¹⁴⁴

Bowman’s exhaustion is more like “quasi”-exhaustion, prohibiting uses that implicate patented knowledge even when exhaustion otherwise applies.

137. *Id.*

138. *Id.* at 288–89.

139. *Id.* at 289.

140. *Id.*

141. *Id.*

142. *Bowman v. Monsanto Co.*, 569 U.S. 278, 289 (2013).

143. *See id.* at 285 (listing Bowman’s affirmative acts to save and replant seed).

144. This explains why the Court noted that the seeds were exhausted *as to particular uses*. *Id.* at 284 (noting that exhaustion did not prevent Bowman from reselling, eating, or feeding the grain elevator seeds to animals). Such use-acts do not require the patented knowledge. But exhaustion did cover “uses” involving “making,” *id.* at 284–85 (“the exhaustion doctrine does not enable Bowman to make additional patented soybeans without Monsanto’s permission”), because the “making” that Bowman conducted necessarily implicated the patented knowledge. The Court did not state that in so many words, but its description of Bowman’s “making” actions, *id.* at 285 (“He took the soybeans he purchased home; planted them in his fields at the time he thought best; applied glyphosate to kill weeds (as well as any soy plants lacking the Roundup Ready trait); and finally harvested more (many more) beans than he started with.”) (emphasis added), clearly points out how the application of glyphosate was crucial to the process. Seeds would not be planted and grown this way without knowledge of Monsanto’s proprietary Roundup Ready® technology.

Conversely, there is some degree of permissible “making,” as far as patented DNA is concerned: genetic material will inherently reproduce even in authorized single plantings as the seeds germinate and resulting plants grow.¹⁴⁵ But *Bowman*’s exhaustion compromise makes sense. Other options would be the extremes of either making exhaustion completely unavailable for seeds or absolving all generations after the first sold seed from possibly infringing. But those alternatives constitute an unfairly arbitrary carve-out in the case of the former, or are economically perverse in the case of the latter. Instead, *Bowman* splits the difference. And *Bowman* does not result in a bizarre situation in modern industrial agriculture. The patent-derived need to purchase seed anew each season mirrors how the hybrid seed market, with its biologically-driven prohibition against replanting seeds, has functioned for a century,¹⁴⁶ showing that “use” and “making” of seeds can be meaningfully decoupled without restraining farmers from alienating themselves from their chattels.¹⁴⁷

C. *IMPRESSION PRODS. V. LEXMARK INT’L, INC.*

The Supreme Court’s next exhaustion case was *Lexmark*. Chief Justice Roberts delivered the majority opinion,¹⁴⁸ which references *Univis*,¹⁴⁹ *Quanta*,¹⁵⁰ and *Kirtsaeng*,¹⁵¹ but, curiously, not *Bowman*.¹⁵² *Lexmark* was not about self-replication, but about refilling and reselling printer cartridges.¹⁵³ *Lexmark*

145. Perhaps uncoincidentally, limited permission for such “micro-makings” (if you will) is implicitly granted by the seed’s licensing agreement, which typically permits crop production and sale (but not replanting saved seed, breeding, or research). *See supra* at note 92 (explaining the content of a licensing agreement).

146. *See supra* text accompanying notes 77–89 (summarizing the hybrid seed market).

147. *See Bowman*, 569 U.S. at 288 (noting that “nonreplicating use of [] commodity beans”—that is, for consumption—is “not just available, but standard fare. And in the more ordinary case, when a farmer purchases Roundup Ready seed *qua* seed—that is, seed intended to grow a crop—he will be able to plant it.”) (alteration in original).

148. The decision was issued almost exactly four years after *Bowman*, on May 30, 2017. Justices Kennedy, Thomas, Breyer, Alito, Sotomayor, and Kagan joined; Justice Ginsburg concurred in part and dissented in part, and Justice Gorsuch did not take part. *Impression Prods. v. Lexmark Int’l, Inc.*, 137 S. Ct. 1523, 28 (2017).

149. *See, e.g., id.* at 1531.

150. *See, e.g., id.*

151. *See, e.g., id.* at 1532.

152. However, CropLife International, a trade association whose membership includes Monsanto and many other large seed companies, did write an amicus brief in support of *Lexmark*. *See* Brief of Amici Curiae Biotechnology Innovation Organization and CropLife International in Support of Respondent at 1, *Lexmark*, 137 S. Ct. 1523 (No. 15-1189), *available at* https://www.scotusblog.com/wp-content/uploads/2017/03/15-1189_amicus_resp_biotechnology_innovation_organization.pdf.

153. *Lexmark*, 137 S. Ct. at 1529.

“designs, manufactures, and sells toner cartridges,”¹⁵⁴ on which it owns several patents.¹⁵⁵ To encourage customers to return (rather than refill and reuse) the cartridges, Lexmark created a “Return Program” that offered cartridges at a reduced price—but with contractual single-use/no sale restrictions.¹⁵⁶

Despite these contracts, a number of “remanufacturer” companies acquired empty cartridges, refilled them, and resold them.¹⁵⁷ The remanufacturers also imported cartridges they had acquired abroad and resold them in the United States.¹⁵⁸ When Lexmark got wind of this, they sued the remanufacturers for patent infringement.¹⁵⁹

Lexmark argued that the remanufacturers infringed its patents by (1) refurbishing and reselling the “Return Program cartridges that Lexmark sold within the United States”,¹⁶⁰ and (2) importing, without authority, cartridges that Lexmark sold overseas.¹⁶¹ The remanufacturers countered that their actions were shielded from infringement because “Lexmark’s sales, both in the United States and abroad, exhausted its patent rights in the cartridges”¹⁶²

The Federal Circuit found no exhaustion¹⁶³—but the Supreme Court reversed.¹⁶⁴ The Court said the Federal Circuit “got off on the wrong foot”¹⁶⁵ by framing exhaustion as “a presumption about the authority that comes along with a sale [rather than] a limit on the scope of the *patentee’s rights*.”¹⁶⁶ The

154. *Id.*

155. *Id.*

156. Specifically, the contract required customers to agree to use the cartridge “only once and to refrain from transferring the empty cartridge to anyone but Lexmark.” *Id.* at 1530.

157. *Id.* at 1529.

158. *Id.* at 1530.

159. *Id.*

160. Lexmark justified this by the Return Program’s express prohibition on “reuse and resale.” *Id.*

161. *Id.*

162. *Id.*

163. *Id.* For Lexmark’s first point, the Federal Circuit relied on its own precedent from *Mallinckrodt* that violations of single-use/no-resale restrictions that are themselves “lawful and clearly communicated to the purchaser” are enforceable under patent law as “infringing conduct” *Lexmark Int’l, Inc. v. Impression Prods.*, 816 F.3d 721, 726 (Fed. Cir. 2016). The court reasoned that exhaustion did *not* apply because “a patentee’s decision to sell an item ‘presumptively grant[s] ‘authority’ to the purchaser to use it and resell it.’ But . . . the patentee does not have to hand over the full ‘bundle of rights’ every time.” *Impression Prods. v. Lexmark Int’l, Inc.*, 137 S. Ct. 1523, 1533–54 (2017) (quoting *Lexmark*, 816 F.3d at 741–42) (emphasis in original). Similarly, the Court held that “foreign sale[s] do[] not trigger patent exhaustion unless the patentee expressly or implicitly transfers or licenses its rights.” *Id.* at 1535 (citations omitted) (internal quotation marks omitted).

164. It also remanded. *See id.* at 1538.

165. *Id.* at 1533.

166. *Id.* at 1534 (citations omitted) (internal quotation marks omitted).

single-use restrictions “may have been clear and enforceable under contract law, but”¹⁶⁷ were not under patent law; the Court held that “Lexmark exhausted its patent rights in these cartridges *the moment it sold them.*”¹⁶⁸ Regarding the overseas sales, the Court held that “[a]n authorized sale outside the United States, just as one within the United States, exhausts all [patent] rights.”¹⁶⁹

In its reasoning, the Court did cite a string of its own precedent on the application of exhaustion,¹⁷⁰ but mostly moved past the details of particular fact patterns and took a deep dive into dissecting the theory of exhaustion, laying bare its inner workings. The Court leaned heavily on “the common law principle against restraints on alienation,”¹⁷¹ going back to Lord Coke in the seventeenth century.¹⁷² Translating this to the modern day, the Court used a theoretical “illustration” of a used car shop to stress that the “smooth flow of commerce would sputter if companies that make the thousands of parts that go into a vehicle could keep their patent rights after the first sale.”¹⁷³

The Court explained that exhaustion takes place upon *sale* because a sale fulfills patent law’s purpose of ensuring that “the patentee has received his reward for the use of his invention”¹⁷⁴ Once that “reward” transfers, there remains “no basis for restraining the use and enjoyment of the thing sold.”¹⁷⁵ Contrasting sales and licenses, the Court explained that exhaustion is concerned with the passing of title on tangible goods, whereas licensing “is about changing the contours of the patentee’s monopoly”¹⁷⁶ As far as sales by licensees are concerned, “if a patentee has not given authority for a licensee to make a sale, that sale cannot exhaust the patentee’s rights.”¹⁷⁷

167. *Id.* at 1531.

168. *Id.* (emphasis added). Justice Ginsburg concurred on this point. *Lexmark*, 137 S. Ct. at 1538 (Ginsburg, J., concurring in part and dissenting in part).

169. *Id.* at 1535. The Court noted that “[t]his question about international exhaustion of intellectual property rights has also arisen in the context of copyright law,” referring to *Kirtsaeng*. *Id.* at 1535–36. The Court revisited its reasoning from that case to conclude that “[a]pplying patent exhaustion to foreign sales is just as straightforward.” *Id.* at 1536. Justice Ginsburg dissented on this point and would have found no exhaustion of U.S. patent rights upon foreign sales. *Id.* at 1538 (Ginsburg, J., concurring in part and dissenting in part).

170. *See Id.* at 1531–33, 1537–38.

171. *Id.* at 1531, 1538.

172. *Id.* at 1532.

173. *Id.*

174. *Id.* (citations omitted) (internal quotations marks omitted).

175. *Id.* (citations omitted) (internal quotations marks omitted).

176. *Id.* at 1534 (citations omitted). This makes perfect sense when you consider that patents are drawn to intangible knowledge. *See* Syed, *supra* note 30, at 9.

177. *Lexmark*, 137 S. Ct. at 1535.

Where does that leave Lexmark and others who want to impose post-sale restrictions? According to the Court, there is

only one answer: Lexmark cannot bring a patent infringement suit . . . to enforce the single-use/no-resale provision Once sold, the . . . cartridges passed outside of the patent monopoly, and whatever rights Lexmark retained are a matter of the contracts with its purchasers, not the patent law.¹⁷⁸

In sum, “[e]xhaustion does not depend on whether the patentee receives a premium for selling in the United States, or the type of rights that buyers expect to receive [R]estrictions and location are irrelevant; what matters is the patentee’s decision to make a sale.”¹⁷⁹ Exhaustion is not about “dealings between the parties, which can be addressed through contract law.”¹⁸⁰ Rather, it is about the principle of what a sale is: that the patentee has decided “to give up title to an item in exchange for payment.”¹⁸¹ In what is perhaps simultaneously the most colorful, succinct, and useful summary of the entire decision, the Court concluded that “[a]llowing patent rights to stick remora-like to that item as it flows through the market would violate the principle against restraints on alienation.”¹⁸²

Interestingly, by dismissing the importance of a “premium” for the patentee as a basis for exhaustion, the Court implicitly clarified its own opinion from four years earlier in *Bowman*. *Bowman* did not clearly articulate the difference between (1) whether the sale price of an item provides fair compensation to the patentee in a given market and (2) why exhaustion does not apply to newly-made items.¹⁸³ *Lexmark* clarifies that the point is “that the patentee receives one reward—of whatever amount the patentee deems to be ‘satisfactory compensation,’—for every item that passes outside the scope of the patent monopoly,”¹⁸⁴ not that the patentee should be “guarantee[d] a

178. *Id.* at 1533.

179. *Id.* at 1538.

180. *Id.*

181. *Lexmark*, 137 S. Ct. at 1538. The U.C.C. defines “sale” as “the passing of title from the seller to the buyer for a price.” U.C.C. § 2-106(1) (AM. LAW INST. & UNIF. LAW COMM’N 2019).

182. *Id.* A remora is a type of fish that uses a suction cup on its head to attach to sharks and other larger sea creatures. Their attachment mechanism is so strong that they “stay attached as water rushes past them [and] can even hold tight as their hosts try to scrape them off on rocks.” Carl Zimmer, *What Good Is Half a Sucker?*, NAT. GEO. (July 17, 2013), <https://www.nationalgeographic.com/science/phenomena/2013/07/17/what-good-is-half-a-sucker/#close>.

183. *See* *Bowman v. Monsanto Co.*, 569 U.S. 278, 285–86 (2013).

184. *Lexmark*, 137 S. Ct. at 1537 (quoting *Keeler v. Standard Folding Bed Co.*, 157 U.S. 659, 661 (1895)).

particular price”¹⁸⁵ Items exit the scope of the patent monopoly when they are sold (thus removing restraints on the purchaser’s ability to later alienate the item), not by whether the patentee ultimately makes a profit in a given market.¹⁸⁶

Preliminary data imply that the success rate of exhaustion defenses may be dropping post-*Lexmark*.¹⁸⁷ But given the pace of litigation, there are still too few cases to measure *Lexmark*’s impact with statistical significance.¹⁸⁸ It will be interesting for future scholars to investigate whether exhaustion defenses are more, less, or equally popular and effective in a post-*Lexmark* world.

Scholars and practitioners have criticized *Lexmark* for imprecision and lack of guidance.¹⁸⁹ Although *Lexmark* provided a welcome dissection of precisely what patent exhaustion means, it failed to provide a clear path forward for patentees to determine if their “licenses” might be construed as “sales,” or whether there remains a viable way to limit downstream use and resale.¹⁹⁰

Shubha Ghosh and Irene Calboli take issue with what they see as a “confounding” and overly-broad “misstatement” in the Court’s announcement that “[a] patentee’s decision to sell a product exhausts *all* of its

185. *Id.*

186. *See id.* at 1531.

187. *Lexmark* was decided on May 30, 2017. *Lexmark*, 137 S. Ct. at 1523. According to data from darts-ip collected on October 12, 2019, exhaustion defenses were accepted in forty-five percent of patent infringement cases in the United States between May 30, 2010, and May 29, 2017, versus in eighteen percent of cases between May 31, 2017, and October 12, 2019 (when the author ran the search). However, the sample sizes—especially for the post-*Lexmark* set—are quite small, representing only seventy-six and eleven total cases in which exhaustion was argued before and after *Lexmark*, respectively. Unsurprisingly, the difference between these proportions is insignificant at $p < 0.05$ when evaluated using Fisher’s Exact Test (test statistic value: 0.1138). Thus, there is a low likelihood that the two populations (of cases before versus after *Lexmark*) are significantly different. The author performed the Fisher’s Exact Test on the above data from darts-ip using a free online calculator provided by Social Science Statistics. *See* Jeremy Stangroom, *Easy Fisher Exact Test Calculator*, SOC. SCI. STAT., <https://www.socscistatistics.com/tests/fisher/default2.aspx> (last visited Dec. 11, 2019).

188. This is according to a search run on darts-ip on October 12, 2019. The lack of significance is unsurprising given that, in the roughly two-and-a-half years between the *Lexmark* decision and when the author collected the data, there have only been eleven U.S. cases in which an exhaustion defense has been raised (based on the search on darts-ip).

189. *See infra* notes 190–193. For further critiques of *Lexmark*, see also Hye Jin Kim, *Avoiding Patent Exhaustion at Home and Abroad: Impression Products v. Lexmark International*, 33 BERKELEY TECH. L.J. 945, 963–80 (2018); Kobak, Jr., *supra* note 18, at 610.

190. Brian Kacedon & Kevin D. Rodkey, *The Aftermath of Impression Products v. Lexmark*, LAW360 (Nov. 13, 2017), <https://www.finnegan.com/en/insights/the-aftermath-of-impression-products-v-lexmark.html>.

patent rights.”¹⁹¹ First, Ghosh and Calboli point out that “the act triggering exhaustion is *the actual sale*, not just simply the *decision*.”¹⁹² Second, they stress that exhaustion does not exhaust “all” patent rights: it is limited to the particular copy that was sold, and does not traditionally apply to the right to make.¹⁹³ But their second point reflects a misunderstanding of the Court’s statement. The next three words of the quote, which Ghosh and Calboli omitted in their analysis, are crucial. What the Court actually said is that “[a] patentee’s decision to sell a product exhausts all of its patent rights *in that item . . .*”¹⁹⁴ And the opinion clarifies that the rights transferred by sale are only “the right to use, sell, or import,”¹⁹⁵ those being the rights associated with “ownership.”¹⁹⁶ Thus, the Court did not imply that the right to make is excused by exhaustion. While the use of the word “all” in that sentence is needlessly confusing, in proper context it is not incorrect.

D. WHY *BOWMAN* STILL STANDS, POST-*LEXMARK*

A superficial reading of *Lexmark* and *Bowman* might lead one to think that *Lexmark* overturned *Bowman*: they both involved single-use licenses and downstream actions by purchasers. But the *Bowman* Court found that exhaustion did not apply, whereas the *Lexmark* court did. Furthermore, *Bowman* emphasized the importance of the value of the patent to the patentee¹⁹⁷ in a way that *Lexmark* dismissed.¹⁹⁸

However, the actual decision in *Bowman* does not depend on the content of the agreements, whether one categorizes the accompanying transaction as a license or a sale, or the amount of sales profit to which the patentee may or may not be entitled. The point is that saved seeds and their progeny cannot be exhausted because they were never sold by the patentee; *they were made* by the farmer. Furthermore, *Bowman*’s act of “making” depended on the patented knowledge of the function of the Roundup Ready® trait. He knew he could

191. Shubha Ghosh & Irene Calboli, EXHAUSTING INTELLECTUAL PROPERTY RIGHTS: A COMPARATIVE LAW AND POLICY ANALYSIS 101 (2018) (quoting *Lexmark*, 137 S. Ct. at 1529) (emphasis in original).

192. *Id.* (emphasis added).

193. *Id.*

194. *Lexmark*, 137 S. Ct. at 1529 (emphasis added).

195. *Id.* at 1534.

196. *Id.*

197. See Daryl Lim, *Argibiotech Patents in the Food Supply Chain: A U.S. Perspective*, GLOBAL FOOD VALUE CHAINS AND COMPETITION LAW, 8 (Ioannis Lianos, Alexey Ivanov & Dennis Davis eds., 2018) (stating that “the heart of *Bowman* was the policy concern that patent protection had to protect Monsanto’s innovation-linked investment and expense”) (citing *Bowman v. Monsanto Co.*, 569 U.S. 278, 285–89 (2013)).

198. See *Lexmark*, 137 S. Ct. at 1538.

obtain only Roundup Ready® plants from mixed commodity seed *because only Roundup Ready® plants in the mix would survive* herbicide application.¹⁹⁹

Exhaustion boils down to a rejection of restraints on alienation of chattels,²⁰⁰ but that only really makes sense when applied to chattels already in existence, even if they are self-replicating.²⁰¹ Thus, “‘a second creation’ of the patented item ‘call[s] the monopoly, conferred by the patent grant, into play for a second time.’”²⁰²

Daryl Lim suggests that, because of *Lexmark*, seed “sales” with use restrictions will be “contrary to federal patent exhaustion rules,” meaning that seed patentees who wish “to avoid triggering exhaustion must *avoid an authorized sale* through carefully structured licenses so that they retain control over subsequent uses by *stopping short of selling their goods*.”²⁰³ However, he suggests a tension between *Bowman* and *Lexmark* that does not actually exist. Seed sales are still perfectly compatible with licenses on self-reproducing technologies. Even if a seed sale truly is an authorized “sale,”²⁰⁴ an accompanying “license” is still valid post-*Lexmark* so long as that license does not claim to curtail uses contained to the particular set of seeds that were subject to the sale and/or uses that only incidentally implicate the patented knowledge. Even post-*Lexmark*, exhaustion does not apply to *new copies* of the invention or to uses of any copy that *more than incidentally implicate the patented knowledge*. Indeed, this is precisely what “seed wrap” licenses typically cover.²⁰⁵

199. See *Bowman*, 569 U.S. at 282. Dissection of “making” in this manner—by focusing on the implicated knowledge space, see Syed, *supra* note 30, at 10–11, sidesteps the dictionary-driven approach that Justice Kagan deploys in the *Bowman* opinion. See 569 U.S. at 285 (quoting Webster’s Dictionary for “how to ‘make’ a new product . . . when the original product is a seed”) (internal quotations marks and citation removed). The literalist approach is only of limited use, as it is inherently limited by the space a dictionary editor chooses to devote to a given word. That being said, the *Bowman* opinion’s analysis does not actually hang on the dictionary and, like *Lexmark*, is primarily driven by policy. See, e.g., 569 U.S. at 283–84 (discussing the goals of the exhaustion doctrine as developed by classic Supreme Court case law).

200. See *Lexmark*, 137 S. Ct. at 1532 (quoting *Kirtsaeng v. John Wiley & Sons, Inc.*, 568 U.S. 519, 538 (2013)).

201. For example, there is no guarantee that any given seed will necessarily germinate and produce future generations of seeds. See *Knox*, *supra* note 54. Thus, ownership of a given physical set of seeds does not necessarily mean ownership of their progeny, because there may not ever be progeny.

202. *Bowman*, 569 U.S. at 284 (alteration in original) (quoting *Aro Mfg. Co. v. Convertible Top Replacement Co.*, 365 U.S. 336, 346, (1961)).

203. Lim, *supra* note 197, at 19–20 (emphases added).

204. Including, as he puts it, “dispossession of the seeds” that “indicate[s] that the companies intended the seeds be sold through a physical transfer of with no intention of return, indicating a sale rather than a license.” Lim, *supra* note 197, at 19–20.

205. See *Janis*, *supra* notes 92–93.

Thus, the content of Monsanto's (or any company's) seed licenses merely fills in any gaps between the purchased physical group of starting seeds (which are themselves exhausted) and the patented underlying knowledge implicated in any "second creations" or "micro-makings"²⁰⁶ involved in getting from that starting seed to certain permitted end products (such as a single year's crop). The seed licenses thus excuse from infringement defined and limited amounts of "making" that implicate that knowledge, such as the amount of DNA and cell replication necessary to grow a plant from a legitimately purchased seed under the farmer's stewardship. But other actions related to those freshly-made copies (such as Bowman's pipeline of selectively propagating Roundup Ready® seeds) remain unexhausted and unlicensed. *Bowman*, therefore, still stands post-*Lexmark*.

V. BOWMAN'S OPEN QUESTIONS, POST-*LEXMARK*

The *Bowman* Court recognized that self-reproducing "inventions are becoming ever more prevalent, complex, and diverse,"²⁰⁷ and are not necessarily limited to the seed market. It is not clear precisely what technologies they are alluding to here, but it is likely that they had in mind technologies such as engineered cell lines for expressing biopharmaceuticals,²⁰⁸ vaccines,²⁰⁹ or software that can be copied.²¹⁰ In any case, the Court was clear that the *Bowman* decision "is limited—addressing the situation before us"²¹¹—that is, a farmer knowingly and purposefully planting "patented soybeans solely to make and market replicas of them"²¹² without authorization from the

206. See *supra* note 145 and accompanying text.

207. *Bowman*, 569 U.S. at 289.

208. See Brief for the American Intellectual Property Law Association as Amicus Curiae in Support of Affirmance at 34–35, *Bowman v. Monsanto Co.*, 569 U.S. 278 (2013) (No. 11-796).

209. See Transcript of Oral Argument, *Bowman v. Monsanto Co.*, 569 U.S. 278 (2013) (No. 11-796).

210. See *id.* at 30. Relatedly, the Supreme Court recently denied certiorari in the Second Circuit's *Redigi* case, which involved the question of the application of the copyright first sale doctrine to digital files. See *Capitol Records, LLC v. ReDigi Inc.*, 910 F.3d 649 (2d Cir. 2018), *cert. denied*, *ReDigi Inc. v. Capitol Records, LLC*, 139 S. Ct. 2760 (2019). In that case, the Second Circuit affirmed the district court's holding that the defendant's platform, which transmitted a single digital copy of the purchased song to the purchaser while deleting the seller's original copy, "infringed the Plaintiffs' exclusive rights under 17 U.S.C. § 106(1) to reproduce their copyrighted works." *Id.* at 652–53.

211. *Bowman*, 569 U.S. at 289.

212. *Id.*

patentee—“rather than every [situation] involving a self-replicating product.”²¹³

The Court then went out of its way to explicitly leave open “whether or how the doctrine of patent exhaustion would apply”²¹⁴ to self-reproducing technologies in two particular situations. The questions the Court leaves open here—described in more detail below—are vague, and the opinion declines to explain precisely what they might refer to.

Lexmark did not explicitly tackle any of *Bowman*’s questions. But unlike *Bowman*, *Lexmark* did not limit its holding, and spoke widely and clearly about its conception of patent exhaustion. Furthermore, as explained above, *Bowman*’s holding remains good law post-*Lexmark*. Thus, although these questions remain apparently open, I argue that *Bowman* and *Lexmark* between them provide all the tools necessary to derive answers to *Bowman*’s loose ends.

This Part unpacks *Bowman*’s open questions, lays out the “Bowmark” framework, applies that framework to answer whether and how exhaustion might apply in each situation, and finally evaluates whether that answer is acceptable or whether affirmative changes are needed to reach a different result.

A. WHAT QUESTIONS *BOWMAN* LEFT OPEN, AND HOW TO APPLY “BOWMARK” TO CLOSE THEM

What I refer to as “*Bowman*’s two questions” are “whether or how the doctrine of patent exhaustion would apply”²¹⁵ when an item’s self-reproduction (1) “occur[s] outside the purchaser’s control,”²¹⁶ or (2) is “a necessary but incidental step in using the item for another purpose.”²¹⁷

Other scholars have suggested constructing novel paradigms to handle the types of questions that *Bowman* left open.²¹⁸ But no one has taken advantage of the Court’s recent clarity in *Lexmark* of how it conceives of exhaustion as a broad theory (as opposed to *Bowman*’s stubbornly limited holding) to revisit

213. *Id.*

214. *Id.*

215. *Id.*

216. *Id.*

217. *Id.*

218. See Adanna Uwazurike, *Remaking Making: Integrating Self-Replicating Technologies with the Exhaustion Doctrine*, 59 B.C.L. REV. 389, 416–18 (2018) (summarizing “a number of proposed solutions to the problem of self-replicating technologies and the threat to innocent infringers,” such as “expanding the use of contract law or property law in the seed industry or applying copyright regimes”); see also *id.* at 418–421 (suggesting that the definition of “making” for infringement of self-replicating technologies be limited to acts of making in which the actor knowingly infringed).

these questions and finally provide conclusive, explicit answers.²¹⁹ I propose to do so, using the “Bowmark” framework to guide the analysis.

The application of Bowmark proceeds in two steps. First, it determines whether there is infringement: Is there an act that implicates the underlying intangible patented knowledge?²²⁰ If not, there is no infringement, and thus no reason to dive into an exhaustion analysis. If there is infringement, the second step evaluates the situation through the lens of *Bowman* and *Lexmark*’s collective exhaustion themes.

The Bowmark guiding themes explain both *how* and *why* exhaustion is applied in the second step of the framework. The “hows” of exhaustion, according to Bowmark, are that: (1) exhaustion applies automatically upon authorized sale;²²¹ (2) sales are only “authorized” if they are conducted directly by the patentee or indirectly through a licensee in compliance with their license;²²² (3) exhaustion cuts off the patentee’s monopoly on “use,” “sale,” and “importation”²²³ (but *not* on active “making” that requires use of the underlying patented knowledge);²²⁴ (4) exhaustion applies separately and individually to each particular sold item;²²⁵ and (5) parties cannot contract around exhaustion.²²⁶

Bowman and *Lexmark* list a variety of justifications for exhaustion,²²⁷ but they are largely redundant. So “Bowmark” boils them down to just two:

219. Lim discusses *Bowman* post-*Lexmark* but does not fully address or unpack *Bowman*’s two open questions. See generally Lim, *supra* note 197, at 7–20; see also *infra* notes 252–253, 255 and accompanying text (discussing Lim’s analysis of *Bowman* in greater detail).

220. See Syed, *supra* note 30, at 8.

221. See *Impression Prods. v. Lexmark Int’l, Inc.*, 137 S. Ct. 1523, 1534–38, 1566 (2017) (stating that: restraints on alienation should be avoided; the patentee is entitled to only one “reward”—but not necessarily to the highest possible price—per item sold; the flow of commerce should not be impeded by “remora-like” patent rights; the doctrines of patent exhaustion and copyright first sale have the same, deep common law roots; exhaustion should be conceived as “a limit on the scope of the patentee’s rights”; licenses are about “exchanging rights,” whereas sales are about “passing title to a product”; “a license does not implicate the same concerns about restraints on alienation as a sale”; “only the patentee can decide” when to release an item from their use, sell, or import monopoly via exhaustion; and a sale authorized by the patentee is the affirmative manifestation of that decision) (internal quotations and citations removed); see also *Bowman v. Monsanto Co.*, 569 U.S. 278, 283–84 (2013).

222. See *Lexmark*, 137 S. Ct. at 1535.

223. See *id.* at 1534.

224. See *Bowman*, 569 U.S. at 288–89 (explaining that *Bowman* “was not a passive observer” and “controlled the reproduction” of the seeds in a way that actively utilized their patented trait).

225. See *Lexmark*, 137 S. Ct. at 1537; see *Bowman*, 569 U.S. at 283–84.

226. See *Lexmark*, 137 S. Ct. at 1538.

227. See *id.* at 1532.

(1) restraints on alienation of chattels should be avoided and (2) the patentee is entitled to only one “reward” per item sold.

B. QUESTION 1: SELF-REPLICATION OUTSIDE THE PURCHASER’S CONTROL

The first open question in *Bowman* is how exhaustion would be applied, if at all, when an item self-reproduces “outside the purchaser’s control.”²²⁸ To what situations could this question apply? I have come up with three plausible scenarios, which I have termed “set it and forget it,” “wind,” and “already rolling.”²²⁹ I will discuss each in turn below, applying the Bowmark framework to determine first if there is infringement, and then if there is exhaustion.

1. “Set It and Forget It”

Based on the oral arguments in *Bowman*, the Court may have had something along these lines in mind: a farmer obtains some amount of Monsanto seeds²³⁰ and “throw[s] the seeds on the ground,”²³¹ exerting (for the sake of argument) no “control” over, and doing no further work towards, their subsequent germination or growth. Nevertheless, “one or two of the [seeds] . . . grow . . . [.]”²³² I will refer to this situation as “set it and forget it.”

The application of the Bowmark framework first asks if there is any direct infringement. The infringement analysis in this situation is straightforward and was actually addressed briefly during oral argument. Yes, the act of throwing seeds on the ground such that some of those seeds later sprout can constitute infringement if it implicates the underlying intangible patented knowledge.²³³

228. *Bowman*, 569 U.S. at 289.

229. Note that, by shooting down Bowman’s “blame the bean” defense, the Court made it clear that they did *not* view this question to apply to a farmer applying routine levels of care towards their fields to usher new generations of crops into existence, when that care implicates patented knowledge. *See id.* at 288–89. As far as the Court is concerned, such a situation does represent “control” by the farmer, even if the farmer is not personally manipulating every cycle of DNA replication and every cell division. *See id.*

230. The scenarios posed in this and subsequent Sections use Monsanto as the hypothetical patentee, but note that the same analyses could apply to other entities holding patents on seeds, biotechnology, or other self-reproducing technology. These scenarios also all assume that the underlying patents are valid, and that every claimed element is present in the accused product or process.

231. Transcript of Oral Argument, *supra* note 209, at 12–13.

232. *Id.*

233. *See id.* at 13. Exactly how this act might implicate patented knowledge will depend on the particular patented trait(s) in question. Hypothetical traits for which a “set it and forget it” type act could implicate patented knowledge include drought resistance or growth patterns engineered to thrive unassisted in a particular soil composition. If the farmer has an assorted collection of seeds and knows that some have been engineered in such a way, then they could selectively propagate only the engineered ones merely by “tossing” them into an environment

It may not be the most efficient “use” of seeds, but it is still an act that uses the seeds without authority of the patentee (assuming no license) and could easily implicate underlying patented knowledge.²³⁴ Does the act of throwing these seeds on the ground constitute unauthorized “making”? Yes, to the extent that the “use” of tossing the seeds implicates the patented knowledge and that the seeds’ landing on the ground initiates at least one round of replication of patented DNA.

The exhaustion analysis is similarly straightforward: if the initial seeds were the subject of an authorized sale, then they were exhausted, and the farmer’s seed-tossing “use” is excused from infringement. However, any subsequent iterations of the patented products, including DNA sequences, are not exhausted. Thus, any growth from the “tossed” seeds, when the farmer “tossed” them based on underlying patented knowledge, constitutes unexhausted infringement. This is not different from the outcome in *Bowman*.²³⁵

2. “Wind”

The Court also may have had in mind the “wind” scenario, which looks something like this: Farm 1 is growing Roundup Ready® plants from seed that was subject to a Monsanto-authorized sale. Wind blows some seed from Farm 1 onto the neighbor’s farm, Farm 2, which is only growing conventional (not genetically modified) seed and has no relationship with Monsanto. The seed from Farm 1 then grows on Farm 2.²³⁶ Perhaps neither farmer is even aware of this wind-borne cross-contamination.²³⁷

they know would not support unengineered seeds, exactly in the same way that *Bowman* applied glyphosate to kill off unengineered plants.

234. See *Bowman v. Monsanto Co.*, 569 U.S. 278, 282 (2013) (explaining that *Bowman* used the knowledge about glyphosate resistance from Monsanto’s patents to direct his infringing acts).

235. For example, the farmer might choose to “ignore” the seeds specifically because of knowledge of the trait; for instance, knowing that certain seeds are drought tolerant might lead the farmer to decide to conserve water and selectively only water other plants.

236. See Transcript of Oral Argument, *supra* note 209, at 38.

237. This “wind” scenario is a common demon in the public discourse surrounding Monsanto, though, contrary to popular belief, Monsanto has yet to sue a farmer for patent infringement based on inadvertent cross-contamination. See Dan Charles, *Top Five Myths Of Genetically Modified Seeds, Busted*, NATIONAL PUBLIC RADIO: THE SALT (Oct. 18, 2012), <https://www.npr.org/sections/thesalt/2012/10/18/163034053/top-five-myths-of-genetically-modified-seeds-busted>. The two closest cases to this situation are *Monsanto Can. Inc. v. Schmeiser*, [2014] S.C.R. 902 (Can.), and *Organic Seed Growers & Trade Ass’n v. Monsanto Co.*, 718 F.3d 1350 (Fed. Cir. 2013), *cert. denied*, 134 S. Ct. 901 (2014).

In the *Schmeiser* case, Monsanto sued a Canadian canola farmer who was found to be growing Roundup Ready® seed without a license. See *Schmeiser*, [2014] SCC at 912. In that

Applying Bowmark here is complicated by the presence of multiple parties. There is Farmer 1, who purchased the seeds, and Farmer 2, in whose field some of the seeds sprouted. Is there direct infringement? Assuming Farmer 2 is performing acts on their field that result in plant growth and implicate the patented knowledge,²³⁸ then yes, Farmer 2's actions constitute direct infringement for the same reasons as in the "set it and forget it" scenario. But if the plants are just randomly growing in a field that Farmer 2 owns—maybe "Farmer 2" isn't even a farmer at all. Maybe they are just the next landowner over and couldn't tell soybeans from sugar beets; then perhaps there are no acts of direct infringement. In that scenario, we do not even reach the question of exhaustion.

If there is direct infringement, we may then also ask about indirect infringement, which could implicate Farmer 1's actions. In order to be liable for *inducing* Farmer 2's infringement, Farmer 1 must have (1) actively induced Farmer 2's infringement and (2) either known that Farmer 2's actions constituted infringement or been willfully blind thereto.²³⁹ This could go either way, depending on the facts. If the two farms are very close to one another—say, directly abutting—and Farmer 1 knows that Farmer 2 does not grow Monsanto²⁴⁰ seed, then it may be relevant whether, for example, Farmer 1's farming practices align with the stewardship practices put forth in Monsanto's Technology Use Guide for reducing cross-contamination.²⁴¹ If Farmer 1 is doing everything within reason to reduce pollen spread, yet it still occurs, then

case, nearly all of Schmeiser's crop was found to be Roundup Ready®, indicating that, even if the initial seed had been blown in by a neighboring farm, Schmeiser went out of his way to select and propagate the plants embodying Monsanto's patents by spraying Roundup® weed killer and saving and replanting seed from the surviving plants. *See id.* at 912, 929. In that case, the Supreme Court of Canada found Schmeiser liable for patent infringement. *See id.* at 937.

In contrast, in *Organic Seed Growers*, a group of farmers and affiliates sued Monsanto with the aim of obtaining declaratory judgements that they could not be held liable for infringing Monsanto's patents for inadvertent contamination in their fields. *See Organic Seed Growers*, 718 F.3d at 1352. The district court dismissed the case for lack of case or controversy because Monsanto had "made binding assurances that it will not 'take legal action against growers whose crops might inadvertently contain traces of Monsanto biotech genes (because, for example, some transgenic seed or pollen blew onto the grower's land)'" *Id.* (citations omitted). The Federal Circuit affirmed. *See id.*

238. For example, maybe Farmer 2 knows that Farmer 1 next door is growing plants exhibiting a patented trait conferring resistance to a particular pest. Then, after a particularly windy planting, Farmer 2 decides to not apply their usual pesticide for that pest to the edge of their field that borders Farmer 1's field, knowing that any seed that blew over will not need it.

239. *See Glob.-Tech Appliances, Inc. v. SEB S.A.*, 563 U.S. 754, 766–71 (2011).

240. I am using "Monsanto" here as a stand-in for convenience, but the hypothetical would work equally well with any other entity selling patented seed.

241. *See Bayer, supra* note 92, at 13–14.

Farmer 1 will likely not be held to be an induced infringer. After all, even Monsanto's own Technology Use Guide acknowledges that "[i]t is generally recognized in the industry that a certain amount of incidental, trace level pollen movement occurs, and it is not possible to achieve 100% purity of seed or grain in any crop production system."²⁴²

But if Farmer 1's practices are unreasonable such that they likely knew or were at least willfully blind to the fact that Farmer 2's acts would infringe—maybe they told Farmer 2 all about how to use their special seeds, and then set up giant fans specifically to blow pollen into Farmer 2's fields (or, less wildly, maybe they merely sowed their seeds unreasonably close to the property line)—then it is possible that Farmer 1 will be liable for induced infringement.

Depending on how the patent claims are structured and what actions Farmer 2 takes, it is possible that Farmer 1 may also be liable for contributory infringement. Continuing from the situation above, perhaps Farmer 1 is aware that their farming practices spread a significant amount of seed into Farmer 2's field, and specifically into open containers of seed that Farmer 2 is about to plant. Furthermore, suppose the patents covering Farmer 1's seed contain a claim similar to claim 130 of U.S. Patent No. RE39,247E in *Bowman*, which covers a "method for selectively controlling weeds" comprising planting Roundup Ready® seeds in a field and spraying the field with Roundup® weed killer.²⁴³ Finally, suppose that after Farmer 2 plants their seed, Farmer 1 mentions to them that they might get good results if they spray Roundup® weed killer, which Farmer 2 then does. In this scenario, Farmer 1 may be liable for contributory infringement for providing Farmer 2 with the seeds necessary to infringe that claim.

Having found infringement, the second step under the *Bowman* framework is whether or not patent rights have been exhausted. First, any seeds that Farmer 1 purchased with authorization from Monsanto are themselves exhausted. However, any subsequent generations—including "micro-makings" of patented items on the molecular and cellular scales—are not exhausted. Thus, any of the "wind" infringement events that involve seed or DNA copies that were not physically subject to an authorized sale are not exhausted. Those are newly infringing items, and, as such, would need to be the subject of an authorized sale or license to excuse their use or making from infringement.²⁴⁴

242. *Id.* at 13.

243. *See* *Monsanto Co. v. Bowman*, 657 F.3d 1341, 1344 (Fed. Cir. 2011).

244. Indeed, Justice Breyer acknowledged this in the *Bowman* oral arguments. Counsel for *Bowman* categorized Monsanto's argument as "any cell division is patent infringement" regardless of how one obtains the seeds. Transcript of Oral Argument, *supra* note 209, at 13.

3. “Already Rolling”

Alternatively, “outside the purchaser’s control” could also encapsulate a hypothetical situation in which the self-replication was already underway at the time of the purchase, which I will refer to as “already rolling.” This is similar to the “wind” scenario in that more than one party is involved, but is further complicated by the element of time. In this situation, self-replication has already been set in motion when the authorized sale takes place.

For instance, imagine that a farmer purchases a field that is already growing a Roundup Ready® crop. At the time of sale, the field contains some seeds that have not sprouted, some immature first-generation plants, some mature first-generation plants that have begun to set seed, and some second-generation seed. This situation fits squarely into those described above: any physical embodiments of a patent that are already present in that field are immediately exhausted upon the sale, assuming the sale is authorized. Thus, any uses that are contained to the copies that existed at the time of sale are excused from infringement. In effect, the seller sold whatever generations existed in the freeze-frame of the time of sale, whether those generations were what one might think of as “whole” (containing fully-formed plants or seeds) or “partial” (imagine, say, an ovule in a mother plant’s flower that was in the process of being fertilized into a second generation at the split second that the sale went into effect). It may be easier to imagine the items implicated in this situation, not as individual “plants” or “seeds” per se, but as a distinct collection of cells containing a given number of copies of particular DNA sequences, frozen in time at the time of sale. Regardless, whatever existed at the time of sale is exhausted—and whatever grows thereafter is not, leaving the purchaser open to potential infringement liability.

However, any embodiments that come into existence after the sale—even immediately after the sale—are not themselves subject to an authorized sale.²⁴⁵ And any subsequent acts that apply patented knowledge to tend to the newly-purchased field, when those acts usher additional copies into existence, are newly infringing acts of “making.” Because copies that came into existence after the sale are not themselves the subject of an authorized sale, they are not exhausted. Therefore, post-sale acts of farming that implicate the patented

Justice Breyer then agreed, and suggested that, for the same reason, it “is just as much a violation” when someone “buys generation 1 from Monsanto . . . , [and] they plant it in the ground and lo and behold up comes generation 2 But I think . . . that [Monsanto’s] response [to] that is, yes, you’re right, it is just as much of a violation. That’s why we, Monsanto, give the buyer a license to do it.” *Id.* at 13–14.

245. This depends on the exact terms of the sale agreement.

knowledge, such as spraying Roundup® weed killer on the field, constitute unexhausted infringement by unauthorized “making.”

Interestingly, if there are any copies that come into existence during a gap between the sale and the onset of the purchaser’s acts of farming, then those copies²⁴⁶ would not be derived from an infringing act.²⁴⁷ This is similar to the version of the “wind” scenario in which Farmer 2 does nothing, and the seeds grow on their land anyway. Here, the purchaser similarly did not act on the new iterations. Thus, the generation of these “gap” copies does not constitute unauthorized “making,” so their existence does not represent infringement.

In contrast, subsequent “use” of these new copies that implicates the patented knowledge—regardless of whether the new copies were “made” by an infringing act—*does* constitute infringement. Furthermore, that infringement is not exhausted because these copies (“gap” or otherwise) were not the subjects of an authorized sale.

The field as a whole, then, might have several different infringement and non-infringement, and exhaustion and non-exhaustion, scenarios at play at once. Oddly, these various scenarios may all even be present in microcosm within a single plant, if that plant’s existence spanned the sale and the onset of the purchaser’s activities. Such a plant might end up being a chimera of at least the following situations: (1) cells and nucleic acids that existed before the sale, the purchaser’s unauthorized “use” of which does not implicate the patented knowledge but would be an infringing act but for the sale’s exhaustive effect; (2) same as (1), but the “use” *does* more than incidentally implicate the patented knowledge, so exhaustion does not apply; (3) cells and nucleic acids that came into existence during a gap between the sale and the onset of the purchaser’s acts, the purchaser’s patented-knowledge-implicative use of which constitutes unexhausted infringement; and (4) cells and nucleic acids whose existence is derived from the purchaser’s farming activities, the use *and* making of which, to the extent that they implicate patented knowledge, constitute infringement by the purchaser. Perhaps those distinctions would not be useful in practice, or perhaps they would be of some help during damages calculations. Either way, the point still stands that such a situation, bizarre as it may seem, can exist but is fully capable of dissection by the Bowmark framework.

Yet there is still more to this “already rolling” scenario: The seller’s actions, too, should be analyzed for infringement. Any of the above acts by the purchaser that constitute direct infringement could make the seller liable for

246. Consider, say, the resulting DNA from cell divisions that were in progress at the time of sale.

247. This is only if there is such an actual gap, and not, say, a contractual instantaneous transfer of control over ongoing farming activities.

indirect infringement. This analysis is essentially the same as that for Farmer 1 in the above “wind” scenario, with the result being that the seller may very well be liable. For example, Monsanto’s limited use license specifies that if a licensed grower sells land containing its patented technology, the grower must notify the purchaser that (1) the contents of the field are subject to the limited use license with Monsanto and that (2) the purchaser “must have or obtain their own Monsanto Technology Stewardship Agreement to harvest or use, transfer or sell the harvested crop.”²⁴⁸ This indicates that the seller is aware that the purchaser will need a license to avoid infringement for farming that field. Therefore, if the seller does *not* inform the purchaser that they need to obtain their own the licensing agreement, and the purchaser does not obtain a license and is thus found to infringe, then the seller may be liable for inducing and/or contributing to that infringement. Furthermore, if the seller did *not* give such requisite notice to the purchaser, then the sale would not be an *authorized* sale in the first place, since it did not comply with the terms of the seller’s own license. In this case, the seller would be liable for direct infringement (unauthorized selling) as well as the aforementioned indirect infringement. Further, absent an authorized sale, any infringement in the “already rolling” scenario would enjoy no exhaustion whatsoever.

4. *Summary and Analysis of Question 1’s Answers*

Lim states that “[f]uture cases must clarify the role of intent.”²⁴⁹ But application of *Bowmark* in the above situations shows that the case law we already have can dissect how unexhausted infringement might arise on the part of someone who may not consider themselves in “control” of a patented product’s self-replication. The facts may be complicated, but the equation is simple: Unauthorized acts of making, using, selling, and importing a patented item infringe when they implicate underlying patented knowledge, and exhaustion applies *only* to the closed set of physical copies that were subject to an authorized sale. Whether it is actually worth the patentee’s time, money, and effort to pursue litigation in any of the above situations does not change the underlying fact that many of them represent acts of unexhausted infringement.²⁵⁰

248. Bayer, *supra* note 92, at 62 part 1(j).

249. See Lim, *supra* note 197, at 26.

250. For example, in the *Bowman* oral arguments, counsel for Monsanto stated that “[t]he point that there may be many farmers . . . that may have some inadvertent Roundup Ready [crops] in their fields may be true, although [] it is not well documented. There would be inadvertent infringement if the farmer was cultivating a patented crop, *but there would be no enforcement of that. The farmer wouldn’t know, Monsanto wouldn’t know, and in any event, the damages would be zero* because you would ask what the reasonable royalty would be, and if the farmer

Is this the “right” outcome, from a policy perspective? How “should” exhaustion work in the context of self-replication “outside the purchaser’s control?” Self-replication is still replication, meaning it still conjures a new embodiment of the patented knowledge into existence, one that was never subject to an authorized sale. By the policy embraced in *Lexmark*,²⁵¹ such an event is completely unrelated to exhaustion. Regarding the lack-of-control aspect, the results are still fair because infringement, as always, still demands the alleged infringer to have committed *an act* that implicates the patented knowledge.²⁵² Such an act results in infringement as soon as it intersects with even microscopic movements of the DNA-replicating enzymes as they physically execute and “control” the self-replication; the infringer never needs to know they were infringing in order to infringe.²⁵³

C. QUESTION 2: SELF-REPLICATION AS A NECESSARY BUT INCIDENTAL STEP TOWARDS ANOTHER PURPOSE

The second open question in *Bowman* is how exhaustion would be applied, if at all, when “the article’s self-replication . . . [is] a necessary but incidental step in using the item for another purpose.”²⁵⁴ Lim reads this question as answering itself; he states that “the Court qualified that multiplication per se would not negate exhaustion since reproduction as an incidental and integral

doesn’t want Roundup Ready technology and isn’t using Roundup Ready technology to save costs and increase productivity, [] the royalty value would be zero.” Transcript of Oral Argument, *supra* note 209, at 40–41 (emphasis added).

251. See *supra* note 221 (explaining the underlying policy of *Lexmark*).

252. Lim argues that “the better solution [is] for licensees to deal with these issues” rather than farmers, because “[a]gribiotech patentees and their immediate downstream licensees are most knowledgeable about the technology” and are better equipped to handle litigation. Lim, *supra* note 197, at 18. However, farmers who are growing crops from proprietary seeds (as opposed to using the seeds as commodities for feed) *are* themselves properly licensees, and infringement on the farmer’s part still boils down to the same basic principles as infringement by anyone else. Thus, it is not helpful to distinguish between “farmers” and “licensees” in terms of the infringement analysis.

253. *But cf.* Lim, *supra* note 197, at 17–18 (arguing that direct liability for farmers and seed companies “should not rest on knowledge. The better reading is for farmers to be found liable as active inducers of the patent infringement their though [*sic*] willful blindness or actively encouraging the [seed companies’ direct] infringement [of selling for uses outside the terms of their license]. . . . Farmers’ liability springs not from their knowledge of the restrictions, but the inducement of the company’s infringement,” and that it is better if “the burden rest[s] upon patentees and their licensees rather than farmers”). But Lim’s discussion of “knowledge” here confuses the issue, which is not about knowledge *of the infringement*, but use of the knowledge covered by the patent. See Syed, *supra* note 30, at 10–11. Farmers may or may not be “willfully blind” to the existence of a relevant patent while still directly infringing by using the knowledge covered by the patent.

254. *Bowman v. Monsanto Co.*, 569 U.S. 278, 289 (2013).

step in using the item for a lawful purpose was acceptable.”²⁵⁵ However, the *Bowman* opinion made no such “qualification”; it explicitly left open “*whether or how* the doctrine of patent exhaustion would apply in such circumstances.”²⁵⁶

In a “typical” exhaustion situation, making a patented article is not incidental to using a patented article, so making is not exhausted. But *Bowman* Question 2 begs the question: What degree of “making” is “incidental” for self-replicating technology? The opinion gives only one clue: a cryptic “CP” citation to, and parenthetical quote of, 17 U.S.C. § 117(a)(1): “[I]t is not [a copyright] infringement for the owner of a copy of a computer program to make . . . another copy or adaptation of that computer program provide[d] that such a new copy or adaptation is created as an essential step in the utilization of the computer program.”²⁵⁷ But that hint is only of limited value: it describes “use” in the singular (“*the* utilization”), whereas *Bowman*’s Question 2 is only directed towards uses that are “for *another* purpose,”²⁵⁸ implying that there may be different possible uses at play.

What, then, is “another” purpose? Answering this question requires first defining a “primary” purpose for the self-reproducing item in question. This would likely be fact specific, depending on the type of technology in question. One might answer this question at the market level by looking to the subject of the license (if there is one) with the patent holder. Is the user a licensed seed supplier who is “supposed to be” using the technology to produce seeds to sell to farmers? Or is the user a farmer who is “supposed to be” using the technology to produce crops to sell as commodities? Alternatively, one might define the “primary purpose” as the goal of the underlying invention, such as easing weed management²⁵⁹ or extending shelf-life for consumers.²⁶⁰ What uses might be “secondary” to these “primary” uses? Some possibilities are research, breeding, and the right to repair. I consider each in turn below.

1. *Research and Breeding*

Research is a “secondary” use because it aims to produce *new knowledge from* a self-replicating item, rather than to (re)produce the item itself. Although “research” is not mentioned in the *Bowman* opinion, it arose briefly several

255. Lim, *supra* note 197, at 8 (citing *Bowman*, 569 U.S. at 289).

256. *Bowman*, 569 U.S. at 289 (emphasis added).

257. *Id.* (alteration in original).

258. *Id.* (emphasis added).

259. As would be the case for Roundup Ready® crops. See *Bowman*, 569 U.S. at 280–81.

260. For example, the Arctic Apple®, which has been engineered to resist browning. See Okanagan Specialty Fruits, *Frequently Asked Questions*, ARCTIC APPLES, <https://www.arcticapples.com/arctic-apples-r/faq/> (last visited Nov. 27, 2019).

times in oral argument,²⁶¹ indicating that the Court may have considered it when writing *Bowman's* Question 2. Activities in this category could be anything from using a patented gene sequence as a starting point to design a new product, to simply conducting research on a topic unrelated to the patent but using samples that contain a proprietary gene. The “research” question could thus be divided up between research that is specifically *about* that underlying knowledge, as in the former example, versus research that just happens to *involve an item* embodying the patented knowledge,²⁶² as in the latter example.

Similarly, “another purpose” could encompass using proprietary plants in a breeding program to create new varieties. Neither the *Bowman* opinion nor the oral argument transcript explicitly mentions breeding. But both discuss PVPA exemptions,²⁶³ and the PVPA research exemption also includes breeding.²⁶⁴ Thus, the Court may have considered breeding in addition to research.

That being said, the discussion of *J.E.M. Ag Supply*²⁶⁵ in *Bowman*²⁶⁶ strongly suggests that the Court would *not* envision exhaustion to absolve breeding or research uses from infringement, at least for self-replicating goods for which PVP protection is available. As described earlier, PVP certificates have exemptions for saving seed and for breeding and research, but patents do not.²⁶⁷ The *Bowman* Court explicitly used the saved-seed distinction between the PVPA and patent law to justify declining to apply exhaustion to patented seeds saved for replanting.²⁶⁸ This strongly suggests that the Court would be equally unwilling to allow the other PVPA exemption to sneak into patent law via exhaustion.

Aside from the PVPA, the Bowmark framework can still be applied to breeding and research. First, breeding and research activities that are

261. See Transcript of Oral Argument, *supra* note 209 at 22–23.

262. Indeed, Justice Kagan posed a hypothetical situation during oral argument in *Bowman* about “a 10-year-old who wants to do a science project of creating a soybean plant, and he goes to the supermarket and gets some edamame, and it turns out that it’s Roundup seeds.” Transcript of Oral Argument, *supra* note 209, at 38.

263. See *Bowman v. Monsanto Co.*, 569 U.S. 278, 286 (2013) (discussing the PVPA’s seed saving exemption); Transcript of Oral Argument, *supra* note 209, at 22–23 (discussing the PVPA’s seed saving and research exemptions).

264. See 7 U.S.C. § 2544 (2018) (stating that “[t]he use and reproduction of a protected variety for *plant breeding or other bona fide research* shall not constitute an infringement of the protection provided under this chapter”) (emphasis added).

265. *J.E.M. Ag Supply, Inc. v. Pioneer Hi-Bred Int’l, Inc.*, 534 U.S. 124 (2001).

266. See *Bowman*, 569 U.S. at 286.

267. *Supra* notes 70–72 and accompanying text.

268. See *Bowman*, 569 U.S. at 286.

specifically directed to expanding or applying the underlying knowledge of the patent in question would likely be considered infringing conduct as unauthorized “uses.” If the breeding or research merely involves a sample containing, say, a proprietary gene sequence, but is not directed *per se* to the knowledge embodied in that sequence, then there might still be infringement. The presence of the proprietary sequence, even if itself not the subject of the research, may affect the properties of the whole system such that any acts to produce new knowledge inextricably implicate the patented knowledge.²⁶⁹

If it is infringement, then exhaustion is still only available for items subject to an authorized sale. Thus, research or breeding that did *not* derive from such an item (perhaps a researcher synthesized DNA that included a patented sequence, or a grape breeder²⁷⁰ obtained proprietary parent lines from a third party who was not authorized to sell them) would not be “saved” from infringement by exhaustion.²⁷¹

2. *Repair*

In general, repair can be subject to exhaustion in as much as it is a “use” of an item that was the subject of an authorized sale.²⁷² Indeed, “*Lexmark* clearly considered reuse or repair of a purchased item a normal incident of ownership protected by the exhaustion doctrine.”²⁷³ However, “courts have drawn a distinction between repair and reconstruction” such that “*repair* is permissible, [but] *reconstruction* of a patented product amounts to the making of a new article and thus constitutes patent infringement.”²⁷⁴ Although “[c]ourts

269. In the same way that the choice of model organism or specific cell line is so important in biology research. The outcome here would likely be highly fact-dependent as to how the trait or sequence in question affects the overall structure and/or function of the organism, and one may imagine various different fact patterns that could come out either way.

270. Although “most grapes are propagated from cuttings, and not grown from seeds,” new grape varieties may still be produced by breeding. Bruce I. Reisch & Philip Stewart, *Grape Breeding Procedures*, CORNELL GRAPE BREEDING: GRAPE FLOWERS (2001), <http://www.hort.cornell.edu/reisch/grapegenetics/breeding/crossing1.html>.

271. Indeed, a district court, citing *Lexmark*, recently found that exhaustion did not apply to sales of grape vine cuttings by a third party where the patentee had authorized the third party to grow and sell grapes from vines owned by the patentee, but not to sell the vines themselves or cuttings thereof, on which the patentee retained ownership. *Int’l Fruit Genetics, LLC v. Orcharddepot.com*, No. 4:17-CV-02905-JSW, slip op. at 4–5 (N.D. Cal Feb. 12, 2018).

272. See Grinvald & Tur-Sinai, *supra* note 22, at 100.

273. Kobak, Jr., *supra* note 18, at 622; see *Impression Prods. v. Lexmark Int’l, Inc.*, 137 S. Ct. 1523, 1532 (2017) (stating that “a shop that restores and sells used cars . . . is free to repair and resell those vehicles”).

274. Grinvald & Tur-Sinai, *supra* note 22, at 100 (emphases added). With that in mind, it is rather odd that *Lexmark* referred to “the parties . . . who refilled and resold the toner containers . . . as ‘remanufacturers’ rather than repairers.” Kobak, Jr., *supra* note 18, at 622.

have struggled in drawing th[is] line,”²⁷⁵ courts discussing exhaustion “have clarified that repair may entail the replacement of spent elements and, yet, still be permissible”²⁷⁶—*unless “the replacement part itself is protected by a utility or design patent.”*²⁷⁷ That caveat will become crucial.

But what does “repair” mean for a self-reproducing technology?²⁷⁸ Perhaps a farmer who wants to grow only conventional crops aims to “repair” a field by removing inadvertent contamination from nearby transgenic fields. Alternatively, perhaps a farmer identifies a genetic mutation in some of their plants and wishes to “correct” it back to the patented sequence.²⁷⁹

First, do either of these scenarios represent infringement? To do so, the infringing “act” would need to both use the patented knowledge *and* cross the line from permissible “repair” to impermissible “reconstruction.”²⁸⁰ For the former case, the farmer arguably uses the patented knowledge by selectively plucking out all seedlings that emerge “too early,” when they know their conventional seed should germinate later than their neighbor’s proprietary seed.²⁸¹ However, such an act arguably falls short of “reconstruction”—if anything, it’s *de*construction—and is therefore likely non-infringing.

275. Grinvald & Tur-Sinai, *supra* note 22, at 100.

276. *Id.* at 112 (first citing *Aro Manufacturing Co. v. Convertible Top Replacement Co.*, 365 U.S. 336, 345–46 (1961); then citing Joshua D. Sarnoff, *White Paper on Protecting the Consumer Patent Law Right of Repair and the Aftermarket for Exterior Motor Vehicle Repair Parts: The PARTS Act, S. 812, H.R. 1879, 115th Congress 3* (Unpublished White Paper, Nov. 2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3082289).

277. Grinvald & Tur-Sinai, *supra* note 22, at 112 (emphasis added). “[W]hen parts are protected by patents and they need to be replaced in the course of repair, *it is only the patent holder who can make and supply those parts.* Registration of a patent over a part of a product could, thus, be used to circumvent the application of the exhaustion doctrine that would otherwise sanction repair of the product.” *Id.* at 112–13 (emphasis added) (citations omitted).

278. Interestingly, as of this writing, there appear to be no reported cases involving the right to repair in plant patents, or, more generally, in the contexts of seeds, DNA, or genes.

279. Note that this situation is not likely common today, but given the rapid advancement of genetic engineering technologies, this sort of thing is likely to be within even a layman’s toolbox in the not-too-distant future. I do not delve into whether or how any regulatory issues associated with the creation of new genetic modifications in crops would come into play here.

280. *See* Grinvald & Tur-Sinai, *supra* note 274 and accompanying text.

281. Note that this is a hypothetical scenario; conventional seed does not necessarily germinate on a different schedule than patented seed. This scenario would essentially be the inverse of what Bowman did. Bowman applied the patented knowledge of glyphosate resistance to select *for* plants covered by Monsanto’s patent. *See* *Bowman v. Monsanto Co.*, 569 U.S. 278, 285 (2013). Bowman’s identification of glyphosate resistant plants necessarily killed the non-resistant plants. But one might imagine other traits, the knowledge of which would allow one to just as easily select *against* those plants. Hypothetical (but realistic) examples of such traits include differences in height, color, or maturation time.

For the latter case, the farmer also uses the patented knowledge of the function of the gene to guide the repair process. This repair is arguably a permissible “replacement of spent elements.”²⁸² Such repair would, therefore, be safe from infringement, unless the sequence of the gene to be “repaired” is itself covered by a patent. If so, such an act would become infringement because of the repair caveat about proprietary replacement parts.²⁸³ Thus, “repair” of a self-reproducing technology could become infringement.²⁸⁴

The first repair scenario did not represent infringement, so there is no need to bring exhaustion into that analysis. But would the gene repair scenario, which infringed through the use of a proprietary replacement gene, be absolved by exhaustion? That comes down to whether the replacement gene had been subject to an authorized sale. At this point, the situation is similar to the research situation described earlier, which entailed the creation and use of additional copies of a patented gene sequence. Here, it would be the replacement part with the “correct” sequence. Like in the research case, the genetic repair here would be ineligible for exhaustion if the replacement gene was not obtained through an authorized sale. Furthermore, unauthorized creation of copies of the replacement gene would also be unexhausted. Thus, infringing “repair” of a self-reproducing technology could be ineligible for the exhaustion exception. But note that “fixing” does not itself present the exhaustion issue here; rather, the exhaustion problem comes from creating new copies of the replacement gene. Absent that issue, genetic “fixes” accomplished by tinkering with the innards of the seed are likely to be exempted as permissible repairs, even though they may use patented knowledge.

3. *Summary and Analysis of Question 2’s Answers*

The application of *Bowmark* to Question 2 shows that exhaustion may, but does not necessarily, apply to self-reproduction that might occur in contexts in which production of the patented good is not necessarily the user’s goal, such as research, breeding, and repair. Is that a good policy? While I have highlighted particular fact patterns that do represent unexhausted infringement, it would be equally easy to tweak those facts and reach the opposite result. This multitude of possible fact patterns itself demonstrates that, even if some activities are “closed” to unauthorized parties, there are still

282. See Grinvald & Tur-Sinai, *supra* notes 276–277 and accompanying text.

283. See *supra* notes 276–277 and accompanying text, stating that repairs, although typically permissible, are *not* permissible when they involve the insertion of a part that is itself the subject of patent protection.

284. Whether a patent owner would necessarily *sue* over such activities is beside the point.

a variety of other possible and permissible uses. Furthermore, depending on the facts, the actor might find an out through another exemption.²⁸⁵ This is entirely consistent with how exhaustion has always worked and is therefore equally as fair as the application of exhaustion in any other, non-self-replicating scenario.

VI. CONCLUSION

The Bowman Court ultimately concluded that, in applying patent exhaustion, it is fair to draw a line between “use” and “making,” even for self-replicating goods, if only for the narrow set of facts of that case. But the Supreme Court openly refused to wade into the application of exhaustion in other more complicated scenarios involving questions of control over, or alternative uses of, a patented self-replicating technology.²⁸⁶ Although *Bowman*’s two questions remain open on their face, the theoretical underpinnings of *Bowman* and *Lexmark* (what I have termed “Bowmark”) are fully capable of handling them.

Deploying Bowmark on *Bowman*’s questions reveals that there are fact patterns squarely within each question that constitute unexhausted infringement. The examples dissected above are not exhaustive.²⁸⁷ The point is not that *Bowman*’s questions are *always* or *never* exhausted, but that the framework can work through even complex fact patterns such as those above.²⁸⁸ *Bowman*’s open questions may appear unassailable when imagined as monolithic, theoretical quandaries, but they are just big buckets for small fact patterns. And those fact patterns, although perhaps technical, do not require new law.²⁸⁹

Even if the old law *can* handle exhaustion of self-replicating technologies, is the confusion and frustration surrounding the issue²⁹⁰ itself a symptom that

285. This could, for example, include experimental use, but that exemption only applies in the very narrow context of uses “solely for amusement, to satisfy idle curiosity, or for strictly philosophical inquiry” *Madey v. Duke Univ.*, 307 F.3d 1351, 1362 (Fed. Cir. 2002).

286. *See Bowman v. Monsanto Co.*, 569 U.S. 278, 289 (2013).

287. Pun intended.

288. Bowmark would work equally well on different facts and, depending on the facts, may lead to different outcomes in terms of the presence or absence of exhaustion.

289. Thus, contrary to what some have proposed, there is no need to impose new carve-outs or spin-offs of the exhaustion doctrine specific to self-replicating technologies. *But see* Uwazurike, *supra* note 218, at 422 (proposing a new definition for “making” for the application of exhaustion to self-replicating technologies, which would contain a knowledge requirement); Ghosh & Calboli, *supra* note 191, at 173 (characterizing *Bowman* as “tailoring” the exhaustion doctrine to have special and different rules for self-replicating technologies).

290. Monsanto has been dubbed “one of the most-hated large companies in the world,” due, in large part, to the very technology that was at issue at *Bowman*. Caitlin Dewey, *Why*

society is bristling for something like a fair use exception for patent law? Others have discussed fair use in patent law in more detail,²⁹¹ but it is noteworthy that the PVPA, which was tailor-written for self-replicating technology, contains exemptions for saving seed, breeding, and research.²⁹²

Perhaps there should be more room for fair use in patent law. But, if so, that should be an explicit policy choice through legislation, not piecemeal alterations of the exhaustion doctrine, which, it must be remembered, is actually quite narrow. The purpose of exhaustion is to smooth the flow of commerce of chattels while ensuring that the patentee receives payment for their invention.²⁹³ Exhaustion, unlike fair use, is not about supporting rival or replicative activity, “or the type of rights that buyers expect to receive.”²⁹⁴

Even if the narrow application of exhaustion seems harsh or unfair for self-replicating items, it is noteworthy that it aligns with other orthogonal policies that apply to many of the same goods. Entirely aside from patent law, we generally do not allow people to simply make self-replicating products like genetically modified crops or biologically-produced drugs without fairly stringent oversight.²⁹⁵ And farmers have purchased new seed each season for certain crops long before the advent of recombinant DNA technology.²⁹⁶ With that in mind, is the potential inapplicability of patent exhaustion to particular scenarios involving self-replicating goods so bizarre?

‘Monsanto’ is no more, WASH. POST (June 4, 2018), <https://www.washingtonpost.com/news/wonk/wp/2018/06/04/why-monsanto-is-no-more/>. That being said, Monsanto’s (now Bayer’s) Roundup Ready® technology is merely one example of a self-replicating trait (even if a dominant one), and it is not necessarily “morally” representative of the entirety of self-replicating goods.

291. See, e.g., Maureen A. O’Rourke, *Toward a Doctrine of Fair Use in Patent Law*, 100 COLUM. L. REV. 1177 (June 2000).

292. See 7 U.S.C. §§ 2543–44 (2018). Furthermore, in 2005, the Supreme Court created a narrow fair-use-like exception for the pharmaceutical industry. See *Merck KGaA v. Integra Lifesciences I, Ltd.*, 545 U.S. 193, 208 (2005) (holding “that the use of patented compounds in preclinical studies is protected . . . as long as there is a reasonable basis for believing that the experiments will produce the types of information that are relevant to” applications to the Food and Drug Administration for approval of human clinical trials or marketing of new drugs) (citations omitted) (internal quotation marks omitted).

293. See *Impression Prods. v. Lexmark Int’l, Inc.*, 137 S. Ct. 1523, 1538 (2017).

294. *Id.*

295. See, e.g., USDA, 6. *What are the roles of government in agricultural biotechnology?*, BIOTECHNOLOGY FAQs, <https://www.usda.gov/topics/biotechnology/biotechnology-frequently-asked-questions-faqs> (last visited Nov. 27, 2019) (summarizing regulations on genetically modified organisms); United States Food & Drug Administration (FDA), *Development & Approval Process*, DRUGS, <https://www.fda.gov/drugs/development-approval-process-drugs> (last visited Nov. 27, 2019) (describing drug approval requirements).

296. See Kloppenburg, *supra* note 79, at 99.

Along similar lines, perhaps it is good public health policy (albeit on a private level) to have self-reproducing goods constantly tied back to the patent owner. This is because this implicitly encourages the expectation that propagation of the product would be controlled, as much as possible, by the patent owner, which implies a degree of responsibility for that product, for better or for worse.²⁹⁷ When these technologies inherently have the ability to propagate themselves, the presence of strict stewards is crucial. Having something like a heightened mental state requirement for infringement of self-replicating goods²⁹⁸ might absolve “innocent” infringers and avoid some of the odd infringement scenarios described above. But not having a heightened standard incentivizes legitimate purchasers and users to be good stewards of the technology, and to actively work towards minimizing its “leakage,” “contamination,” or “escape.” Regardless of whether the incentive comes from fear of an infringement suit or some feeling of environmentalism, the positive result on an ecological level is the same. Inevitably, however, some level of leakage is likely, and may just be an occupational hazard. But the exhaustion doctrine helps align and balance incentives to maintain control of the self-replicating technology on both sides of the equation. The patentee wants to maintain its monopoly, on one side, and a farmer does not want their seeds to spread downwind, on the other.

Furthermore, as DNA sequencing becomes cheaper, faster, and easier, tracking the spread of biological self-replicating technologies—and the identities of their patent owners—will be trivial, hardly conjuring the “remora” image that Justice Roberts used to explain the necessity of exhaustion.²⁹⁹ On the other hand, in some situations it may be advantageous to allow the purchaser-user to continue using and/or making a patented self-replicating technology through one or more use-make cycles after only a single initial purchase.³⁰⁰ But that, too, does not require altering the exhaustion doctrine.

297. Interestingly, corn farmers sued seed companies for allowing the companies’ genetically modified seed to contaminate the farmers’ corn supply, resulting in economic loss for the farmers. *In re Starlink Corn Prods. Liab. Litig.*, 212 F. Supp. 2d 828, 833 (N.D. Ill. 2002).

298. See Uwazurike, *supra* note 218, at 422 (proposing a new definition for “making” for exhaustion of self-replicating technologies, which would contain a knowledge requirement).

299. See *Impression Prods. v. Lexmark Int’l, Inc.*, 137 S. Ct. 1523, 1526 (2017) (stating that “[a]llowing patent rights to stick remora-like to that item as it flows through the market would violate the principle against restraints on alienation”).

300. Such as, for example, in the space industry, or in the context of microbial production systems. See, e.g., Kalil, *supra* note 7 (discussing self-replicating structures for space exploration); Waschulin & Specht, *supra* note 10 (discussing microbial bioreactors); Service, *supra* note 11 (discussing “living” cement).

Licenses, as Lexmark was happy to remind the nation, are still an available means of enforcement.³⁰¹ Exhaustion is fine as it is.

301. *See Lexmark*, 137 S. Ct. at 1538 (stating that “dealings between [] parties . . . can be addressed through contract law”).

CAPRICIOUS PATENT BOUNDARIES: THE “RE-LIBERALIZATION” OF POST-ISSUANCE PATENT CLAIM AMENDMENTS

Julien Crockett[†]

I. INTRODUCTION

The Patent Trial and Appeal Board’s (PTAB) recent “Preliminary Guidance”¹ in *Mylan Pharmaceuticals v. Sanofi-Aventis* on Sanofi’s motion to amend (MTA) is an important reminder that, unlike physical objects, intellectual property deals with abstract objects that have no clear boundaries. French global pharmaceutical company Sanofi sued German pharmaceutical company Mylan in October 2017 for infringement of its insulin pen patents.² Mylan subsequently petitioned the PTAB for *inter partes* review (IPR),³ a trial-like⁴ proceeding created in 2011 by the America Invents Act (AIA)⁵ for

DOI: <https://doi.org/10.15779/Z38W08WH48>

© 2020 Julien Crockett.

† J.D., University of California, Berkeley, School of Law, 2020. The author thanks Talha Syed for his encouragement and guidance as well as Kenneth Bamberger, Peter Menell, and Wayne Stacy for their helpful comments.

1. Mylan Pharm. Inc. v. Sanofi-Aventis Deutschland GmbH, No. IPR2018-01680, Paper No. 65 (P.T.A.B. Oct. 16, 2019). This was the first “Preliminary Guidance” ever issued. *PTAB Gives Sanofi First-Ever Early Feedback on Amendments*, LAW360 (Oct. 18, 2019), <https://www.law360.com/articles/1210934/ptab-gives-sanofi-first-ever-early-feedback-on-amendments>.

2. Complaint for Patent Infringement, *SanofiAventis U.S. LLC v. Mylan GmbH*, No. 2:17-cv-09105-SRC-CLW (D.N.J. Oct. 24, 2017), <https://storage.courtlistener.com/recap/gov.uscourts.njd.357534.1.0.pdf>.

3. *Mylan Pharm. Inc. v. Sanofi-Aventis Deutschland GmbH*, No. IPR2018-01682, Paper No. 2 (P.T.A.B. Sept. 10, 2018).

4. “Trial-like” because it adopts characteristics from district court litigation (such as limited discovery, expert evidence, cross-examination, oral argument, and rules of evidence and procedure) and characteristics from examinational proceedings (such as technical judges and the possibility of claim amendments). This “hybrid” nature, however, is controversial. *See Oil States Energy Servs., LLC v. Greene’s Energy Grp., LLC*, 138 S. Ct. 1365 (2018) (holding that IPRs are constitutional); *Saint Regis Mohawk Tribe v. Mylan Pharm.*, 896 F.3d 1322 (Fed. Cir. 2018) (holding that tribal sovereign immunity does not apply to IPR proceedings); *Regents of the Univ. of Minn. v. LSI Corp.*, 2019 WL 2479596 (Fed. Cir. June 14, 2019) (holding that state sovereign immunity is similar to Native American tribal sovereign immunity and does not apply to IPRs challenging state-owned patents).

5. The AIA was the first comprehensive patent bill to be enacted since the Patent Act of 1952 (“Patent Act”) and was intended to “establish a more efficient and streamlined patent system that w[ould] improve patent quality and limit unnecessary and counterproductive

revisiting the validity of a patent's claims. Sanofi then moved to amend some of its patent claims and requested preliminary guidance on its motion.⁶

This dance presents one of the oddest aspects of patent law. Claims that are meant to act as a patent's boundaries—to inform “the public during the life of the patent of the limits of the monopoly asserted”⁷—are shape-shifting entities. Seeking to amend patent claims both pre- and post-issuance is commonplace.⁸ And policy makers are only making it easier. In seeking preliminary guidance on its motion, Sanofi took advantage of a new strategy for patent owners offered by the PTO's Motion to Amend Pilot Program (“Pilot”). As of March 2019, patent owners may request preliminary guidance from the PTAB on their motions to amend and, based on this guidance and any opposition by the petitioner, revise their motions.⁹ The Pilot is not, however, alone in easing patent owners' abilities to amend their claims. It is part of a larger movement within the patent system that is adjusting the lever of claim amendments in defining the balance between patent owners and the public.

Claim amendments during examination¹⁰ and post-issuance have been an important part of the patent system since at least 1836¹¹—and for good reason. Patents have two primary parts: (1) a written description, or specification, that discloses the invention, and (2) claims that define the precise scope of the invention.¹² After a patent is filed—giving a patent its priority date—the

litigation costs.” 157 Cong. Rec. S1361 (daily ed. Mar. 8, 2011) (statement of Sen. Leahy), https://www.uspto.gov/sites/default/files/aia_implementation/20110308-debate_s23_part-b_s1360-s1394.pdf.

6. *Mylan Pharm. Inc. v. Sanofi-Aventis Deutschland GmbH*, IPR2018-01682, Paper No. 32 (P.T.A.B. June 25, 2019).

7. *Permutit Co. v. Graver Corp.*, 284 U.S. 52, 60 (1931).

8. Amendments “are a common practice in prosecution of patent applications.” *Hughes Aircraft Co. v. U.S.*, 717 F.2d 1351, 1363 (Fed. Cir. 1983). As described below, the PTO has also historically liberally granted post-issuance claim amendments. Greg Reilly, *Amending Patent Claims*, 32 HARV. J.L. & TECH. 1, 16 (2018) (“On first glance, the AIA extends the traditional liberal approach to claim amendments into these new post-issuance invalidity proceedings.”).

9. Notice Regarding a New Pilot Program Concerning Motion to Amend Practice and Procedures in Trial Proceedings Under the America Invents Act Before the Patent Trial and Appeal Board, 84 Fed. Reg. 9497 (Mar. 15, 2019) [hereinafter Notice Regarding a New Pilot Program] (providing (1) the opportunity to receive preliminary guidance on their motions to amend, and (2) the option to revise amended claims in light of a petitioner's opposition and the preliminary guidance).

10. Examination, or patent prosecution, is the process by which the PTO determines whether an inventor's application satisfies the various statutory requirements for the issuance of a patent.

11. Reilly, *supra* note 8, at 2 (“Patent claim amendments have been a largely uncontroversial part of the patent system since 1836.”).

12. JANICE M. MUELLER, PATENT LAW 11 (5th ed. 2016).

written description may not be amended to add any new matter.¹³ This requirement reflects an important policy decision that prevents inventors from filing before they have completed their inventions. However, claims are often extensively modified during examination.¹⁴ Justifications for this permissiveness include that claims do not add matter beyond what is described in the written description, and that inventors should be able to obtain patent protection even if the initial claims are rejected. Yet, even during examination, too liberal of a claim amendment policy can be problematic for the patent system. It creates perverse incentives for inventors, for example, to draft overbroad or vague claims and fails to incentivize them to avoid drafting mistakes.¹⁵

Post-issuance amendments are even harder to justify. While they may similarly protect a patent owner's investment while enabling claim correction, post-issuance amendments burden the PTO with more examination and increase uncertainty for competitors, follow-on innovators, and courts in parallel proceedings.¹⁶ In designing post-issuance amendment avenues, actors within the patent system have struggled to balance exclusive rights that incentivize invention against public enjoyment and use of those inventions. Too rigid of a claim amendment practice limits the universe of patents, incentivizing overly narrow claims. Too liberal of a practice fails to give the public notice regarding what is claimed.¹⁷ The Pilot is only the most recent

13. *Id.* at 190–92.

14. Those modifications generally taking the form of (1) narrowing amendments that distinguish claims from prior art or make the claims commensurate with the disclosure in the specification; (2) broadening amendments, provided they are supported by the specification's disclosure; or (3) clarifying amendments that neither narrow nor broaden claim language but more clearly specify the invention, distinguish it from prior art, or better track the disclosure in the specification. Reilly, *supra* note 8, at 10–11.

15. See Tun-Jen Chiang, *Fixing Patent Boundaries*, 108 MICH. L. REV. 523 (2010) (arguing that examination amendments create undesirable incentives for patent owners to draft overly broad and vague terms).

16. See *id.* (arguing that post-issuance amendments permit ever changing patent boundaries, resulting in uncertainty for both patent owners and competitors). As will be discussed below, limitations on post-issuance claim amendments vary with the method used (e.g., “intervening rights” are more strenuously applied to reissuance and reexamination than continuation applications).

17. The doctrine of intervening rights applies when post-issuance proceedings change the scope of a patent. “This doctrine recognizes that third parties may rely on the claims of an issued patent and thus provides a safe harbor to parties practicing subject matter covered by the amended claims.” PETER S. MENELL, LYNN H. PASAHOW, JAMES POOLEY, MATTHEW D. POWERS, STEVEN C. CARLSON, JEFFREY G. HOMRIG, GEORGE F. PAPPAS, CAROLYN CHANG, COLETTE REINER MAYER & MARC DAVID PETERS, PATENT CASE MANAGEMENT JUDICIAL GUIDE 14–22 (3rd ed. 2016), <https://www.law.berkeley.edu/wp-content/uploads/2016/05/Chapter-8-Final.pdf>. There are two types of intervening rights: (1) absolute intervening rights

design choice—promulgated by an increasingly assertive PTO¹⁸—and is a reaction to the most recent avenues for amending claims post-issuance: AIA trials.¹⁹

AIA trials, and IPRs in particular, were intended to serve as “relatively low-cost and prompt alternative[s]” —or even “complete substitute[s]” —to the validity portion of district court litigation.²⁰ In furtherance of this goal, the PTO restricted patent owners’ ability to amend claims during AIA trials.²¹ It did so by tinkering with the burden of persuasion for claim amendment proceedings. With other post-issuance procedures, the PTO or petitioner bear the burden of persuasion to prove the unpatentability of claim amendments.²² However, with AIA trials, the PTO reversed course and placed the burden on patent owners to prove the validity of their proposed amended claims.²³ Patent owners had to show patentability over all prior art known to the patent owner, and not just over the references applied by the petitioner against the original

that protect the right of an accused infringer to continue using, selling, or offering to sell infringing products made before the date of the amended claim in reissue or IPR, and (2) equitable intervening rights, which are forward-looking, and may permit the infringer to continue using products made after the amended date. *Id.*

18. See Clarisa Long, *The PTO and the Market for Influence in Patent Law*, 157 U. PA. L. REV. 1965, 1965–66 (2009) (discussing the PTO’s growing role as a player in the market for legal influence over patent law).

19. Figure 1 illustrates the relationship between the six avenues for amending claims post-issuance.

20. See Joel Sayres & Julie Wahlstrand, *To Stay or Not to Stay Pending IPR? That Should be a Simpler Question*, 17 CHI.-KENT J. INTELL. PROP. 52, 63 (2018) (citing the statutory framework and associated legislative history, including 157 Cong. Rec. S1376 (daily ed. Mar. 8, 2011) (statement of Sen. Kyl) (“Ideally, extending could-have-raised estoppel to privies will help ensure that if an inter partes review is instituted while litigation is pending, that review will completely substitute for at least the patents-and-printed publications portion of the civil litigation.”); 157 Cong. Rec. S5409 (daily ed. Sept. 8, 2011) (statement of Sen. Schumer) (stating that the AIA “streamlines review of patents to ensure that the poor-quality patents can be weeded out through administrative review rather than costly litigation.”); H.R. REP. NO. 112-98, pt. 1, at 48 (2011) (stating the purpose of IPRs is to “provid[e] quick and cost effective alternatives to litigation.”)).

21. See Supplemental Brief on Rehearing En Banc for Intervenor: Director of the USPTO at 28, *In re Aqua Prods., Inc.*, 833 F.3d 1335 (Fed. Cir. 2016) [hereinafter Supplemental Brief on Rehearing] (arguing that Congress did not want to “simply carry forward” the process for addressing amendments in inter partes reexamination to the “new proceedings”).

22. See, e.g., inter partes reexamination, the proceeding that the AIA trials replaced.

23. See Supplemental Brief on Rehearing, *supra* note 21, at 28 (arguing that Congress did not want to “simply carry forward” the process for addressing amendments in inter partes reexamination to the “new proceedings”); Reilly, *supra* note 8, at 3 (“[A]fter the America Invents Act of 2011 (AIA) created new post-issuance proceedings to review and cancel issued patents that fail the requisite statutory criteria, the Patent Office made it virtually impossible to amend claims in these proceedings.”).

patent claims. The resulting ninety-two-percent denial rate for claim amendments—a marked shift from the liberal grant rate in inter partes reexamination, the proceeding the AIA trials replaced²⁴—drew protest from the patent bar.²⁵ And beginning in 2017, three developments have worked to “re-liberalize” post-issuance claim amendment practice at the PTAB.

First, the Federal Circuit ruled in *Aqua Products v. Matal* that the patent owner does not bear the burden of persuasion with respect to demonstrating the patentability of requested amended claims.²⁶ This change effectively reversed what had been the PTO’s practice since 2012. Second, in October 2018, the PTO changed the standard by which claims are construed in post-issuance proceedings, including for both “unexpired patent claims and substitute claims proposed in a motion to amend,” from the “broadest reasonable interpretation”—the standard used by PTO examiners to determine the validity of a patent claim—to the more narrow *Phillips* standard, or “plain and ordinary meaning” standard, used in federal courts.²⁷ Because of this change, issued patents now retain a presumption of validity both in court and at the PTAB. And third, in March 2019, the PTO launched the above-mentioned Pilot, which gives patent owners new flexibility when amending claims, including both the opportunity to receive preliminary guidance on motions to amend, as well as the opportunity to revise amended claims in light of a petitioner’s opposition and the preliminary guidance.²⁸

This Note argues that the “re-liberalization” of post-issuance claim amendments during AIA trials hurts the patent system, particularly for patents involved in concurrent district court litigation. It undermines the AIA’s goal of mitigating burdensome litigation, creates perverse drafting and strategic incentives for patent owners, and is redundant with other post-issuance claim

24. See Reilly, *supra* note 8, at 15 (“In fact, claims are amended in approximately two-thirds of reexaminations, a significantly more common outcome than either complete confirmation or cancellation of claims.”).

25. See e.g., Neal Solomon, *The Problem of Inter-Partes Review (IPR)*, IPWATCHDOG (Aug. 8, 2017), <https://www.ipwatchdog.com/2017/08/08/problem-inter-partes-review-ipr/id=86287/> (“The combination of these features suggest that due process rights are not available to patent holders in the PTO compared to the district courts.”).

26. *Aqua Prods., Inc. v. Matal*, 872 F.3d 1290 (Fed. Cir. 2017).

27. Changes to the Claim Construction Standard for Interpreting Claims in Trial Proceedings Before the Patent Trial and Appeal Board, 83 Fed. Reg. 51,340, 51,340–41 (Oct. 11, 2018) (“This final rule revises the rules for IPR, PGR, and CBM proceedings that implemented provisions of the Leahy-Smith America Invents Act (“AIA”) providing for trials before the Office, by replacing the BRI standard for interpreting unexpired patent claims and substitute claims proposed in a motion to amend with the same claim construction standard that would be used to construe the claim in a civil action under 35 U.S.C. 282(b).”).

28. Notice Regarding a New Pilot Program, *supra* note 9.

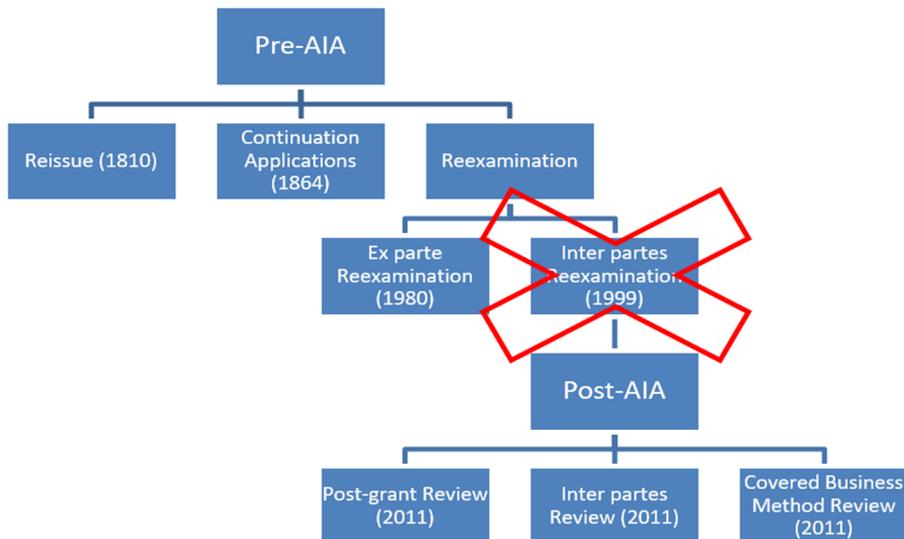
amendment avenues including reissue. This Note further illustrates these ill effects empirically. Not only have rates of claim amendment motions filed and granted in AIA trials increased since October 2017, but there has also been a decrease in the rate of petitions for IPR and an increase in district court stays pending IPR, demonstrating that IPRs are losing their ability to mitigate burdensome litigation.

Part II begins with a history of claim amendments at the PTO, including the AIA's restriction on claim amendments during AIA trials. Part III documents recent moves toward re-liberalizing post-issuance claim amendments in AIA trials, focusing on the Federal Circuit's *Aqua Products* decision, the PTO's embrace of the *Phillips* standard, and the PTO's Motion to Amend Pilot Program. Part IV analyzes the costs and benefits of a liberal approach to post-issuance claim amendments in AIA trials, concluding that the costs outweigh the benefits. And Part V moves beyond theory by examining the rate of claim amendments filed and granted during AIA trials since 2013, the rate of petitions for IPR since 2015, and the rate of district court stays granted pending IPR since 2013, empirically supporting the theoretical analysis in Part IV. This Note concludes by offering solutions to counteract the ill effects of re-liberalizing post-issuance claim amendments in AIA trials.

II. HISTORY OF POST-ISSUANCE CLAIM AMENDMENTS

This Part discusses the history of post-issuance claim amendments, detailing the six approaches for amending claims post-issuance displayed in Figure 1. It begins by discussing the origins of the PTO and the three pre-AIA methods—reissuance, continuation applications, and reexamination. It concludes by discussing post-AIA methods. In discussing each method's development, it explores the tension at the heart of post-issuance claim amendments: balancing incentives for inventors and costs for society.

Figure 1: Historical Development of Post-Issuance Proceedings



A. PRE-AIA

1. *Origins of the PTO*

The purpose of the patent system is “[t]o promote the progress of science and useful arts.”²⁹ After a brief start judging patent validity in 1790, the patent system became a pure registration system.³⁰ Patent validity became the prerogative of district courts, and only when patent owners tried to enforce their patents.³¹ As will be discussed below with reexamination and AIA trials, the close relationship between the PTO and district courts remains to this day.

29. U.S. CONST. art. I, § 8, cl. 8.

30. OREN BRACHA, *OWNING IDEAS: THE INTELLECTUAL ORIGINS OF AMERICAN INTELLECTUAL PROPERTY 1790-1909*, at 201 (Cambridge Univ. Press 2016) (“The 1790 regime was short lived. There were some complaints from patentees, but, as Jefferson observed years later, the main reason for its demise was that the board members were overwhelmed by the work load.”).

31. The Patent Office, created in 1802, was the first independent agency in the United States. Due to limited availability of federal institutions, the initial 1790 Patent Act—which invested the newly created Patent Board with discretionary power to weigh social costs and benefits underlying each patent grant—quickly gave way to the 1793 Patent Act’s registration system. *See* BRACHA, *id.* at 202 (“In the absence of meaningful prior review, the task of deciding the validity of patents now fell to the courts and was postponed entirely to the stage after the patent was issued.”). That district courts can determine validity is a distinctive U.S. practice, still in play today. In most countries, validity, even post-patent issuance, continues to be the province of the national patent office, with courts typically limited to patent enforcement issues (infringement, remedies). Robert P. Merges, *The Hamiltonian Origins of the U.S. Patent*

As could be expected under a pure registration system, patent quality was a serious issue with the 1793 Patent Act. The 1836 Patent Act shifted the issuance framework to an examination system.³² It was a step towards setting standards for universality and uniformity. Patent owners now had a right to receive a patent upon meeting patentability requirements. Along with this right came the ability to appeal decisions by the patent office and to amend one's patent claims both pre- and post-issuance.³³

As discussed above, patents have two primary parts: (1) a written description, or specification, that discloses the invention, and (2) claims that define the precise scope of the invention for which protection is sought. After a patent is filed, the written description may not be amended to add any new matter. Claims, however, are often extensively modified during examination. The patent system has struggled with designing avenues for post-issuance claim amendments that balance flexible claim adjustments for patent owner protection with public protection from unfair adjustments. The three pre-AIA avenues for amending claim amendments post-issuance are reissue, continuation applications, and reexamination.

2. *Reissue*

Reissue is only available to correct errors in a patent arising without “deceptive intention” where, as a result of the error, the patent is deemed

System, and Why They Matter Today, 104 IOWA L. REV. 2559, 2566 (2019) (“The resulting system, whereby federal judges determine patent validity in the context of a defense to patent infringement, was and is a distinctive aspect of U.S. patent practice.”).

32. BRACHA, *supra* note 30, at 209 (“The 1836 Patent Act was legislated after a long period of dissatisfaction with the existing framework. Complaints about the registration system varied, but most of them revolved around the insecurity of registered patents and the vagueness surrounding them. Issuing patents on demand with no prior examination resulted in their validity being highly questionable. Patentees often discovered their patents were invalid or that other patents overlapping with theirs had been issued. Purchasers of patented goods were harassed with dubious patents or paid licensing fees only to discover later that the goods they bought were covered by other conflicting patents.”). Unlike the semi-political board created by the 1790 Patent Act, here, the bureaucratic examiners had a clear prerogative: certifying the satisfaction of standard patentability criteria. *Id.* (“The act specifically established the Patent Office as a subdivision of the State Department, defined its structure, and established a corps of examiners. . . . [T]he newly organized Patent Office was meant to be nothing like the old Patent Board. It was set up not as a semipolitical forum with discretionary powers to grant privileges, but rather as a bureaucracy whose role was to certify the satisfaction of standard patentability criteria.”).

33. *Id.*

“wholly or partly inoperative or invalid.”³⁴ It proceeds much the same way as initial examination.

Without express statutory authority, the PTO has been granting reissued patents since 1810.³⁵ The Supreme Court endorsed this practice a few years before the statutory grant in *Grant v. Raymond*.³⁶ In *Grant*, the defendants raised an issue central to the balance between patent owners and the public: intervening rights.³⁷ What should courts do when a third party who was lawfully making, using, or selling a given article before a reissue now finds himself to be infringing under a modified patent? The Court in *Grant* declined to take on the topic and it wasn’t until *Sontag Chain Stores v. National Nut Co. of California* in 1940 that the Supreme Court endorsed the idea that a defendant could enjoy a personal “intervening right” to continue otherwise infringing activity after a valid reissue.³⁸

In response to concerns that reissue proceedings were used to expand claim scope to capture competitor’s products, the Supreme Court introduced limits on “broadening reissues” throughout the nineteenth century, including a requirement of diligence.³⁹ However, despite limiting broadening reissues to a two-year window after the grant of the original patent with the 1952 Patent Act,⁴⁰ actors within the patent system have heavily favored patent owners’ ability to amend their claims post issuance through reissue. For example, unlike the Acts of 1836 and 1870, which understood reissued patents as “new

34. 35 U.S.C. § 251. The error requirement is satisfied by a patent owner’s failure to previously present narrower claims. *See, e.g., In re Tanaka*, 640 F.3d 1246, 1251 (Fed. Cir. 2011).

35. While codified in the Patent Act of July 3, 1832, Pub. L. No. 22-162, § 3, without express statutory authority, the PTO began granting replacement amended patents, or reissued patents, in 1810, and the Supreme Court endorsed this practice in *Grant v. Raymond*, 31 U.S. 218, 243 (1832). MUELLER, *supra* note 12, at 569; Reilly, *supra* note 8, at 12.

36. *Grant v. Raymond*, 31 U.S. at 243.

37. *Id.* (“An objection much relied on is, that after the invention has been brought into general use, those skilled in the art or science with which it is connected, perceiving the variance between the specification and the machine, and availing themselves of it, may have constructed, sold and used the machine without infringing the legal rights of the patentee, or incurring the penalties of the law. The new patent would retro-act on them, and expose them to penalties to which they were not liable when the act was committed.”).

38. *Sontag Chain Stores Co. v. Nat’l Nut Co. of Cal.*, 310 U.S. 281 (1940). There currently are two types of intervening rights stemming from 35 U.S.C. § 252: (1) absolute intervening rights, whereby a party is shielded from liability if the products were made or used before the patent reissues, and (2) equitable intervening rights, which are a matter of judicial discretion and can shield a party from liability even for accused products made or used after the patent reissues.

39. 4A DONALD S. CHISUM, CHISUM ON PATENTS: A TREATISE ON THE LAW OF PATENTABILITY, VALIDITY AND INFRINGEMENT § 15.02[5] (LEXIS 2019).

40. 35 U.S.C. § 314(a) (2011).

patents” that were only effective as to “causes thereafter arising,”⁴¹ the 1928 Patent Act allowed patentees to recover “in so far as the claims of the original and reissued patents are identical.”⁴² In addition, courts have continued to interpret reissue based on the fundamental notions of equity and fairness to inventors recognized in the Supreme Court’s 1892 *Topliff* decision⁴³ and have construed the reissue statute liberally when they find no deceptive intent on the part of the inventor.⁴⁴ The Federal Circuit took this track recently in *In re Staats*, when it allowed a continuing reissue application to broaden claims after § 251’s two-year limit, finding that the two-year time limit only applied to the initial broadening reissue application.⁴⁵

3. Continuation Applications

The second method by which patent owners can alter their claims was developed in the mid-nineteenth century: continuation applications. Continuation applications are very popular and today constitute about one-

41. Therefore, the patentee could not recover for acts of infringement occurring before reissue.

42. 4A CHISUM, *supra* note 39, at § 15.02[5].

43. In finding that the patent owner could retain the broadened reissued patent, the Supreme Court in *Topliff v. Topliff*, 145 U.S. 156, 171 (1892), discussed the importance of post-issuance claim amendments:

The specification and claims of a patent, particularly if the invention be at all complicated, constitute one of the most difficult legal instruments to draw with accuracy, and in view of the fact that valuable inventions are often placed in the hands of inexperienced persons to prepare such specifications and claims, it is no matter of surprise that the latter frequently fail to describe with requisite certainty the exact invention of the patentee, and err either in claiming that which the patentee had not in fact invented, or in omitting some element which was a valuable or essential part of his actual invention. Under such circumstances, it would be manifestly unjust to deny him the benefit of a reissue to secure to him his actual invention, provided it is evident that there has been a mistake and he has been guilty of no want of reasonable diligence in discovering it, and no third persons have in the meantime acquired the right to manufacture or sell what he had failed to claim. The object of the patent law is to secure to inventors a monopoly of what they have actually invented or discovered, and it ought not to be defeated by a too strict and technical adherence to the letter of the statute, or by the application of artificial rules of interpretation.

44. *See, e.g., In re Weiler*, 790 F.2d 1576, 1579 (Fed. Cir. 1986) (noting that the reissue statute “should be construed liberally”).

45. *In re Staats*, No. 2010-1443 (Fed. Cir. Mar. 5, 2012).

quarter of all issued patents.⁴⁶ Continuation applications allow an inventor to rely, in a subsequent application, on the filing date of a prior co-pending application disclosing the same invention.⁴⁷ Although they didn't gain express statutory recognition until 1952, continuation applications were first recognized by the Supreme Court in *Godfrey v. Eames* in 1864, which found that an inventor's second application was entitled to the filing date of a first application because the second was analogous to an amendment and together they could be constituted as "one continuous application."⁴⁸

Throughout the nineteenth century and into the twentieth century, case law clarified the meaning of a "continuous application." In *Crown Cork & Seal v. Ferdinand Gutmann*, the Supreme Court emphasized the importance of continuity of disclosure rather than continuity of claims.⁴⁹ The lower courts further explained that the standard of continuity of disclosure was, in effect, equated with the enablement requirement.^{50, 51}

Many of these decisions, however, were not unanimous and the varying opinions demonstrate the unease some had (and continue to have)⁵² with continuation applications. The dissent in *Harder v. Haywood*, for example,

46. Mark A. Lemley & Kimberly A. Moore, *Ending Abuse of Patent Continuations*, 84 B.U.L. REV. 63, 66 (2004).

47. Menell, *supra* note 17, at 14–16 ("A continuation application is a second application for an invention claimed in a prior application. To qualify as a continuation application and claim the benefits of the earlier "parent" application's priority date, the application must be filed while the parent is still pending (i.e., not issued or abandoned), expressly refer to the parent application, identify at least one common inventor, and encompass the same disclosure of the parent application without adding any new matter. The same invention must be claimed, but the scope of the claims can vary." (internal citations omitted)).

48. *Godfrey v. Eames*, 68 U.S. (1 Wall.) 317, 325–26 (1864).

49. *Crown Cork & Seal Co. v. Ferdinand Gutmann Co.*, 304 U.S. 159, 165 (1938) ("[A]s Warth's disclosure was continuously before the Patent Office, clearly without any adverse significance to the petitioner is the fact that Warth formally canceled one disclosure from his first application and with it claims thought by the Circuit Court of Appeals broad enough to cover the disclosure. The continuity so maintained shows that Warth intended to retain, not to abandon, the disclosed invention.").

50. 4A CHISUM, *supra* 39, at § 13.02 (citing *Victor Talking Mach. Co. v. Am. Graphophone*, 145 F. 350, 351 (2d Cir. 1906) (holding that claims in a continuation application were entitled to the benefit of the filing date of the first (parent) application if the disclosures of the parent "were full enough to warrant the making of the claims"); *In re Febrey*, 135 F.2d 751 (C.C.P.A. 1943) (holding that the right to file a continuation application was created by *Godfrey v. Eames* and is dependent only upon copendency and continuity of disclosure)).

51. The enablement requirement requires that the specification "describe the invention in such terms that one skilled in the art can make and use the claimed invention." 2164 *The Enablement Requirement* [R-11.2013], USPTO, <https://www.uspto.gov/web/offices/pac/mpep/s2164.html> (last revised June 2020).

52. Lemley & Moore, *supra* note 46.

disapproved of an applicant keeping “his application alive indefinitely.”⁵³ The dissent also flagged the inappropriate incentives continuation applications create, in particular regarding continuation-in-part applications after allowance.^{54, 55} The dissent asserted that whatever purpose continuations-in-part have would be better solved by “proper attention” during the drafting of the patent.⁵⁶

The Patent Act of 1952 later codified the *Godfrey* doctrine entitling continuation applications to the filing date of prior applications.⁵⁷ Under § 120, an applicant can file a continuation application at any time before the PTO actually issues the patent or before the applicant abandons the application.

Here, again, courts interpret continuing applications in a manner favorable to patent owners. In *Kingsdown Medical Consultants v. Hollister*, for example, the Federal Circuit explicitly endorsed the practice of keeping an application pending in the PTO just to monitor developments in the marketplace and subsequently add claims during the continuation process to cover ideas learned from a competitor.⁵⁸ This practice appears particularly appealing in the pharmaceutical and biotechnology industries.⁵⁹ While Congress and courts

53. *Harder v. Haywood*, 150 F.2d 256, 265 (C.C.P.A. 1945) (Jackson, J., dissenting) (“The formal allowance of an application by the Patent Office is a holding by it that the patent which is granted on it is valid. Consequently, a continuation-in-part is as much of a device for delaying the issuance of a patent as is a continuation, and in my opinion both are subject to the same rule. Proper attention given to the drafting of specifications and claims in an application by reasonably diligent and skillful counsel would certainly reveal to him before allowance the scope of the necessary protection.”).

54. *Id.*

55. Unlike continuation applications, continuation-in-part applications also add additional matter. MUELLER, *supra* note 12, at 75.

56. *Haywood*, 150 F.2d at 265 (“The formal allowance of an application by the Patent Office is a holding by it that the patent which is granted on it is valid. Consequently, a continuation-in-part is as much of a device for delaying the issuance of a patent as is a continuation, and in my opinion both are subject to the same rule. Proper attention given to the drafting of specifications and claims in an application by reasonably diligent and skillful counsel would certainly reveal to him before allowance the scope of the necessary protection.”).

57. 35 U.S.C. § 120.

58. *Kingsdown Med. Consultants v. Hollister*, 863 F.2d 867, 874 (Fed. Cir. 1988) (“It should be made clear at the outset of the present discussion that there is nothing improper, illegal or inequitable in filing a patent application for the purpose of obtaining a right to exclude a known competitor's product from the market; nor is it in any manner improper to amend or insert claims intended to cover a competitor's product the applicant's attorney has learned about during the prosecution of a patent application. Any such amendment or insertion must comply with all statutes and regulations, of course, but, if it does, its genesis in the marketplace is simply irrelevant and cannot of itself evidence deceitful intent.”).

59. Lemley & Moore, *supra* note 46, at 71.

have created a number of patent doctrines designed to combat the misuse of continuation applications—including changing the term of patents, ending the secrecy of most patent applications, reviving the doctrine of written description, and creating a defense of prosecution laches—some commentators have raised concerns with this practice and propose further solutions to stem abuse, including banning the practice altogether.⁶⁰

4. Reexamination

Congress created the third approach for amending patent claims post-issuance in 1980: reexamination.⁶¹ Established to “strengthen[] investor confidence in the certainty of patent rights by establishing a system of administrative reexamination of doubtful patents,” reexamination was envisioned as a substitute for patent litigation.⁶² As the legislative history explains, “[r]eexamination will permit efficient resolution of questions about validity of issued patents without recourse to expensive and lengthy infringement litigation.”⁶³

In reexamination proceedings, the PTO assesses the patentability of the claims of an issued patent in view of prior art patents and publications.⁶⁴ Prior to the AIA’s enactment, there were two types of reexamination procedures: *ex parte* reexamination (still used today) and *inter partes* reexamination (replaced under the AIA with post-grant review, inter partes review, and covered business method proceedings). Both types of reexamination require a threshold showing that a substantial new question of patentability exists for one or more claims.⁶⁵

Ex parte reexamination may be brought by the patent owner, a third party, or the Director of the PTO.⁶⁶ If initiated by a third party, the third party has

60. *Id.* at 66 (offering solutions to stem abuse of continuation applications, including requiring publication of all applications, putting a time limit on the addition of new claims that broaden the scope of the patent, and creating a defense for infringers who independently developed the patented inventions before it was added to the patent claims).

61. Act of Dec. 12, 1980, Pub. L. No. 96-517, 94 Stat. 3015 (codified as amended in scattered sections of 35 U.S.C.).

62. H.R. REP. NO. 96-1307, pt. 1, at 4 (1980). Much scholarship has dissected the concerns that led patent policy makers to reallocate decision making authority away from the courts and into the PTO, including capitalizing on the institutional competence of the expert agency over the relative inexpertise of courts. *See, e.g.,* Saurabh Vishnubhakat, *The Youngest Patent Validity Proceeding: Evaluating Post-Grant Review*, 24 TEX. INTELL. PROP. L.J. 333, 345 (2016).

63. H.R. REP. NO. 96-1307, pt. 1, at 4.

64. 35 U.S.C. § 304.

65. 35 U.S.C. §§ 302–03.

66. 35 U.S.C. § 301.

limited involvement and the proceedings are conducted in a manner similar to examination.⁶⁷

Ex parte reexamination, however, was criticized as biased in favor of patent owners because of the extremely limited opportunities for third-party participation, which led to the creation of inter partes reexamination in 1999. However, inter partes reexamination was ineffective because the proceedings were often hampered by lengthy delays, severe estoppel provisions, a lack of decisive results, and a permissiveness for claim amendments.⁶⁸ In addition, involvement by third parties was limited to responding to the patent owner's responses to the PTO as part of the reexamination.

As noted above, the PTO took and continues to take a liberal stance on claim amendments during reexamination. In fact, claims are amended in approximately two-thirds of reexaminations.⁶⁹ However, unlike reissue, reexamination cannot be used to broaden the claims of an issued patent.⁷⁰

B. THE AIA

As the first comprehensive patent bill to be enacted since the Patent Act of 1952 (hereinafter the "Patent Act"), the AIA intends to "establish a more efficient and streamlined patent system that w[ould] improve patent quality and limit unnecessary and counterproductive litigation costs."⁷¹ Responding to mounting pressures regarding global patent harmonization, patent over-granting, and aggressive nonpracticing entities (i.e., patent trolls⁷²), the AIA enacted two major administrative changes to the patent system: shifting to a

67. 35 U.S.C. § 305.

68. See Brian J. Love & Shawn Ambwani, *Inter Partes Review: An Early Look at the Numbers*, 81 U. CHI. L. REV. DIALOGUE 93 (2014); Teresa Stanek Rea, *Building a Better Post Grant*, USPTO (May 15, 2012), https://www.uspto.gov/blog/director/entry/building_a_better_post_grant.

69. Reilly, *supra* note 8, at 15.

70. 35 U.S.C. § 305 ("no proposed amended or new claim enlarging the scope of a claim of the patent will be permitted in a reexamination proceeding under this chapter.").

71. See 157 CONG. REC. S1380 (daily ed. Mar. 8, 2011) at S1361 (statement of Sen. Leahy).

72. See Ted Sichelman, *Commercializing Patents*, 62 STAN. L. REV. 341, 368 (2010) (noting that non-practicing entities are pejoratively referred to as "trolls" because they "tend to exploit litigation and licensing market defects to extract unwarranted rents from commercializers, usually on patents that the commercializer was completely unaware of before the NPE's demand for payment"). But see FED. TRADE COMM'N, PATENT ASSERTION ENTITY ACTIVITY: AN FTC STUDY (2016), https://www.ftc.gov/system/files/documents/reports/patent-assertion-entity-activity-ftc-study/p131203_patent_assertion_entity_activity_an_ftc_study_0.pdf (arguing that a label like "patent troll" is "unhelpful because it invites pre-judgment about the societal impact of patent assertion activity without an understanding of the underlying business model that fuels such activity").

modified first-to-file system⁷³ and expanding post-issuance review proceedings.⁷⁴

Building on its goal of creating an inexpensive alternative to court challenges, Congress replaced inter partes reexamination with a three-tiered framework: broad post-grant review (PGR) in a patent's infancy, followed by narrower inter partes review (IPR) thereafter, with a limited exception for broad review of older covered business method patents (CBM). Unlike inter partes reexamination, these new proceedings take place as adjudicatory proceedings before three-member panels of administrative judges (ALJs)⁷⁵ from the PTAB.⁷⁶ They are thus "hybrid,"⁷⁷ adopting characteristics from district court litigation (such as limited discovery, expert evidence, cross-examination, oral argument, and rules of evidence and procedure) and characteristics from examinational proceedings (such as technical judges and the possibility of claim amendments). However, under Title 35, the PTO does not have the authority to decide infringement or to grant relief for violation of a patentee's rights.⁷⁸ Such disputes are decided in civil actions filed in federal district court.

73. "Modified" because the AIA regime still permits an inventor a one-year grace period under certain conditions. Because of this, the U.S. regime still diverges from other first-to-file systems.

74. See Changes to Implement Inter Partes Review Proceedings, Post-Grant Review Proceedings, and Transitional Program for Covered Business Method Patents, 77 Fed. Reg. 48,680 (Aug. 14, 2012) (codified at 37 C.F.R. §§ 42.100 *et seq.*). See also Jeffrey P. Kushan, *The Fruits of the Convoluted Road to Patent Reform: The New Invalidity Proceedings for the Patent and Trademark Office*, 30 YALE L. & POL'Y REV. 385 (2012) (arguing that, given the excessive cost of litigation, administrative proceedings before the PTO play a crucial role in providing a way to contest patents).

75. 35 U.S.C. § 6. The ALJs also hear appeals from adverse examiner decisions in patent applications and reexamination proceedings, and render decisions in interferences. *Id.* Appointment of APJs is a controversial topic: the Federal Circuit recently held that appointment of APJs by the Secretary of Commerce violates the Appointments Clause. See *Federal Circuit Says PTAB Judges Are Not Constitutionally Appointed*, IPWATCHDOG (Oct. 31, 2019), <https://www.ipwatchdog.com/2019/10/31/federal-circuit-says-ptab-judges-not-constitutionally-appointed/id=115556/>; Gene Quinn, *USPTO Admits to Stacking PTAB Panels to Achieve Desired Outcomes*, IPWATCHDOG (Aug. 23, 2017), <https://www.ipwatchdog.com/2017/08/23/uspto-admits-stacking-ptab-panels-achieve-desired-outcomes/id=87206/>.

76. Steve Brachmann, *Supreme Court Petition Challenges PTAB's Constitutionality Under the Takings Clause*, IPWATCHDOG (Aug. 20, 2018), <https://www.ipwatchdog.com/2018/08/20/supreme-court-petition-challenges-ptabs-constitutionality-takings-clause/id=100401/>.

77. See *supra* note 4.

78. While Title 35 does not give the PTO the authority to decide infringement, in *Oil States Energy Services, LLC v. Greene's Energy Group, LLC*, the Court determined that patents are a government franchise rather than a vested property right. See 138 S. Ct. 1365, 1375. Invoking the "public rights doctrine," the Court found that patent validity trials need not take place in Article III courts, nor do they violate the Seventh Amendment right to a jury trial. *Id.* at 1379.

This Section addresses (1) the three new post-issuance proceedings, (2) their purpose, and (3) the PTO's approach to claim amendments before the "re-liberalization" of post-issuance claim amendments.

1. *AIA trials*

As discussed above, with these new procedures, the AIA created a three-tiered framework with broad PGR in a patent's infancy, followed by narrower IPR thereafter, with a limited exception for broad review of older CBM patents. In contrast with pre-AIA inter partes reexamination, PGR, IPR, and CBM review take place as adjudicatory proceedings before three-member panels of ALJs from the PTAB.⁷⁹ The parties can conduct limited discovery, respond to each other's arguments, and have the right to an oral hearing.⁸⁰

a) Post-Grant Review

Considered "the first window" for challenging patent validity, PGR may be sought within the first nine months after a patent is issued by any person who is not the patent owner.⁸¹ By statute, PGR should be completed by the PTO within twelve months from the date of institution, although this can be extended an additional six months for good cause, or as needed in the case of joinder.⁸² According to a Judiciary Committee Report, "[t]he intent of the post-grant review process is to enable early challenges to patents, while still protecting the rights of inventors and patent owners against new patent challenges unbounded in time and scope."⁸³ Thus, the third party standing requirements are broad, as is the scope of challenges—encompassing all

The Court's holding was narrow but left open the possibility of future constitutional challenges, including regarding the interplay between infringement and validity, and whether the public rights doctrine applies to patent infringement.

79. 35 U.S.C. § 6(c).

80. 35 U.S.C. § 316(a)(5), (8), (10), (13).

81. 35 U.S.C. § 321(a)–(c). Because of this timing restraint, PGR applies only to patents subject to the AIA's modified first-to-file priority rule, enacted eighteen months from the AIA's enactment (i.e., only to patents from applications filed after March 16, 2013). Pub. L. No. 112-29, § 3(n)(1), 125 Stat. 284, 293 (2011). In large part because of this structural constraint, during the first four years in which the procedure became available, the PTO received only twenty-one petitions for PGR, in contrast with over four thousand for IPR and four hundred for CBM. See Vishnubhakat, *supra* note 62, at 354. As of September 30, 2018, the filing statistics are as follows: 8,476 IPRs, 560 CBMs, and 134 PGRs. USPTO, TRIAL STATISTICS: IPR, PGR, CBM (2018), https://www.uspto.gov/sites/default/files/documents/trial_statistics_20180930a.pdf. For a further analysis as to why IPRs are more popular than PGRs, see Jonathan Stroud, *Patent Post-Grant Review After Alice*, 69 BAYLOR L. REV. 56 (2017).

82. See 35 U.S.C. § 326(a)(11).

83. H.R. REP. NO. 112-98, at 58 (2011).

grounds that could be asserted to render a patent invalid or unenforceable⁸⁴—and the permitted evidence, which includes expert opinion.⁸⁵

b) Inter Partes Review

To avoid overlap with PGR, IPR can only be requested by a third party after the nine-month window following the issue date of the patent or after termination of PGR.⁸⁶ As with PGR, only third parties can petition for IPR and they “must demonstrate that there is a reasonable likelihood that the petitioner would prevail with respect to at least one of the claims challenged in the petition.”⁸⁷ In *SAS Institute, Inc. v. Iancu*, the Supreme Court ended the PTAB’s practice of “partial institution,” holding that if the PTAB institutes review, it must issue a final written decision addressing every patent claim challenged in an IPR petition.⁸⁸ Unlike PGR, the grounds for IPR are limited to lack of novelty⁸⁹ and nonobviousness,⁹⁰ and the grounds for challenges are limited to patents and printed publications.⁹¹ If the PTAB grants review, a final determination must be issued “not later than one year” after the petition is granted.⁹² The one-year period may be extended for good cause for up to six months, although “[e]xtensions of the one-year period are anticipated to be rare.”⁹³

c) Covered Business Method Patents

While PGR and IPR are the primary new forms for challenging issued patents before the PTAB, CBM was added as a transitional program set to expire on September 16, 2020,⁹⁴ to address the “flood of poor quality business

84. Whereas IPR limits the grounds by which claims can be challenged to just novelty and obviousness over patents and printed publications, PGR permits challenges by those too, as well as by written description (except for best mode), enablement, indefiniteness, claim broadening on any prior art statutorily available, and subject matter eligibility. 35 U.S.C. § 321(b).

85. 35 U.S.C. §§ 321(a), 322(a)(3).

86. 35 U.S.C. § 311(c).

87. 35 U.S.C. § 314(a).

88. *SAS Inst., Inc. v. Iancu*, 138 S. Ct. 1348 (2018).

89. 35 U.S.C. § 102.

90. 35 U.S.C. § 103.

91. 35 U.S.C. § 311 (2013).

92. 35 U.S.C. § 316(a)(11) (2014).

93. 77 Fed. Reg. 48,680, 48,695 (Aug. 14, 2012). Extensions of the one-year period have, in fact, been rare, but may become more frequently invoked in light of *SAS Institute, Inc. v. Iancu*’s holding that the PTAB must issue a final written decision addressing every patent claim challenged in an IPR petition, ending the PTAB’s practice of “partial institution.”

94. On March 20, 2018, Congress held a hearing to determine the future of CBM and whether it should be extended beyond its current September 16, 2020, sunset date (Congress has considered whether to extend CBM several times, including in June 2015). *See* Gene Quinn

method patents” unleashed by the Federal Circuit’s decision in *State St. Bank & Trust Co. v. Signature Financial Group* in 1998.⁹⁵ State Street categorically did away with the business method exception to patentability, allowing patents over business methods for the first time.⁹⁶ Because of the lack of prior art references and scholarship around methods of conducting business, patent examiners had little guidance for weeding out undeserving applications.⁹⁷ Following State Street, courts quickly recognized the growing problem of vague and suspect business method patents and began to articulate new standards for obviousness,⁹⁸ but a body of “problematic” CBM patents were already issued. To help “reduce the burden placed on courts and the economy by this back-and-forth shift in judicial precedent,” Congress created a “temporary administrative alternative for reviewing business method patents,”⁹⁹ within the AIA. As with PGR and IPR, the transition program “is designed to provide a cheaper, faster alternative to district court litigation,”¹⁰⁰ but with an even greater emphasis on not burdening courts.¹⁰¹

& Steve Brachmann, *The House IP Subcommittee: A Bunch of Fiddling Neros Watching the U.S. Patent System Burn* (Mar. 29, 2018), IP Watchdog, <https://www.ipwatchdog.com/2018/03/29/house-ip-subcommittee-fiddling-neros-watching-patent-system-burn/id=95263/>. The March 2018 hearing took place a week after the U.S. Government Accountability Office issued an assessment of CBM that highlighted the general decline of CBM filings since 2015. *Id.* This decline makes sense given that the program was created, as chair of the House Intellectual Property Subcommittee Representative Durrell Issa stated in his introduction, to address “a series of patents that got out that the court today clearly has said that the PTO should not grant.” *Id.* CBM remains slated to sunset after September 16, 2020. *Id.*

95. See *State St. Bank & Trust Co. v. Signature Fin. Grp.*, 149 F.3d 1368 (1998); H.R. REP. NO. S23, pt. S1363 (2011).

96. H.R. REP. NO. S23, pt. S1363.

97. *Id.*

98. See, e.g., *Bilski v. Kappos* 561 U.S. 593 (2010) (acknowledging the need for a limiting principle on business method patents and signaling that abstract business methods are not patentable); *eBay Inc. v. MercExchange, L.L.C.* 547 U.S. 388, 397 (2006) (Kennedy, J., concurring) (citing the “potential vagueness and suspect validity” of business method patents).

99. H.R. REP. NO. S23, pt. S1363.

100. *Id.*

101. Some commentators and courts have argued that because the “underlying impetus and rationale of CBMs largely mirror those of IPRs, there is little reason to apply a different standard for the two proceedings with respect to stays pending litigation.” Sayres & Wahlstrand, *supra* note 20, at 64–65 (citing *Murata Mach. USA v. Daifuku Co.*, 830 F.3d 1357, 1362 (Fed. Cir. 2016) (concluding that district courts may properly consider the fourth CBM factor in considering motions to stay pending IPR because the “legislative history confirms that Congress’s desire to enhance the role of the PTO and limit the burden of litigation on courts and parties was not limited to the CBM review context”) (internal quotation marks omitted); *NFC Tech. LLC v. HTC Am., Inc.*, No. 2:13-CV-1058-WCB, 2015 WL 1069111, at *5 (“The legislative history indicates that Congress recognized that the same underlying policy

2. *Goal of AIA Trials*

The AIA trials' overlap with civil litigation goes to the heart of the AIA: mitigating burdensome litigation. Although patent litigation takes place in federal district courts, it is common for parties to initiate parallel proceedings in the PTO to challenge the validity of an asserted patent or to confirm that a patent is valid in light of potential challenges.^{102, 103} Based on recent scholarship, an estimated seventy percent of instituted IPRs are brought by parties involved in district court litigation.¹⁰⁴ Congress believed that IPR, in particular, could serve as a "complete substitute" for district court litigation with respect to validity based on patents and printed publications under 35 U.S.C. § 102 (novelty) and § 103 (nonobviousness).^{105, 106}

considerations that apply to CBM review apply to inter partes review as well."); Vishnubhakat, *supra* note 62, at 64 ("In general, the AIA's legislative history indicates Congress wanted both IPRs and CBM reviews to serve as a substitute for Article III litigation over patent validity.").

102. Although historically, the PTO's job has focused on examining patent applications, Congress established and expanded the PTO's authority to reexamine previously issued patents over the past forty years. The AIA's post-issuance administrative procedures were instantly popular with patent challengers for at least three reasons: (1) they are generally much less expensive than district court patent infringement litigation; (2) they can be invoked by anyone wishing to challenge a patent, not just those who have a litigation-worthy "case or controversy" under federal court procedure; and (3) issued patents in these proceedings do not enjoy the "presumption of validity" that district courts must afford during the adjudication of challenges to patent validity. *See* ROBERT PATRICK MERGES & JOHN FITZGERALD DUFFY, *PATENT LAW AND POLICY: CASES AND MATERIALS* 18 (7th ed. 2017).

103. Five main venues are available to resolve disputes involving U.S. patents, beginning with U.S. District Courts, the U.S. International Trade Commission, and the PTO. Regardless of the first venue selected, all cases can then be appealed to the U.S. Court of Appeals for the Federal Circuit, and subsequently to the U.S. Supreme Court. It is worth noting that the "substantial evidence" standard on appeal from the PTO is more deferential than the "clearly erroneous" standard for appeals from district courts. *See Dickinson v. Zurko*, 527 U.S. 150 (1999).

104. Saurabh Vishnubhakat, Arti K. Rai & Jay P. Kesan, *Strategic Decision Making in Dual PTAB and District Court Proceedings*, 31 *BERKELEY TECH. L.J.* 45, 49 (2016).

105. *See* Sayres & Wahlstrand, *supra* note 20, at 52, and accompanying parenthetical.

106. The PTO took an additional step in this direction by replacing its previous broadest reasonable interpretation standard on October 11, 2018, with the same *Phillips* standard for claim construction as district courts. *See* Changes to the Claim Construction Standard for Interpreting Claims in Trial Proceedings Before the Patent Trial and Appeal Board, 83 Fed. Reg. 51,340 (Oct. 11, 2018). Its stated ambition was that "the change will lead, among other things, to greater consistency and harmonization with the federal courts and the ITC and lead to greater certainty and predictability in the patent system." *PTAB Issues Claim Construction Final Rule*, USPTO (Oct. 10, 2018), <https://www.uspto.gov/patents-application-process/patent-trial-and-appeal-board/procedures/ptab-issues-claim-construction>.

However, while AIA trials, and IPRs in particular, have proven to be popular, they are also controversial.¹⁰⁷ Commentators have criticized IPR's anti-inventor bias,¹⁰⁸ and legislation was introduced that would abolish not only IPR, but also the PTAB.¹⁰⁹ Further, courts have considered IPR's constitutionality¹¹⁰ and overlap with civil litigation.¹¹¹ A less reported but equally important development has been the AIA's unanticipated effect on claim amendments.¹¹²

3. *Tightening Post-Issuance Motions to Amend*

Many assumed that claim amendment procedures wouldn't change with the AIA.¹¹³ As noted above, AIA trials were intended to replace inter partes reexamination, which liberally allowed claim amendments. In addition, the AIA states that "the patent owner may file one motion to amend the patent."¹¹⁴

107. See, e.g., Andre Iancu, Director, USPTO, Keynote Address at the U.S. Chamber of Commerce Patent Policy Conference: Role of U.S. Patent Policy in Domestic Innovation and Potential Impacts on Investment (Apr. 11, 2018), <https://www.uspto.gov/about-us/news-updates/remarks-director-andrei-iancu-us-chamber-commerce-patent-policy-conference> ("Pointing to the high invalidation rates in IPR proceedings, some hate the new system with vigor, arguing that it's an unfair process that tilts too much in favor of the petitioner. Others love the system, and think it's the best tool we have to correct errors, eliminate 'bad patents,' and improve patent quality.").

108. See, e.g., Russell Slifer, *Five Years After the AIA Created the PTAB*, IPWATCHDOG (Sept. 16, 2017), <http://www.ipwatchdog.com/2017/09/16/five-years-after-the-aia-created-the-ptab/id=87994/> (arguing that post-issuance proceedings have substituted defendant abuse in district court litigation with patent owner abuse in the PTAB).

109. H.R. 6264, 115th Cong. (2018) (proposing to abolish the PTAB and eliminate both IPRs and PGRs because they "have harmed the progress of science and the useful arts by subjecting inventors to serial challenges to patents").

110. See *Oil States Energy Servs., LLC v. Greene's Energy Grp., LLC*, 138 S. Ct. 1365 (2018) (holding that IPRs are constitutional); see, e.g., *Regents of the Univ. of Minn. IPR2017-01186*, Paper No. 14 (P.T.A.B. Dec. 19, 2017) (considering whether state sovereign immunity applies to IPR proceedings). See Shashank Upadhye & Adam Sussman, *A Real Separation of Powers or Separation of Law: Can an Article I Administrative Agency Nullify an Article III Federal Court Judgment?*, 25 FORDHAM INTELL. PROP., MEDIA, & ENT L.J. 1 (2014).

111. See, e.g., *Wi-Fi One, LLC v. Broadcom Corp.*, 887 F.3d 1329 (Fed. Cir. 2018) (holding that decisions by the PTAB to institute IPR validity challenges may be appealed when such decisions relate to the timing of the petitions, rather than on the substantive merits).

112. Barbara Clarke McCurdy, Thomas L. Irving & Stacy D. Lewis, *How the PTAB and Federal Circuit Have Responded to Aqua Products*, FINNEGAN (Mar. 28, 2018), <https://www.finnegan.com/en/insights/how-the-ptab-and-federal-circuit-have-responded-to-aqua-products.html> ("The patent bar watched the en banc rehearing decision in *Aqua Products Inc. v. Matal*, with great interest. This was because the lack of availability to amend claims in an inter partes review has frustrated patent owners and was a somewhat unanticipated consequence of the America Invents Act." (internal citations omitted)).

113. *Id.*

114. 35 U.S.C.A. § 316(d)(1).

However, the PTO set new regulations that restricted post-issuance claim amendments in AIA trials, namely by putting the burden of patentability on the patent owner.¹¹⁵ As the PTO stated, it understood Congress as not wanting to “carry forward to the new proceedings” the amendment process of inter partes reexamination.¹¹⁶ But why did it think that?

Before the AIA, Congress explicitly directed the PTO to follow the initial examination approach for reexamination.¹¹⁷ With the AIA, however, Congress removed that explicit instruction. Further, the PTO understood Congress to have “created a new practice in patent law—the motion to amend a patent apart from any sort of reexamination or reissue proceeding before the USPTO.”¹¹⁸ Thus, the PTO took Congress as not wanting to “carry forward” the amendment process from reexamination.

Instead, the PTO placed the burden of proving patentability of amended claims on the patent owner—the person in the best position to be familiar with the prior art.¹¹⁹ In doing so, however, the PTAB severely restricted motions to amend during post-issuance proceedings.¹²⁰ In fact, between 2012 and 2017, the PTAB denied an MTA in ninety-two percent of AIA trials.¹²¹ Beginning in 2017, however, several developments “re-liberalized” post-issuance claim amendments at the PTAB.

III. RELIBERALIZING CLAIM AMENDMENTS

Since 2017, three recent events have “re-liberalized” post-issuance claim amendments: (1) the Federal Circuit’s decision in *Aqua Products*, which clarifies where the burden of proving patentability lies; (2) the PTO’s embrace of the Phillips standard; and (3) the PTO’s Motion to Amend Pilot. They are discussed below in chronological order.

115. See 37 C.F.R. § 42.121.

116. Supplemental Brief on Rehearing, *supra* note 21, at 28. However, in the House Report for the AIA, several representatives noted with approval the high rate of “modification or nullification” of patent claims in inter partes reexamination and their desire to retain this feature in IPRs. H.R. REP. NO. 112-98, pt. 1, at 164 (2011).

117. 35 U.S.C. § 314 (pre-AIA) (providing that the initial examination procedures where amendments are treated like original claims and must be proven unpatentable by the Patent Office of § 132 and § 133 would apply during inter partes reexamination).

118. Supplemental Brief on Rehearing, *supra* note 21, at 14–17.

119. *Id.* at 28.

120. USPTO, PATENT TRIAL AND APPEAL BOARD MOTION TO AMEND STUDY (2018), https://www.uspto.gov/sites/default/files/documents/ptab_%20mta_study_%28installment_5_-_update_through_fy2018%29.pdf (showing that PTAB has denied a motion to amend in ninety percent of AIA trials).

121. *Id.*

A. AQUA PRODUCTS AND ITS PROGENY

The decision in *Aqua Products* was widely anticipated by the patent bar.¹²² Given the ninety-two-percent denial rate of claim amendments, many felt that the PTO had overstepped its role.¹²³

The product at issue in *Aqua Products v. Matal* was a jet-propelled pool cleaner. During litigation, Zodiac Pool Systems, Inc., (“Zodiac”) petitioned the PTO for IPR of Aqua Products’ (“Aqua”) patent.¹²⁴ The PTAB held all instituted claims unpatentable and denied Aqua’s MTA, concluding that Aqua had failed to prove the patentability of the substitute claims.¹²⁵ The PTAB highlighted its standard that the burden is on the patent owner to show that the proposed amendments are patentable over the known prior art.¹²⁶

Aqua appealed the decision to the Federal Circuit and argued that it did not bear the burden of proving the patentability of its proposed substitute claims.¹²⁷ The Federal Circuit in *In re Aqua Products* rejected Aqua’s argument, finding that the court “has upheld the Board’s approach of allocating the patentee the burden of showing that its proposed amendments would overcome the art of record.”¹²⁸ Aqua sought rehearing *en banc* of that panel decision.¹²⁹

In a fractured decision comprising five separate opinions, the Federal Circuit sided with Aqua. The Federal Circuit held that, under the PTO’s current rules and rule-making history, the PTO may not place the burden of persuasion regarding the patentability of proposed amended claims during an IPR on the patent owner.¹³⁰ A plurality of the court ruled that the burden of

122. McCurdy, *supra* note 112, and accompanying parenthetical. *See also* BOB STEINBERG, JONATHAN M. STRANG & MICHAEL J. GERARDI, *AQUA PRODUCTS LEVELS THE PLAYING FIELD AT THE PTAB*, LATHAM & WATKINS (2017), <https://www.lw.com/thoughtLeadership/LW-Aqua-Products-Levels-the-Playing-Field-at-the-PTAB> (“The Federal Circuit’s long-awaited *en banc* opinion . . .”).

123. *See, e.g.*, Gene Quinn, *Industry Reaction to the Federal Circuit’s Decision in Aqua Products v. Matal*, IPWATCHDOG (Oct. 4, 2017), <https://www.ipwatchdog.com/2017/10/04/industry-reaction-federal-circuits-decision-aqua-products-v-matal/id=88866/> (gathering reactions to the *Aqua Products* decision).

124. *Aqua Prods., Inc. v. Matal*, 872 F.3d 1290 (Fed. Cir. 2017).

125. *Id.* at 1297 (citing *Zodiac Pool Sys. Inc. v. Aqua Prods. Inc.*, No. IPR2013-159, 2014 WL 4244016 (P.T.A.B. Aug. 22, 2014)).

126. *Id.*

127. *Id.*

128. *Aqua Prods.*, 872 F.3d at 1297.

129. *Id.*

130. *Id.* at 1327 (“The only legal conclusions that support and define the judgment of the court are: the PTO has not adopted a rule placing the burden of persuasion with respect to the patentability of amended claims on the patent owner that is entitled to deference; and in

proof during IPR regarding the patentability of amended claims is on the petitioner.¹³¹ The court was divided, however, on how to interpret the AIA, particularly on the relationship between 35 U.S.C. § 316(e) and 35 U.S.C. § 316(d), and whether they “unambiguously” place the burden of proof on petitioners for both original and amended claims.¹³²

A majority of the court found the statute “ambiguous,” but in the absence of an interpretation of the statute by the PTO to which the court must defer under *Chevron*, it held that “the most reasonable reading of the AIA is one that places the burden of persuasion with respect to the patentability of amended claims on the petitioner.”¹³³ The PTO’s amendment practice—placing the burden of patentability on the patent owner—therefore violated the AIA.¹³⁴

The *Aqua Products* decision, however, left open the possibility that the PTO could promulgate a rule placing the burden on the patent owner or petitioner. Thus far, the PTO hasn’t done so. In guidance offered after *Aqua Products*, however, the PTO was hesitant to accept that the petitioner had the burden of persuasion.¹³⁵ In fact, the PTO’s understanding of the *Aqua* decision seemed markedly different from industry’s.¹³⁶ The PTO stated that while it would not place the burden of persuasion on the patent owner, “the Board will proceed to determine whether the substitute claims are unpatentable by a preponderance of the evidence based on the entirety of the record, including any opposition made by the petitioner.”¹³⁷ It concluded by noting that the only impact this change would make is if “the entirety of the evidence of record before the Board is in equipoise as to the unpatentability of one or more substitute claims.”¹³⁸ If that were the case, the Board would grant the MTA.

the absence of anything that might be entitled deference, the PTO may not place that burden on the patentee.”).

131. *Id.* at 1296.

132. *Id.*

133. *Id.*

134. *Id.*

135. Memorandum from David P. Ruschke on MTAs in View of *Aqua Products*, USPTO (Nov. 21, 2017), https://www.uspto.gov/sites/default/files/documents/guidance_on_motions_to_amend_11_2017.pdf (noting that the only impact this change would make is if “the entirety of the evidence of record before the Board is in equipoise as to the unpatentability of one or more substitute claims”).

136. See STEINBERG, *supra* note 123, at 1 (“Once the patent owner meets its burden of showing they are responding to the petitioners’ arguments and not broadening their claims, the burden will shift to the petitioner to introduce prior art (including new prior art) and to make arguments concerning the patentability of the amended claims. Aqua’s holding should enable patent owners to more easily amend claims during an IPR. The impact of this change on all stakeholders in the patent system will be significant.”).

137. Memorandum from David P. Ruschke, *supra* note 135, at 2.

138. *Id.*

The PTO later went a step further by designating its *Western Digital Corporation, v. Spex Technologies Inc.* order precedential and de-designating earlier orders on motions to amend in *MasterImage 3D Inc. v. RealD Inc.* and *Idle Free Systems Inc. v. Bergstrom Inc.*¹³⁹ In *Western Digital*, the PTAB declared that “the burden of persuasion will ordinarily lie with the petitioner to show that any proposed substitute claims are unpatentable by a preponderance of the evidence,” although “[t]he Board itself also may justify any finding of unpatentability by reference to evidence of record in the proceeding.”¹⁴⁰

B. PHILLIPS STANDARD

Until October 2018, the PTAB and district courts used different standards for claim construction in validity proceedings. The PTAB used the Broadest Reasonable Interpretation (BRI), which required consideration of the “broadest reasonable meaning of [a claim’s] words in their ordinary usage as they would be understood by one of ordinary skill in the art, taking into account whatever enlightenment by way of definitions or otherwise that may be afforded by the written description contained in the applicant’s specification.”¹⁴¹ This is the standard PTO examiners use to determine whether a patent claim should be issued in the first instance. By contrast, district courts use the *Phillips* approach, wherein claim terms must be given “the meaning that the term would have to a person of ordinary skill in the art in question at the time of the invention.”¹⁴²

BRI was controversial from the outset. It treated issued patent claims the same as unexamined patent claims, essentially ridding the system of the presumption of validity. In addition, different standards for interpreting claims for the purpose of invalidity can lead to gaming of the system.¹⁴³ However, the Supreme Court in *Cuozzo Speed Technologies v. Lee* authorized the PTO to adopt a rule promulgating use of BRI under the logic that patent owners may move

139. Rules of Practice to Allocate the Burden of Persuasion on Motions to Amend in Trial Proceedings Before the Patent Trial and Appeal Board, 84 Fed. Reg. 56,401 (Oct. 22, 2019).

140. The board has since replaced its *Western* decision with *Lectrosonics, Inc. v. Zaxcom, Inc.*, IPR2018-01129, -01130, Paper No. 15, slip op. at 3–4 (P.T.A.B. Feb. 25, 2019) (precedential). See also USPTO, TRIAL PRACTICE GUIDE UPDATE (2019), <https://www.uspto.gov/sites/default/files/documents/trial-practice-guide-update3.pdf>.

141. *In re Morris*, 127 F.3d 1048, 1054 (Fed. Cir. 1997).

142. *Phillips v. AWH*, 415 F.3d 1303, 1313 (Fed. Cir. 2005).

143. See, e.g., *PPC Broadband, Inc. v. Corning Optical Commc’ns RF, LLC*, 815 F.3d 734, 741 (Fed. Cir. 2016) (“This case hinges on the claim construction standard applied—a scenario likely to arise with frequency. And in this case, the claim construction standard is outcome determinative.”).

to amend during an AIA trial.¹⁴⁴ However, courts and the PTAB have used these different standards as an excuse to perform de novo claim construction.¹⁴⁵

Under the PTO's recent move adopting the *Phillips* standard for both "unexpired patent claims and substitute claims," however, the PTAB now takes into consideration any prior claim construction determination that is made in a civil action or a proceeding before the ITC. In seeking to harmonize standards, the PTO hopes that "the change will lead, among other things, to greater consistency and harmonization with the federal courts and the ITC and lead to greater certainty and predictability in the patent system."¹⁴⁶ Former Commissioner for Patents at the PTO, Robert Stroll, commended Director Iancu "for being responsive to the wide majority of parent practitioners who have raised concerns that the PTAB procedures were unfairly tilted against patent owners."¹⁴⁷

However, in eliminating the BRI, the PTO loses its rationale for allowing claim amendments. New and substitute claims that have never passed in front of an examiner under the BRI now never will—and ALJs are not well positioned to act as first review of new or amended claims.¹⁴⁸ In addition, while the PTO's stated purpose is improving consistency between the courts and PTAB, it is unlikely to do so given the complexities of *Phillips*. Further, other

144. *Cuozzo Speed Techs., LLC v. Lee*, 136 S. Ct. 2131, 2145 (2016) (explaining that the patent owner's opportunity to amend its patent in IPRs is what justifies the Board's use of the broadest reasonable interpretation standard in IPRs). *See also In re Rambus, Inc.*, 753 F.3d 1253, 1256 (Fed. Cir. 2014) (finding that, in inter partes reexamination, "the sole basis for the 'broadest reasonable interpretation' rubric is the ability to amend claims"); *In re Prater*, 415 F.2d 1393, 1404–05 (C.C.P.A. 1969) (holding that claims are given their broadest reasonable interpretation during examination "since the applicant may then amend his claims").

145. KEVIN GREENLEAF, SHAUN ZHANG, ERIC COHEN, JOAN ELLIS, DAVID RANDALL, JONATHAN KWOK, JAMES HIETALA, TED NISSLY & LISSA MOJICA, INTELL. PROP. OWNERS ASS'N, *HOW DIFFERENT ARE THE BROADEST REASONABLE INTERPRETATION AND PHILLIPS CLAIM CONSTRUCTION STANDARDS?*, <https://ipo.org/wp-content/uploads/2018/10/BRI-v-Phillips-Final-1.pdf>.

146. USPTO, *supra* note 106.

147. Gene Quinn, *USPTO Publishes Final Rule Adopting Phillips Standard at PTAB*, IPWATCHDOG (Oct. 10, 2018), <https://www.ipwatchdog.com/2018/10/10/uspto-publishes-final-rule-phillips-standard-ptab/id=102210/>.

148. Not only do many ALJs not have the same training as examiners, but AIA trials have statutorily mandated short timelines. *See Stacy Lewis & Tom Irvin, Amending Rather than Cancelling Claims in Inter Partes Review*, 11 BUFF. INTELL. PROP. L.J. 178 (2015). *See also* Gene Quinn, *PTAB Judges Shockingly Inexperienced Compared to District Court Judges*, IPWATCHDOG (Mar. 6, 2018), <https://www.ipwatchdog.com/2018/03/06/ptab-judges-shockingly-inexperienced/id=94438/> (questioning whether PTAB judges can competently fulfill their roles).

proceedings at the PTO such as ex parte reexamination will continue to use the BRI standard, leading to inconsistency within the PTO itself.

C. MOTION TO AMEND PILOT PROGRAM

The PTO's Motion to Amend Pilot Program launched in March 2019.¹⁴⁹ It provides a patent owner with two options not previously available during AIA trials. The first option is that a patent owner may choose to receive preliminary guidance from the PTAB on its MTA.¹⁵⁰ The second option is that a patent owner may choose to file a revised MTA after receiving petitioner's opposition to the original MTA and after receiving the PTAB's preliminary guidance (if requested).¹⁵¹ The Pilot also extends time between due dates (e.g., "rather than 1 month, a patent owner will have 6 weeks to file a reply after receiving an opposition to its original MTA, and a petitioner will have 6 weeks to file a sur-reply in response to that reply").¹⁵²

The Pilot's roots began in 2014, eighteen months after implementing the AIA. The PTO requested comments on all aspects of the new administrative trial provisions of the AIA and many comments focused on which party should bear the burden of proving the patentability or unpatentability of substitute claims proposed in an MTA.¹⁵³ In 2015, the PTAB specifically requested feedback on "[w]hat modifications, if any, should be made to the PTAB's practice regarding motions to amend," and again most comments focused on the burden of proof.¹⁵⁴ The PTAB undertook a Motion to Amend Study in 2016, analyzing (1) the number of MTAs that had been filed in AIA trials, both as a cumulative total and by fiscal year; (2) subsequent developments in each MTA (i.e., whether the motion was decided, rendered moot, withdrawn, or otherwise dismissed); (3) the number of MTAs requesting to substitute claims that were granted, granted-in-part, denied-in-part, and denied; and (4) the reasons the PTAB provided for denying entry of substitute claims.¹⁵⁵ In finding that ninety-two percent of motions to amend are denied, the PTO formulated the Pilot.¹⁵⁶

149. Notice Regarding a New Pilot Program, *supra* note 9.

150. *Id.*

151. *Id.*

152. *Id.*

153. *Id.*

154. *Id.*

155. *Id.*

156. USPTO, *supra* note 120.

D. CURRENT LAW ON MOTIONS TO AMEND & PARALLEL PROCEEDINGS

Together, these three events have rapidly liberalized post-issuance claim amendments in AIA trials. Aqua Products changed the burden from the patent owner to the petitioner. Shifting from BRI to the Phillips standard rid the PTO of its justification for permitting post-issuance claim amendments during AIA trials. And the Pilot has increased the chances of patent owners having their amendments granted. This section will address (1) the current law on motions to amend during post-issuance validity proceedings, and (2) will identify how parallel procedures may also be used.

a) Current Law on Motions to Amend

The AIA provides that a patent holder in an IPR “may file one motion to amend the patent,” either by cancelling any challenged patent claim or by “propos[ing] a reasonable number of substitute claims.”¹⁵⁷ The AIA delegates authority to the Director to “prescribe regulations . . . establishing and governing inter partes review” and to “set[] forth standards and procedures for allowing the patent owner to move to amend the patent” under § 316(d).¹⁵⁸ Invoking this authority, the Director promulgated 37 C.F.R. § 42.121, which permits a patent owner to file one motion to amend after conferring with the Board.¹⁵⁹ Under this regulation, the Board may deny a motion to amend if the amendment does not satisfy the requirements of § 316(d)(3) (i.e., if it expands the claim scope, introduces new matter,¹⁶⁰ or if it “does not respond to a ground of unpatentability involved in the trial”¹⁶¹). As the PTAB demonstrated recently in *Amazon.com, Inc. v. Uniloc*, while IPR is statutorily limited to prior art challenges based on patents and printed publications under § 102 and § 103, a substitute claim, however, must be reviewed for compliance with all of the statutory requirements for patentability.¹⁶²

The Director also set forth 37 C.F.R. § 42.20 to govern all motion practice before the PTAB. This section was at the heart of the *Aqua Products* case

157. 35 U.S.C. § 316(d)(1)(B). There is a presumption that a reasonable number of substitute claims is one substitute claim per challenged claim, 37 C.F.R. § 42.121(a)(3), but a patent owner may rebut this presumption.

158. 35 U.S.C. §§ 316(a)(4), (9).

159. 37 C.F.R. § 42.121(a)(1).

160. 35 U.S.C. § 316(d)(3); 37 C.F.R. § 42.121(a)(2)(ii).

161. 37 C.F.R. § 42.121(a)(2)(i).

162. *Amazon.com, Inc. v. Uniloc Luxembourg S.A.*, IPR2017-00948, Paper No. 34, at 5 (P.T.A.B. Jan. 18, 2019) (“By its terms, § 311(b) limits a petitioner to requesting cancellation of existing claims of a patent only under § 102 and § 103. It does not, however, limit the grounds of unpatentability that can be raised in response to proposed substitute amended claims presented in a motion to amend.” (internal citations omitted)).

discussed above.¹⁶³ In relevant part, Rule 42.20(a) requires that any “[r]elief, other than a petition requesting the institution of a trial, must be requested in the form of a motion.”¹⁶⁴ Rule 42.20(c) states additionally that “[t]he moving party has the burden of proof to establish that it is entitled to the requested relief.”¹⁶⁵ In *Aqua Products*, the PTO claimed that the Board “interpreted Rules 42.20 and 42.121 to place the burden of persuasion on a patent owner to demonstrate, by a preponderance of the evidence, that any proposed amended claims are patentable, that it must do so in light of prior art not already part of the IPR, and that the Director has endorsed that interpretation.”¹⁶⁶ This rule, however, directly conflicts with the penultimate section under § 316 titled “Evidentiary Standards,” which states “in an inter partes review instituted under this chapter, the petitioner shall have the burden of proving a proposition of unpatentability by a preponderance of the evidence.”¹⁶⁷ As discussed, in *Aqua Products* the Federal Circuit held that the PTO may not place the burden of persuasion, regarding the patentability of proposed amended claims during an IPR, on the patent owner.¹⁶⁸ Instead, the burden should generally be placed on the petitioner. The PTAB recently designated *Lectrosonics v. Zaxcom* as precedential and it clarifies this point, stating that:

as a result of the current state of the law and USPTO rules and guidance, the burden of persuasion ordinarily will lie with the petitioner to show that any proposed substitute claims are unpatentable by a preponderance of the evidence. The Board itself also may justify any finding of unpatentability by reference to evidence of record in the proceeding, for example, when a petitioner ceases to participate. . . . Thus, the Board determines whether substitute claims are unpatentable by a preponderance of the evidence based on the entirety of the record, including any opposition made by the petitioner.¹⁶⁹

After the PTAB decides whether to grant any amended claim, the PTAB “shall issue a final written decision with respect to the patentability of any patent claim challenged by the petitioner and any new claim added under section 316(d).”¹⁷⁰ The statute provides that, following the final written

163. See generally *Aqua Prods., Inc. v. Matal*, 872 F.3d 1290 (Fed. Cir. 2017).

164. 37 C.F.R. § 42.20.

165. *Id.*

166. *Aqua Prods.*, 872 F.3d at 1301.

167. 35 U.S.C. § 316(e).

168. *Aqua Prods.*, 872 F.3d at 1296.

169. *Lectrosonics, Inc. v. Zaxcom, Inc.*, No. IPR2018-01129, Paper No. 15 (P.T.A.B. Feb. 25, 2019).

170. 37 C.F.R. § 318(a).

decision and any subsequent appeal, the Director shall incorporate “in the patent . . . any new or amended claim determined to be patentable.”¹⁷¹

Unlike in the examination context, the PTAB proceedings must be completed within a year, and until the PTO’s 2019 Motion to Amend Pilot, there was no time for the back-and-forth between the applicant and the PTO that happens during prosecution. Under the Motion to Amend Pilot, patent owners now have two new opportunities: obtaining preliminary guidance from the Board and filing a revised motion to amend. A disappointed patent owner who believes the PTAB erred in refusing to allow a claim amendment may, of course, seek review from the Federal Circuit.

b) Parallel Proceedings at the PTO

In addition to amending claims during AIA trials, there are parallel proceedings for amending claims, including continuing applications, reissue, and ex parte reexamination. Given the low grant rate of claim amendments during AIA trials, the PTO provided notice to remind patent owners of these parallel proceedings and to describe the factors the PTO considers when determining whether to stay or suspend a parallel proceeding during the pendency of an AIA proceeding.¹⁷²

According to the PTO, eighty-nine percent of patents challenged in AIA proceedings have not had any associated reexams or reissues.¹⁷³ Indeed, less than one percent of the U.S. patents granted annually are reissue patents.¹⁷⁴ Continuation patents, on the other hand, constitute about one-quarter of all issued patents.¹⁷⁵

As others have summarized, those who seek reissues are often successful.¹⁷⁶ For example, over seventy percent of utility reissue patent applications and ninety percent of design reissue patent applications were granted for all applications issued or abandoned since 2014.¹⁷⁷ Further, for reissue patent applications pending in parallel to a post-issuance proceeding challenging claims of the original patent, the data indicates a grant rate of over

171. 37 C.F.R. § 318(b).

172. Notice Regarding Options for Amendments by Patent Owner Through Reissue or Reexamination During a Pending AIA Trial Proceeding, 84 Fed. Reg. 16,654 (Apr. 22, 2019).

173. *Id.*

174. *Reissue Presents A Solid Option For Amending Patent Claims*, LAW360 (Oct. 21, 2019), <https://www.law360.com/articles/1210276/reissue-presents-a-solid-option-for-amending-patent-claims>.

175. *Id.*

176. *See, e.g., id.*

177. *Id.*

fifty percent.¹⁷⁸ One explanation for these increased chances may be that a third party generally cannot actively participate in reissue proceedings. Further, reissues are examined by examiners, not ALJs.

IV. ANALYSIS ON WHETHER THE RELIBERALIZATION OF POST-ISSUANCE CLAIM AMENDMENTS DURING VALIDITY PROCEEDINGS IS BENEFICIAL FOR THE PATENT SYSTEM

As noted above, the U.S. patent system is justified through a utilitarian rationale of maximizing net social welfare—balancing exclusive rights that incentivize invention against public enjoyment and use of those inventions.¹⁷⁹ Claim amendments are an important lever in that balance. They provide inventors with the flexibility to narrow, broaden, clarify, and correct claims.¹⁸⁰ But post-issuance claim amendments also burden the patent system and create uncertainty for competitors, follow-on inventors, and courts in parallel proceedings. This Part weighs the pros and cons of post-issuance amendment procedures during validity proceedings.

A. PROS OF POST-ISSUANCE CLAIM AMENDMENTS

Post-issuance claim amendments during AIA trials serve a potentially cost saving corrective function. Congress noted with the AIA that “[a]llowing narrowing amendments during an IPR helps strengthen and clarify patents. [. . .] [P]roviding a patent owner with a meaningful opportunity to amend subject to minimal statutory and regulatory criteria helps preserve the merited benefits of patent claims better than the win-all or lose-all validity contests in district court.”¹⁸¹ The Supreme Court has also noted that claim drafting is difficult.¹⁸² A more permissive system encourages innovation because it ensures that patentees receive protection even if the original claim is struck down. Further, district court litigation may be stayed to allow the validity portion of trial to

178. *Id.*

179. William Fisher, *Theories of Intellectual Property*, in *NEW ESSAYS IN THE LEGAL AND POLITICAL THEORY OF PROPERTY* 168, 173–74 (Stephen Munzer ed., 2001).

180. Reilly, *supra* note 8, at 20 (“When an original claim covers both matter to which the inventor is entitled and matter to which they are not, narrowing claim amendments offer a middle ground that allows the inventor to receive only the protection that is warranted, rather than being denied any protection whatsoever.”).

181. 152 CONG. REC. 16,834 (2006) (statement of Sen. Leahy on S. 3818) (emphasis added).

182. *Topliff v. Topliff*, 145 U.S. 156, 171 (1892).

take place at the less expensive PTAB.¹⁸³ It is a waste to continue litigating claims in district court when the claims may change mid-proceeding at the PTO. As discussed below in Part V, stay rates have increased significantly over the last two years.¹⁸⁴

In addition, the ability to amend claims during AIA trials is limited. Patent owners cannot broaden claims and their amendment is via a “motion to substitute claims” with limited back-and-forth examination. Therefore, the danger of amending claims to, for instance, capture a competitor’s technology is more limited than in continuation applications. In addition, by making it more difficult to amend patents during AIA trials, patent owners may turn to other methods such as reissue or continuation applications.¹⁸⁵ Moreover, the PTAB invalidates a large percentage of claims during AIA trials, not just because of its past use of BRI, but also because of the lower burden of proof.¹⁸⁶

B. CONS OF POST-ISSUANCE CLAIM AMENDMENTS

While the merits of a patent system that is fair to patent owners is well understood, policy makers have swung the balance too far. Post-issuance claim amendments during AIA trials should be limited for three reasons: (1) they undermine the AIA’s goal of mitigating burdensome litigation; (2) they create perverse drafting and strategic incentives for patent owners; and (3) not only are they redundant with other post-issuance claim amendment avenues

183. Anne S. Layne-Farrar, *The Cost of Doubling Up: An Economic Assessment of Duplication in PTAB Proceedings and Patent Infringement Litigation*, ABA (May 1, 2018), https://www.americanbar.org/groups/intellectual_property_law/publications/landslide/2017-18/may-june/cost-doubling-up/ (“A 2015 American Intellectual Property Law Association (AIPLA) Economic Survey found that the median cost of an IPR through a PTAB hearing was \$275,000 and through appeal was \$350,000. In contrast, even with relatively low stakes (\$1–\$10 million at risk), patent infringement litigation typically costs nearly \$1 million through the end of discovery and \$2 million through final disposition.”).

184. See also Graham C. Phero & Lauren A. Watt, *Success of Motions to Stay Rising, but Why?*, STERNE, KESSLER, GOLDSTEIN & FOX (Feb. 2020), <https://www.sternekessler.com/news-insights/publications/success-motions-stay-rising-why> (attributing these trends to “the all-or-nothing institution approach required under the *SAS* decision and the PTAB’s adoption of the Phillips standard for claim construction”).

185. Some firms, for example, have included this in their guidance to the public and clients. McCurdy, *supra* note 112 (“Maintaining pending continuation applications has long been a recommended practice to protect the patent applicant’s invention while the scope of such protection is being determined.”).

186. An IPR petition only needs to prove unpatentability of challenged claims by a preponderance of the evidence, 35 U.S.C. § 316(e), whereas in a lawsuit, each patent claim is presumed valid, 35 U.S.C. § 282, and the accused infringer must prove invalidity by clear and convincing evidence.

such as reissue and continuations, but also ALJs are in worse positions than examiners to judge the validity of new or amended claims.

First, post-issuance claim amendments during AIA trials undermine the AIA's goal of mitigating burdensome litigation. Congress believed that AIA trials, and IPR in particular, could serve as a "relatively low-cost and prompt alternative"—or even a "complete substitute"—to some district court litigation.¹⁸⁷ However, the re-liberalization of post-issuance claim amendments complicates that goal, given the greater likelihood that amended claims will emerge.¹⁸⁸ Because patents post-issuance create public reliance, post-issuance amendments increase uncertainty for courts in parallel proceedings, competitors, and follow-on innovators. Further, while district court stays pending AIA trials may save costs, stays also delay court proceedings and, because judges inconsistently apply stays, lead to venue shopping.¹⁸⁹

Second, post-issuance claim amendments during AIA trials create perverse drafting and strategic incentives for patent owners. Knowing that if challenged their originally granted claims can be amended to the amount of protection it was entitled to in the first place, patent owners are rewarded for drafting vague and overbroad claims. Although potential infringers may be protected by intervening rights, often, patent claims are amended to cover a competitor's product, undermining competition and follow-on innovation. Knowing this, accused infringers have less incentive to challenge weak patents in post-issuance invalidity proceedings. Further, often missing in the conversation are the effects of claim amendments on the staff at the PTO. As has been shown with continuation applications, drawn out examination wears down

187. See Sayres & Wahlstrand, *supra* note 20, and accompanying parenthetical.

188. *Aqua Products amici* in favor of the PTO have further noted that Congress's goal of "improving patent quality"—an argument that some have argued would be stifled by tightening post-issuance claim amendments—is in fact bolstered through the "low success rate of motions to amend." It is evidence that IPR is operating as Congress intended: to target and cancel especially weak patent claims. En Banc Brief of the Internet Association, the Computer & Communications Industry Association, Dell, Garmin, Intel, Red Hat, Samsung, SAP America, SAS Institute, the Software & Information Industry Association, Symmetry, and Vizio as Amici Curiae in Support of Intervenor, *In re Aqua Prods., Inc.*, No. 2015-1177 (Fed. Cir. Nov. 3, 2016) at 4, 14.

189. See Sayres & Wahlstrand, *supra* note 20, at 53 ("These goals [of IPR substituting for district court validity proceedings], however, have not been fully realized, in part because the variation and inconsistency in applying the factors commonly applied to stay motions during the days of inter partes reexamination have remained."); *How Increased Stays Pending IPR May Affect Venue Choice*, LAW360 (Nov. 15, 2019), <https://www.law360.com/articles/1220066/how-increased-stays-pending-ipr-may-affect-venue-choice> ("The Supreme Court's recent decision in *TC Heartland LLC v. Kraft Foods Group Brand LLC* appears to have amplified SAS Institute's impact by redistributing cases to courts that are more likely to grant stays. These trends are likely to significantly influence litigants' choice of venue in patent cases.").

examiners.¹⁹⁰ Weighing the incentives of patentees and the PTO demonstrates that examiners are not given enough time with each patent and suffer from burnout.¹⁹¹

Third, post-issuance claim amendments during AIA trials are redundant with other post-issuance claim amendment avenues such as reissuance and continuations: if patent owners' claims are overbroad, they can narrow them via reissuance; if their claims are too narrow, they can broaden them by filing a continuation application or seeking reissuance (within two years of issuance). Further, as discussed above, unlike reexamination and reissue, AIA trials take place before ALJs instead of examiners who are better positioned to act as first review of new or amended claims.¹⁹²

In conclusion, while there are benefits to allowing amendments during AIA review, the costs outweigh the benefits, especially in light of the availability of other avenues for amending claims after issuance.

V. EMPIRICAL ANALYSIS

Not only is a liberal post-issuance claim amendment practice bad in theory, but as the data below demonstrates, the problems identified are coming to fruition. The liberalization of post-issuance claim amendments during validity proceedings hurts the patent system and runs contrary to the AIA's goal of improving efficiency of patent litigation and increasing quality of patents. Since October 2017 (*Aqua Products*), we can see that (1) the rate of claim amendments filed and granted during AIA trials has increased; (2) the rate of petitions for IPR has stagnated or decreased; and (3) the rate of stays granted pending IPR has increased. Given the incentives created by a liberal stance towards post-issuance claim amendments, these patterns are expected.

190. Lemley & Moore, *supra* note 46, at 11.

191. See Michael D. Frakes & Melissa F. Wasserman, *Irrational Ignorance at the Patent Office*, 72 VAND. L. REV. 975, 978 (2019) ("If examiners are not given enough time to evaluate applications, they may not be able to reject applications by identifying and articulating justifications with appropriate underlying legal validity. Offering validation for these concerns, recent reports commissioned by the federal government bemoan that examiners believe they are 'fighting for their lives' and are 'not [given] enough time to do a proper job.'). See also David M. Longo & Ryan P. O'Quinn, Note, *Checking the Stats: How Long Is Too Long to Give Adequate Public Notice in Broadening Reissue Patent Applications?*, 10 DUKE L. & TECH. REV. 1, 4 ("The current patent system provides little incentive for patentees or the USPTO to spend sufficient time and money to prosecute a high-quality patent; the result is a kudzu-like thicket of low-quality, under-descriptive patents."); Arti K. Rai, *Addressing the Patent Gold Rush: The Role of Deference to PTO Patent Denials*, 2 WASH. U.J.L. & POL'Y 199, 218 (2000) (arguing for a change in the current patent examiners' incentive system that encourages them to grant patents).

192. See *supra* note 75.

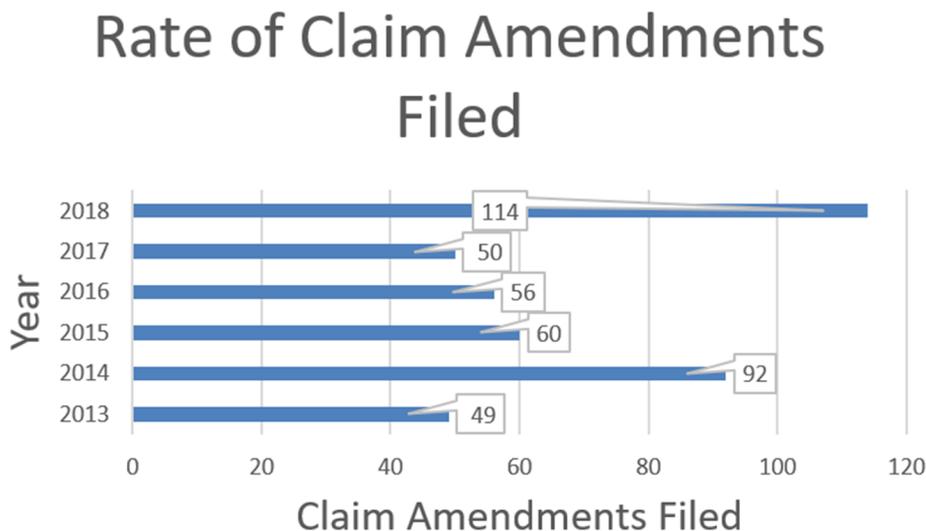
A. RATE OF CLAIM AMENDMENTS FILED AND GRANTED DURING IPR

With the liberalization of post-issuance claim amendments in AIA trials, beginning in October 2017, we can expect to see an increase in claim amendment filings and grant rates during AIA trials.

1. *Rate of Claim Amendments Filed*

Figure 2 demonstrates the increase in claim amendments filed in AIA trials since October 2017 by comparing claim amendment filings since 2013. This data was collected from Docket Navigator¹⁹³ and checked with the PTO's Motion to Amend Study.¹⁹⁴ Figure 2 illustrates that the number of motions to amend filed in 2018 (114) is greater than the number of motions to amend filed in any other year, and more than the number of motions to amend filed in 2017 (fifty) and 2016 (fifty-six) combined.

Figure 2: Rate of Claim Amendments Filed for years 2013–2018



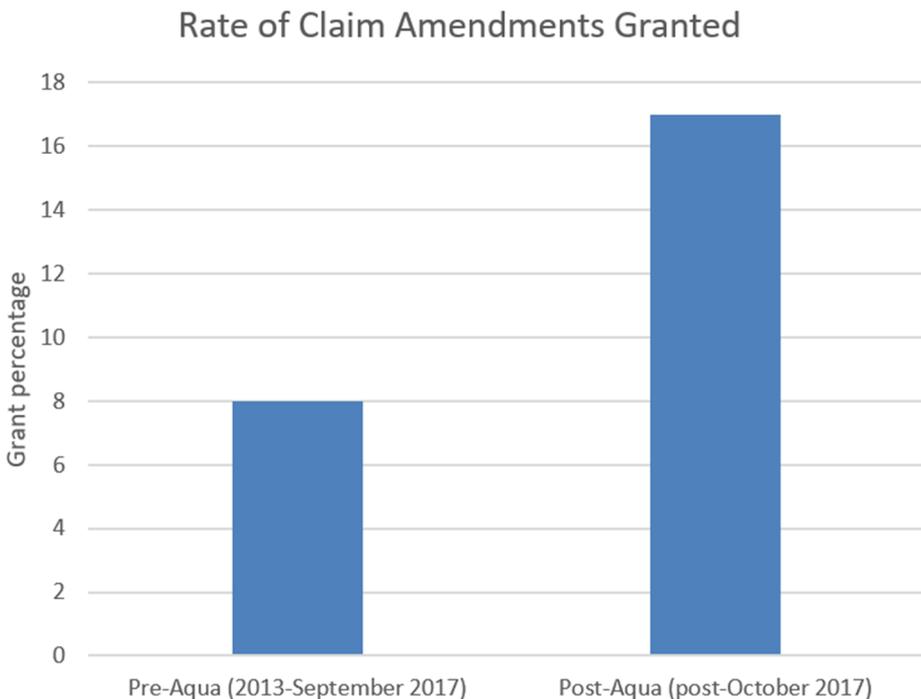
193. DOCKET NAVIGATOR, <https://www.docketnavigator.com/welcome/login>.

194. The data for 2019 was not yet available from the PTO at the time of writing this Note.

2. *Rate of Claim Amendments Granted*

Figure 3 compares the rate of claim amendments granted, or granted-in-part, both before and after October 2017.¹⁹⁵ Since October 2017, the PTAB has decided twenty-nine motions to amend and granted or granted in part five of these motions, resulting in a seventeen-percent grant rate, which is more than two times greater than the pre-Aqua grant rate of eight percent (granting, or granting-in-part, fourteen motions to amend out of 170 total motions). This is a small sample size and, currently, the Fischer Exact Hypothesis test¹⁹⁶ suggests that there is not yet enough evidence to reject the null hypothesis that the rate remains unchanged, but as more data is collected this may change. The author expects that the trend towards a higher grant rate will continue.

Figure 3: Rate of Claim Amendments Granted comparing Pre- and Post-Aqua Products



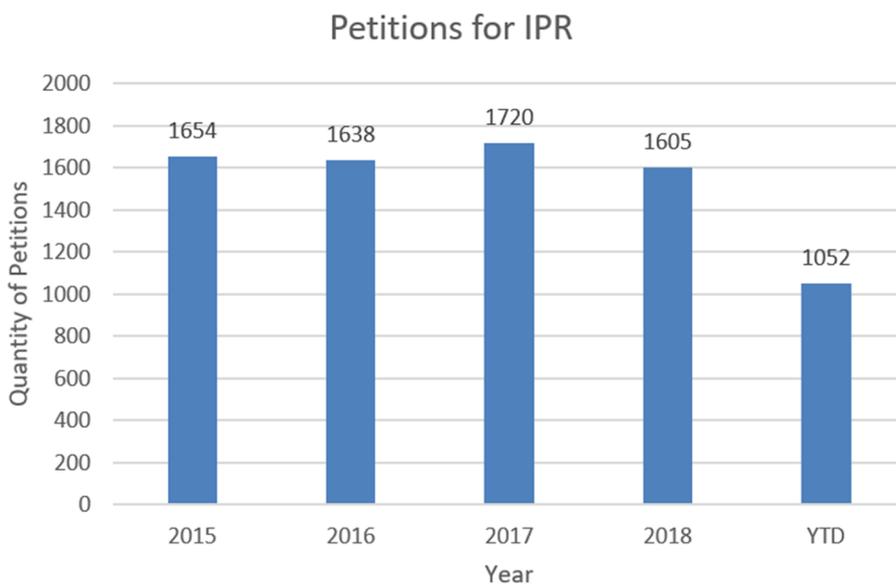
195. DOCKET NAVIGATOR, *supra* note 193.

196. The Fischer Exact Hypothesis test is a “statistical test used to determine if there are nonrandom associations between two categorical variables.” *Fisher’s Exact Test*, WOLFRAM MATHWORLD, <https://mathworld.wolfram.com/FishersExactTest.html> (last visited June 6, 2021).

B. RATE OF PETITIONS FOR IPR

With the liberalization of post-issuance claim amendments in AIA trials, we should expect to see a decrease in the rate of petitions for IPR since October 2017, because they have become less favorable for defendants. Patent owners are now more likely to retain their patent (albeit in amended form). Figures 4 and 5 demonstrate this decrease by comparing petitions both yearly for IPR, since 2015, and monthly.¹⁹⁷ This data was collected from the PTO's database on trial statistics.¹⁹⁸

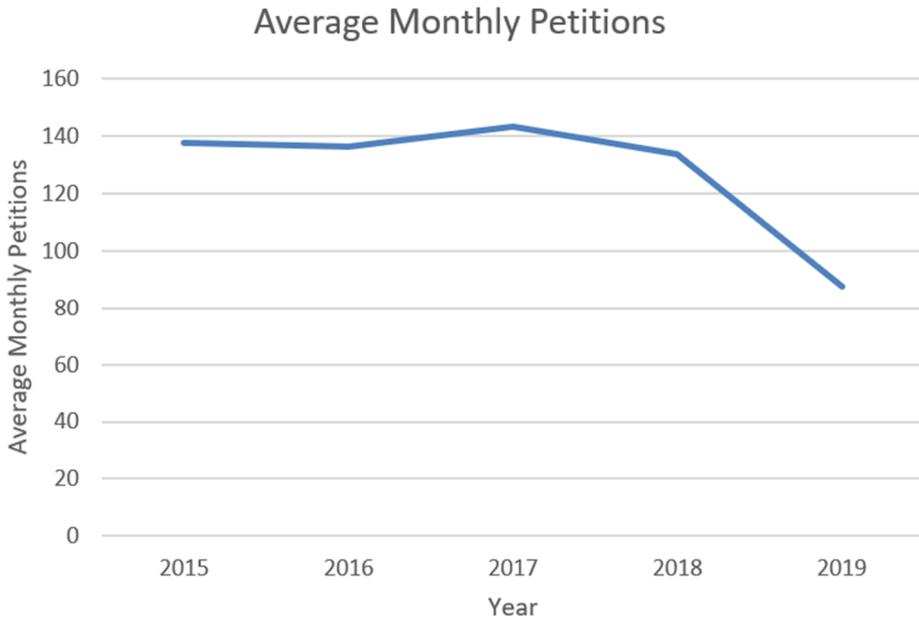
Figure 4: Petitions for IPR by year



197. The data was only available up to October 2019 at the time of writing this Note.

198. *Statistics*, USPTO, <https://www.uspto.gov/patents-application-process/patent-trial-and-appeal-board/statistics> (last visited June 6, 2021).

Figure 5: Average monthly petitions by year



Figures 6 and 7 further demonstrate this decrease by comparing rates of petitions by month. Figure 6 clearly demonstrates the decrease by comparing monthly petitions by year. As Figure 6 illustrates, the largest effect observed is in 2019. One bar demonstrates the number of petitions pre-October 2017 and the second bar the amount post-October 2017. Figure 7 adds another bar for 2019 with data available through November.

Further, a T-Test on pre- and post-Aqua Products petitions resulted in a T statistic of 3.51 with a p value of .004. This passes the significance test at level of .05. Some may argue that the results are skewed given that there were more years pre-Aqua Products, but, if we look at Figure 4, we can see that 2015 and 2016 seem fairly consistent and that the main difference in the data occurring after October 2017.

Figure 6: Comparing Petitions for IPR by Month pre- and post-*Aqua Products*

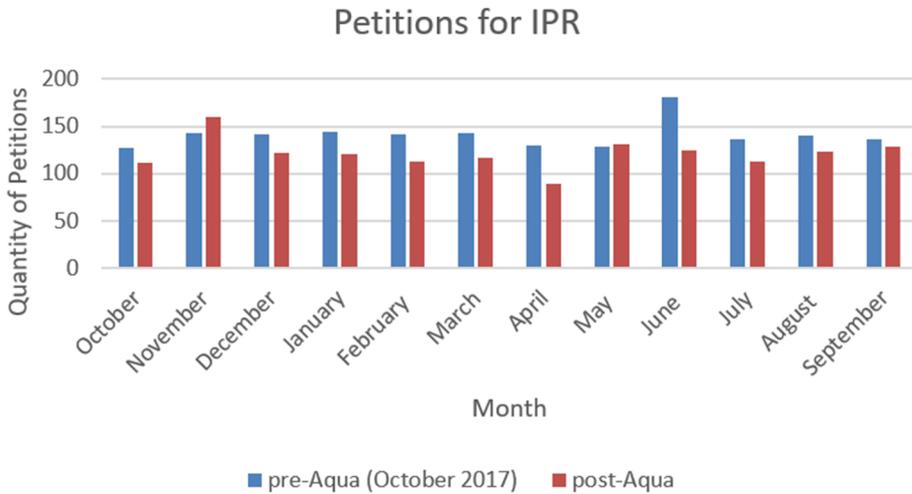
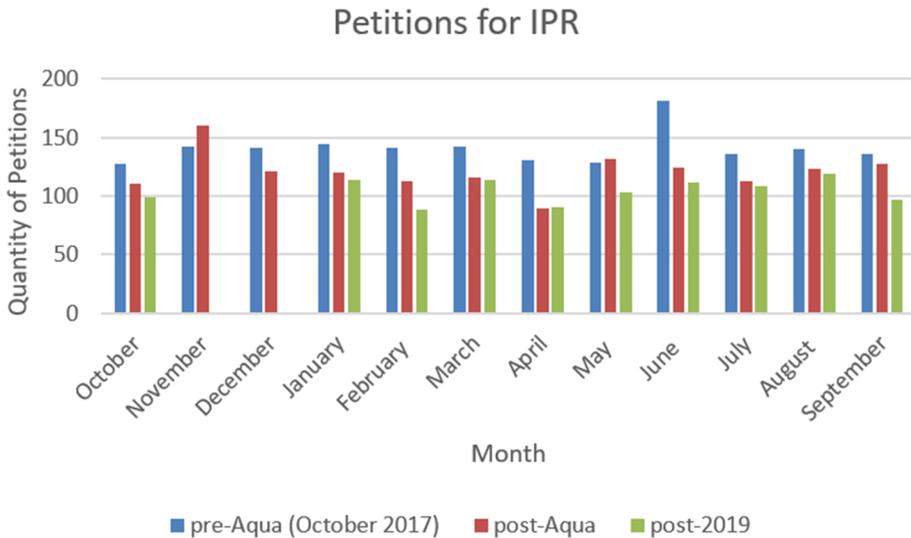


Figure 7: Comparing Petitions for IPR pre- and post-*Aqua Products*, and 2019



C. RATE OF STAYS GRANTED PENDING IPR

With the liberalization of post-issuance claim amendments in AIA trials since October 2017, we should expect to see an increase in the grant of stays pending IPR, because claims are likely to change during an IPR. Figure 8

demonstrates this increase by comparing the percentages of stays granted pending IPR by year, since 2013.¹⁹⁹

Figure 8: Comparing Yearly Percentages of Stays Granted Pending IPR

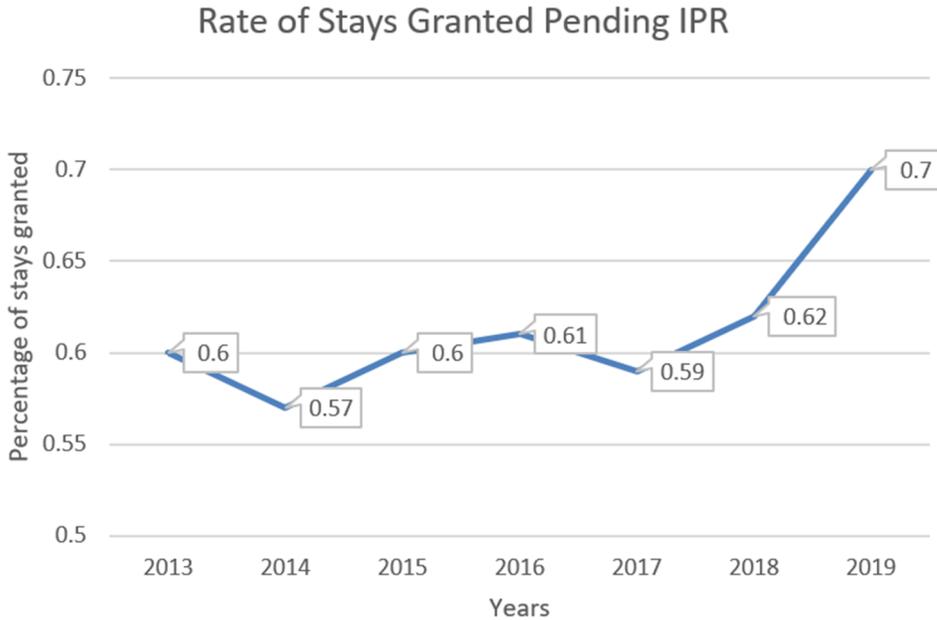
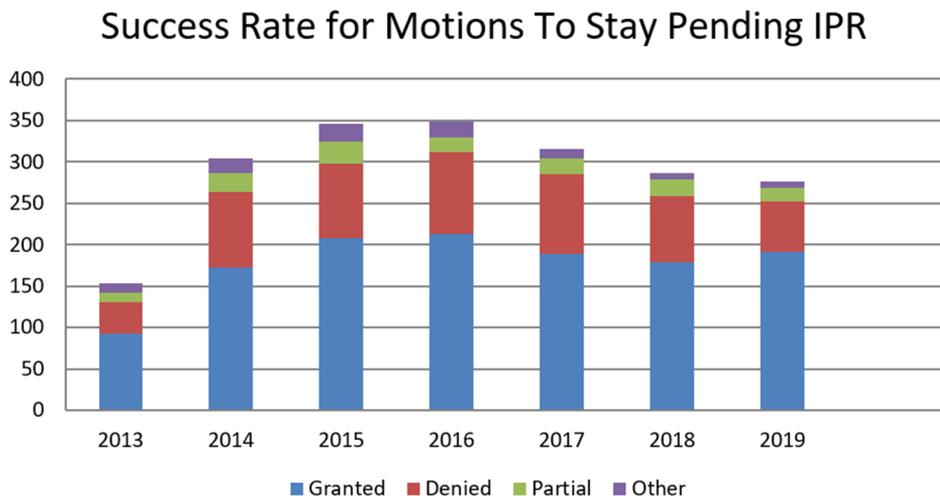


Figure 9, however, displays these totals as a bar graph and demonstrates that this trend may partly be explained because the patent bar is becoming more intelligent with motions to stay and moving for them less often.

199. The data here was collected from Docket Navigator and is complete through 2019.

Figure 9: Success Rate for Motions to Stay Pending IPR



In conclusion, the data demonstrate that the potential adverse effects of permitting post-issuance claim amendments during AIA trials have come to fruition. As demonstrated, since October 2017 (*Aqua Products*), (1) the rate of claim amendments filed and granted during AIA trials increased; (2) the rate of petitions for IPR stagnated or decreased; and (3) the rate of stays granted pending IPR increased.

VI. CONCLUSION

In conclusion, the “re-liberalization” of post-issuance claim amendments during AIA trials hurts the patent system. Claims serve two functions: forcing the patentee to define what she invented at the time of filing to prevent further incorporation of new insights and to give notice to the public about what she is claiming.²⁰⁰ Post-issuance claim amendments are detrimental for this second function. They are particularly detrimental for patents involved in concurrent district court litigation. Post-issuance claim amendments during validity trials undermine the AIA’s goal of mitigating burdensome litigation, create perverse drafting and strategic incentives for patent owners, and are redundant with other post-issuance claim amendment avenues. This Note offers empirical evidence to demonstrate that the hypothesized ill effects are in fact playing out. The rate of claim amendment motions filed and granted in AIA trials increased since October 2017. There has also been a decrease in the rate of

200. Chiang, *supra* note 15, at 560–61.

petitions for IPR and an increase in district court stays pending IPR, demonstrating empirically that IPR is not serving its intended purpose of mitigating burdensome litigation.

There are two obvious avenues for counteracting the ill effects of re-liberalization: abolishing claim amendments during AIA trials or enacting larger penalties for doing so. Along with these avenues, we should give more time to examiners up front to avoid “bad patents.”²⁰¹

Abolishing claim amendments during AIA trials would incentivize narrower claims at the outset during examination. This would not only help competitors more easily navigate patent boundaries, knowing that the boundaries of existing patents are less likely to change post-issuance, but would also help counteract the tremendous confusion surrounding the interpretation of patent claims by judges by pushing for clearer patent claims.²⁰² Rather than completely abolish the practice we could also more narrowly implement it by going after its more serious consequences. For example, we could only abolish post-issuance claim amendments for patents involved in concurrent litigation. We should, at a minimum, require notifying a district court if there is a likelihood of claim amendment. Regarding penalties, we could either increase fees or limit the doctrine of equivalents for amended claims. Either of these practices would help counteract the ill effects of post-issuance claim amendments during AIA trials.

201. Frakes & Wasserman, *supra* note 191, at 981.

202. This approach improves the quality of patent examination, enabling competitors to more easily navigate patent boundaries. It reduces the tremendous confusion surrounding the interpretation of patent claims by judges. See Peter S. Menell, *Reinvigorating Patent Notice Through Faithful Implementation of the Section 112(b) Claim Indefiniteness Provision* (U.C. Berkeley Pub. L. Rsch. Paper No. 2,403,902, 2014) (identifying that the innovation sector is awash with patents of uncertain scope because “the patent system rewards vague claiming of inventions in some technological fields”); Peter S. Menell & Michael J. Meurer, *Notice Failure and Notice Externalities*, 5 J. LEGAL ANALYSIS 1 (2013); Peter Menell, *It’s Time to Make Vague Software Patents More Clear*, WIRED (Feb. 7, 2013), <https://www.wired.com/2013/02/its-time-to-make-vague-software-patents-more-clear/>.

RE-CLASSIFYING GOVERNMENTAL PETITIONERS AS “PERSONS” IN AIA REVIEW PROCEEDINGS

Artin Au-Yeung[†]

I. INTRODUCTION

Does a government agency qualify as a person who can petition for certain proceedings at the U.S. Patent and Trademark Office (PTO)? Although not a natural person, a government agency, like a corporation, acts as a single unit towards a specific purpose. Notably, corporations are considered persons in PTO review proceedings,¹ available under the Leahy-Smith American Invents Act (AIA), and for other purposes.² Perhaps government agencies should then also qualify as persons in these AIA review proceedings, considering the government’s participation in the patent system. “Government agencies can apply for and obtain patents; they can maintain patents; they can sue other parties for infringing their patents; they can be sued for infringing patents held by private parties; [and] they can invoke certain defenses to an infringement lawsuit on the same terms as private parties”³

Nevertheless, in *Return Mail v. United States Postal Service*, the Supreme Court held that a government agency is not a “person” who can initiate AIA review proceedings. In *Return Mail*, the U.S. Postal Service (USPS) initiated an AIA proceeding to invalidate Return Mail’s patent claims. After the claims were invalidated, Return Mail argued that the government should be unable to initiate AIA review proceedings, and the Court agreed. Under the AIA, only a “person” can petition for AIA review proceedings. The Court focused on 1 U.S.C. § 1 (from the Dictionary Act),⁴ which provides a rebuttable presumption that the term “person” includes “corporations, companies,

DOI: <https://doi.org/10.15779/Z382J6853T>

© 2020 Artin Au-Yeung.

† J.D., 2020, University of California, Berkeley, School of Law. I would like to thank Professors Tejas N. Narechania, Talha Syed, and Kenneth A. Bamberger for their invaluable help. All mistakes are my own.

1. *See e.g.*, *Oil States Energy Servs., LLC v. Greene’s Energy Grp., LLC*, 138 S. Ct. 1365, 1368 (2018) (explaining that Greene’s Energy, a corporation, successfully petitioned for inter partes review against Oil States). AIA review proceedings occur after a patent is issued. *See Return Mail, Inc. v. United States Postal Serv.*, 139 S. Ct. 1853, 1860 (2019).

2. *See Citizens United v. Fed. Election Comm’n*, 558 U.S. 310, 365 (2010) (stating that “Government may not suppress political speech on the basis of the speaker’s corporate identity”).

3. *Id.* at 1871 (Breyer, J., dissenting).

4. *Id.* at 1862.

associations, firms, partnerships, societies, and joint stock companies, as well as individuals.”⁵ The Court concluded that government agencies are presumptively not persons because the Dictionary Act does not reference sovereigns.

The Court’s decision in *Return Mail* was incorrect. The PTO promulgated a rule allowing government agencies to petition for *ex parte* reexamination, the precursor to AIA review proceedings. Based on canons of statutory construction, the Court should have applied this administrative interpretation of reexamination to AIA review proceedings. After all, *ex parte* reexamination and AIA review proceedings share the same purposes (to increase patent quality and create an alternative to patent litigation), and government participation in AIA review proceedings furthers these purposes. Importantly, as this Note later discusses, counterarguments from the Court and Judge Newman of the Federal Circuit are not enough to justify upholding the decision in *Return Mail*.

Because the Court’s decision was incorrect, the PTO should rectify the Court’s mistake by promulgating a rule, which would state that government entities qualify as persons who can petition for AIA review proceedings. Under *National Cable & Telecommunications Association v. Brand X Internet Services*,⁶ an agency rule can overcome a prior court decision if the three steps of *Chevron* deference⁷ are satisfied. As this Note later explains, this proposed rule would meet all three steps.

This Note is structured as follows: Part II will explain the background of AIA review proceedings. Part III will then summarize the *Return Mail* decision. Part IV will discuss why the Court’s decision was incorrect and will rebut counterarguments that ostensibly support the Court’s decision. Part V will propose an agency-based solution using the Court’s decision in *Brand X*. The conclusion will summarize the arguments and the stakes of the decision.

II. PROCEEDINGS AT THE PATENT OFFICE

Originally, a third party could only use *ex parte* reexamination or *inter partes* reexamination to invalidate a patent. This changed in 2012 in response to criticism. The AIA created three additional PTO review proceedings that third

5. 1 U.S.C. § 1 (2018).

6. *See Nat’l Cable & Telecommunications Ass’n v. Brand X Internet Servs.*, 545 U.S. 967, 982, 986 (2005).

7. Under *Chevron* deference, if certain requirements are met, a Court must defer to an agency’s interpretation of a statute. *See United States v. Mead Corp.*, 533 U.S. 218, 227 (2001) (citing *Chevron, U.S.A., Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837, 844–45 (1984)).

parties could use to invalidate patents: (1) post-grant review (PGR),⁸ (2) inter partes review (IPR),⁹ and (3) covered business method review (CBM).¹⁰ These AIA proceedings directly descended from reexamination proceedings. As a result, reexamination and AIA review proceedings share the purposes of increasing patent quality and decreasing patent litigation.

A. REEXAMINATION

In 1980, Congress created *ex parte* reexamination, which still exists today.¹¹ The purposes of *ex parte* reexamination are (1) to increase patent quality¹² and (2) to create an alternative to expensive and lengthy district court litigation.¹³ During *ex parte* reexamination, the PTO reexamines the patentability of a patent's claims.¹⁴ To convince the Director to institute *ex parte* reexamination, a petitioner can point to "other patents or printed publications."¹⁵ When proceedings begin, third-party petitioners cannot participate.¹⁶ *Ex parte* reexamination proceedings resemble the PTO's initial examination procedures.¹⁷ To cancel patent claims, the PTO must find invalidity by a preponderance of the evidence.¹⁸ This differs from the higher standard for the invalidity defense in court, which requires clear and convincing evidence.¹⁹

Because *ex parte* reexamination received criticism,²⁰ Congress enacted *inter partes* reexamination in 1999.²¹ Congress noted that *ex parte* "reexamination has been used infrequently since a third party who requests reexamination cannot participate at all after initiating the proceedings."²² As a descendant to *ex parte* reexamination, *inter partes* reexamination shared the

8. 35 U.S.C. § 321 (2018).

9. 35 U.S.C. § 311 (2018).

10. Leahy-Smith America Invents Act (AIA), Pub. L. No. 112-129, § 18, 125 Stat. 284, 329 (2011).

11. 35 U.S.C. § 302 (2018).

12. *See* H.R. REP. NO. 96-1307, pt. 1, at 2–4 (1980) (discussing "creating a system of administrative reexamination of doubtful patents" to "strengthen[] investor confidence in the certainty of patent rights" and "restore confidence in the effectiveness of our patent system.").

13. *See id.* at 3–4.

14. *See* 35 U.S.C. § 302.

15. 35 U.S.C. §§ 301, 302, 303(a), 304 (2018).

16. *Return Mail, Inc. v. United States Postal Serv.*, 139 S. Ct. 1853, 1866 (2019) (citing 36 U.S.C. §§ 302, 303).

17. *See* 35 U.S.C. § 305 (2018).

18. MPEP § 2286 (9th ed. Rev. June 2020).

19. *See Microsoft Corp. v. i4i Ltd. P'ship*, 131 S. Ct. 2238, 2240, 2245 (2011) (citing 35 U.S.C. § 282 (2006)).

20. H.R. REP. NO. 112-98, at 46 (2011) (summarizing critiques of *ex parte* reexamination).

21. *Id.*

22. H.R. REP. NO. 106-464, at 133 (1999).

same goals of increasing patent quality²³ and decreasing patent infringement litigation.²⁴ To better effectuate these goals, Congress made improvements to inter partes reexamination.²⁵ Third parties could now participate by writing comments that addressed issues raised by the PTO or the patent owner's response.²⁶ Also, challengers could appeal unfavorable decisions.²⁷ However, because inter partes reexamination was barely used, inter partes review and other AIA review proceedings replaced it²⁸ on September 16, 2012.²⁹

B. AIA REVIEW PROCEEDINGS

The AIA created three review proceedings: PGR, IPR, and CBM.³⁰ Like their ancestors, they were intended to increase patent quality and create a viable alternative to district court litigation.³¹ Similar to inter partes reexamination, a third-party petitioner can participate in all AIA review proceedings.³² However, given that inter partes reexamination failed, Congress structured AIA review proceedings differently to effectuate the aforementioned purposes.³³ For instance, the Patent Trial and Appeal Board (PTAB) oversees AIA review proceedings,³⁴ which are more adjudicatory and adversarial.³⁵ Parties have access to limited discovery and short oral hearings,³⁶ and a party who is unsatisfied with the decision can appeal to the Federal Circuit.³⁷

The three AIA proceedings share important traits. Similar to reexamination, the standard of proof is a preponderance of the evidence.³⁸ In addition, statutory estoppel provisions preclude petitioners from making arguments, which they already raised during prior AIA review proceedings, in

23. H.R. REP. NO. 107-120, at 3 (2001) (stating that one of the original goals of ex parte reexamination was to “reinforce investor confidence in the certainty of patent rights by affording an opportunity to review patents of doubtful validity”).

24. See H.R. REP. NO. 106-464, at 133.

25. See H.R. REP. NO. 112-98, at 45–46 (2011).

26. 35 U.S.C. § 314(b)(2) (2000).

27. H.R. REP. NO. 112-98, at 46.

28. See *id.* at 46–47.

29. See AIA § 35 (2011).

30. 35 U.S.C. §§ 311, 321 (2018); AIA § 18.

31. See H.R. REP. NO. 112-98, at 40.

32. See *Return Mail, Inc. v. United States Postal Serv.*, 139 S. Ct. 1853, 1866 (2019).

33. See H.R. REP. NO. 112-98, at 45–47.

34. See 35 U.S.C. §§ 316(c), 326(c) (2018).

35. See *Return Mail*, 139 S. Ct. at 1866.

36. See *Regents of the Univ. of Minnesota v. LSI Corp.*, 926 F.3d 1327, 1336 (Fed. Cir. 2019), *cert. denied*, No. 19-337, 2020 WL 129563 (U.S. Jan. 13, 2020); 37 CFR §§ 42.51, 42.70 (2019).

37. 35 U.S.C. § 141(c) (2018).

38. 35 U.S.C. §§ 316(e), 326(e) (2018).

subsequent district court or International Trade Commission (ITC) proceedings.³⁹ Furthermore, statutory estoppel applies to arguments that a petitioner “reasonably could have raised” in a prior PGR or IPR.⁴⁰

The first of the proceedings, PGR, permits any person, other than the patent owner, to petition for review of a patent’s validity based on any grounds a party can rely on in litigation.⁴¹ A petitioner can only initiate a PGR within nine months of the challenged patent’s issuance date.⁴²

IPR, another AIA proceeding, directly replaced *inter partes* reexamination.⁴³ Any person other than the patentee can initiate IPR.⁴⁴ A person can typically petition for IPR nine months after the patent’s issuance date.⁴⁵ IPR proceedings are limited to issues of novelty and nonobviousness, and petitioners can only rely on “prior art consisting of patents or printed publications.”⁴⁶ Notably, parties may initiate PGR and IPR without the constitutional standing⁴⁷ required for asserting an invalidity defense⁴⁸ or requesting a declaratory judgment of invalidity in court.⁴⁹

The AIA also created CBM, which expired on September 16, 2020⁵⁰ and only applied to covered business method patents.⁵¹ Unlike PGR and IPR, a person could not petition for CBM “unless the person or the person’s real party in interest or privity has been sued for infringement of the patent or has been charged with infringement under that patent.”⁵² Similar to PGR,

39. 35 U.S.C. §§ 315(e)(2), 325(e)(2) (2018); AIA § 18(a)(1)(D).

40. 35 U.S.C. §§ 315(a), 325(e) (2018).

41. 35 U.S.C. §§ 321(a)–(b) (2018).

42. *Id.* § 321(c).

43. H.R. REP. NO. 112-98 at 46–47.

44. 35 U.S.C. § 311(a).

45. *See id.* § 311(c)(1).

46. *Id.* § 311(b).

47. *See* *Cuozzo Speed Techs., LLC v. Lee*, 136 S. Ct. 2131, 2143–44 (2016); *see also* *Regents of the Univ. of Minnesota v. LSI Corp.*, 926 F.3d 1327, 1335 (Fed. Cir. 2019) (citing *Cuozzo*, 136 S. Ct. at 2143–44).

48. *See* 35 U.S.C. § 282(b)(2) (2018).

49. *See* Brief of Amicus Curiae Professor Tejas N. Narechania in Support of Respondents at 12–13, *Return Mail, Inc. v. United States Postal Service*, 139 S. Ct. 1853 (2019) (No. 17-1594) (citations omitted).

50. AIA § 18(a)(3)(A).

51. *Id.* § 18(d)(1). A covered business method patent “claims a method or corresponding apparatus for performing data processing or other operations used in the practice, administration, or management of a financial product or service . . .” *Id.*

52. *Id.* § 18(a)(1)(B).

petitioners in CBM could rely on any arguments related to invalidity and could base their arguments on more than just patents and printed publications.⁵³

III. THE *RETURN MAIL* DECISION

A. FACTS OF *RETURN MAIL*

Return Mail, a corporation, owned U.S. Patent No. 6,826,548 (the '548 Patent), which deals with the processing of undeliverable mail items.⁵⁴ In the Court of Federal Claims,⁵⁵ Return Mail alleged that USPS infringed its patent.⁵⁶ After USPS petitioned for CBM of the '548 patent's relevant claims,⁵⁷ Return Mail argued that USPS should have been unable to initiate CBM.⁵⁸ The PTAB disagreed and held that the patent claims were invalid.⁵⁹

B. FEDERAL CIRCUIT DECISION

Return Mail appealed. The Federal Circuit, affirming the PTAB, held that a government agency can petition for CBM.⁶⁰ Because a party must be sued for or charged with infringement to initiate CBM, the Federal Circuit discussed whether the government can be sued for infringement. When the government commits a taking of a patent license, the patentee can only sue the government in the Court of Federal Claims under 28 U.S.C. § 1498(a).⁶¹ The Federal Circuit held that section 1498 suits qualify as patent infringement actions, meaning that government agencies sued under section 1498 can petition for CBM.⁶²

Judge Newman dissented, arguing that the government is not a person who can initiate AIA review proceedings.⁶³ She looked to the Dictionary Act, which states that a "person" presumptively includes certain parties but does

53. MARK A. LEMLEY, PETER S. MENELL & ROBERT P. MERGES, *INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE*: 2018 318, 323 (2018).

54. *Return Mail, Inc. v. U.S. Postal Serv.*, 868 F.3d 1350, 1354 (Fed. Cir. 2017), *rev'd*, 139 S. Ct. 1853 (2019).

55. The Court of Federal Claims is an Article I Court. *About the Court*, UNITED STATES COURT OF FEDERAL CLAIMS, <https://www.usfc.uscourts.gov/about-court> (last visited Feb. 5, 2020). It is the only venue by which a patentee can sue the government for patent infringement. 28 U.S.C. § 1498(a).

56. *Return Mail, Inc. v. United States*, No. 11-130 C, 2013 WL 5569433, at *1 (Fed. Cl. Oct. 4, 2013).

57. *See* *United States Postal Service and United States v. Return Mail, Inc.*, CBM2014-00116 1, 2 (2014).

58. *Return Mail*, 868 F.3d at 1355.

59. *Return Mail*, CBM2014-00116 at 3, 12.

60. *Return Mail*, 868 F.3d at 1365–67, 1371.

61. *Id.* at 1361–63; 28 U.S.C. § 1498(a) (2018).

62. *Return Mail*, 868 F.3d at 1362–63.

63. *Id.* at 1375 (Newman J., dissenting).

not include sovereigns (like the U.S. government) in its definition.⁶⁴ She also pointed to the statutory estoppel provisions of AIA review proceedings, which mentions district court litigation and ITC proceedings,⁶⁵ but does not mention section 1498 patent infringement suits.⁶⁶ According to Judge Newman, Congress did not include section 1498 suits because it did not intend for government agencies to be eligible petitioners.⁶⁷ Otherwise, government agencies would be able to relitigate arguments that failed in an AIA review proceeding.⁶⁸

The majority disagreed. It explained that the statutory purpose, “the context, the subject matter, legislative history, and executive interpretation”⁶⁹ can rebut the presumption against sovereigns. The majority contended that the presumption carried less weight because the statute conferred a “benefit or advantage” on the government.⁷⁰ It also noted that “[t]he AIA does not appear to use the term ‘person’ to exclude the government in other provisions.”⁷¹

Concerning the absence of section 1498 suits in the estoppel provisions, the majority rejected the dissent’s argument because the AIA’s legislative history did not directly address policy concerns of potential re-litigation of arguments.⁷² It then concluded that there was no reason to exclude the government from CBM when, “like a party sued in federal district court or the ITC, it has interests at stake with respect to the patent it has been accused of infringing.”⁷³

C. SUPREME COURT DECISION

The Supreme Court reversed and held that federal agencies cannot petition for AIA review proceedings.⁷⁴ Under the AIA, a “person” can petition for AIA review proceedings.⁷⁵ The Court construed the definition of “person” by applying the presumption from the Dictionary Act, which it interpreted to mean that a person does not include sovereigns.⁷⁶ Thus, the Court presumed

64. *Id.* at 1372 (Newman, J., dissenting) (citing 1 U.S.C. § 1).

65. *Id.* at 1373–74 (Newman, J., dissenting).

66. *See id.*

67. *See id.*

68. *Id.* at 1375 (Newman J., dissenting).

69. *Id.* at 1365 (quoting *Wilson v. Omaha Indian Tribe*, 442 U.S. 653, 667 (1979)).

70. *Id.* (quoting *Wilson*, 442 U.S. at 667).

71. *Id.*

72. *See id.* at 1366.

73. *Id.*

74. *Return Mail, Inc. v. United States Postal Serv.*, 139 S. Ct. 1853, 1859 (2019).

75. 35 U.S.C. §§ 311(a), 321(a); AIA § 18(a)(1)(B).

76. *Return Mail*, 139 S. Ct. at 1861–62.

that government agencies cannot petition for AIA review proceedings.⁷⁷ The Court found that the legislative history, executive interpretation, purpose, subject matter, and context did not rebut the presumption.⁷⁸

The Court rejected USPS's attempts to rebut the presumption. USPS first argued that other usages of the word "person" in the Patent Act included the government.⁷⁹ For instance, 35 U.S.C. § 296(a) stated that any person, "including any governmental or nongovernmental entity," can sue a state in federal court for patent infringement.⁸⁰ The Court was unpersuaded, since other usages of "person" in that Act do not include the government. 35 U.S.C. § 6(a) states that administrative patent judges, who cannot include federal agencies, must be persons.⁸¹ USPS also pointed to a provision stating that federal agencies are "persons" who can obtain patents.⁸² The Court found that the matters of obtaining a patent and petitioning to invalidate a patent are unrelated.⁸³

USPS's second argument centered on the Manual of Patent Examining Procedure (MPEP),⁸⁴ which states that a governmental entity is a "person" who can petition for ex parte reexamination or cite prior art.⁸⁵ Although pre-AIA and AIA proceedings share the same goals,⁸⁶ the Court rejected USPS's argument. It reasoned that, while AIA review proceedings are adjudicative and adversarial, ex parte reexamination is internal and lacks third-party participation.⁸⁷ The Court noted that the government already enjoys unique benefits in a different adversarial proceeding involving patent validity.⁸⁸ Specifically, the government can only be sued in the Court of Federal Claims

77. *Id.* at 1863.

78. *See id.* at 1868 (Breyer, J., dissenting) (citing *Int'l Primate Prot. League v. Administrators of Tulane Educ. Fund*, 500 U.S. 72, 83 (1991)).

79. *Id.* at 1863.

80. *Id.* at 1863 n.3 (citing 35 U.S.C. § 296(a) (2018)).

81. *Id.* at 1863, 1863 n.4 (citing 35 U.S.C. § 6(a) (2018)).

82. *Id.* at 1864.

83. *Id.* at 1864.

84. According to the Foreword of the MPEP, the MPEP is a PTO manual that provides guidance for patent examination. MPEP (9th ed. Rev. June 2020).

85. *See Return Mail*, 139 S. Ct. at 1865 (citing MPEP §§ 2203, 2212 (4th ed. Rev. July 1981)). For purposes of citing prior art, "[a]ny person" may be a corporate or governmental entity as well as an individual." MPEP § 2203 (9th ed. Rev. June 2020). For ex parte reexamination, "Corporations and/or governmental entities are included within the scope of 'any person.'" MPEP § 2212 (9th ed. Rev. June 2020).

86. *See Return Mail*, 139 S. Ct. at 1865–66.

87. *Id.*

88. *Id.*

for infringement,⁸⁹ where trial by jury, injunctions, and punitive damages are unavailable.⁹⁰

The Court also rejected USPS's third argument—that the government's ability to assert an invalidity defense in litigation demonstrates that it should be able to petition for AIA review proceedings.⁹¹ According to the Court, Congress could have decided to only provide AIA review proceedings to nongovernmental actors. Unlike private actors, when the government asserts an invalidity defense, it enjoys the aforementioned benefits under 28 U.S.C. § 1498.⁹²

The Court concluded by expressing a concern that it would be “awkward” for a civilian patent owner to defend its patent in a proceeding that one government agency initiated and another agency (the PTO) oversaw.⁹³

Lastly, in a footnote, the Court addressed, but did not rely upon, Judge Newman's argument in her dissent⁹⁴ regarding statutory estoppel.⁹⁵ The Court recognized that common law estoppel could potentially remedy the situation, writing that “the practical effect of the estoppel provisions' potential inapplicability to the Government is uncertain given that this Court has not decided whether common-law estoppel applies in 28 U.S.C. § 1498 suits.”⁹⁶

IV. THE PROBLEMS WITH THE SUPREME COURT'S DECISION

This Part explains why the Court's holding was in error and addresses counterarguments from the Court and Judge Newman.

Because the PTO permitted government agencies to petition for *ex parte* reexamination, the Court should have similarly allowed government agencies to petition for AIA review proceedings. Under the prior construction canon, Congress presumably incorporates prior administrative interpretations of an existing statutory provision into a new statute that uses the same language. Given that the two types of proceedings share the same purposes and that

89. *Id.* at 1866–67.

90. *Id.*

91. *Id.* at 1866.

92. *Id.* at 1867 (discussing the lack of injunctive relief, jury trials, and punitive damages in the Court of Federal Claims).

93. *Id.*

94. *Return Mail, Inc. v. U.S. Postal Serv.*, 868 F.3d 1350, 1373–75 (Fed. Cir. 2017) (Newman, J., dissenting) (arguing that because statutory estoppel does not apply to section 1498 suits, if the government could petition for AIA review proceedings, it would have the ability to relitigate its arguments, from a failed AIA review proceeding, in court).

95. *Id.* at 1867 n.10.

96. *Id.*

allowing the government to initiate AIA review proceedings furthers those purposes, it makes sense to apply the prior construction canon here.

Furthermore, the counterarguments presented by the Court and Judge Newman are not persuasive. First, although Congress provided the government with unique treatment under 28 U.S.C. § 1498, this special treatment is justified by other legal doctrines, and allowing the government to initiate AIA review proceedings would not chill inventors from obtaining patents. Second, it is not awkward for a civilian to defend itself against a government agency before another government agency. This does not create a conflict between agencies that must be resolved by the President. Moreover, the PTO does not necessarily act with bias in favor of other agencies, and other proceedings exist that involve similar situations. Third, Congress could have meant to subject government agencies only to common law estoppel, instead of the broader statutory estoppel provisions of the AIA. The government has limited resources and thus is less likely to harass patentees, and a government agency's decision to initiate AIA review proceedings can implicate unique public policy concerns.

A. THE ADMINISTRATIVE INTERPRETATION OF A PERSON FOR REEXAMINATION APPLIES TO AIA REVIEW PROCEEDINGS

Although the MPEP (since 1981), in the context of *ex parte* reexamination and citation of prior art, explained that a person included government entities,⁹⁷ the Court in *Return Mail* was not convinced that the Dictionary Act's presumption was rebutted.⁹⁸ The Court reasoned that *ex parte* reexamination differs from AIA review proceedings, as the former is not adversarial.⁹⁹

The Court recognized the prior construction canon¹⁰⁰ that “[w]hen administrative and judicial interpretations have settled the meaning of an existing statutory provision, repetition of the same language in a new statute indicates as a general matter, the intent to incorporate its administrative and judicial interpretations as well.”¹⁰¹ However, it did not apply the canon because

97. MPEP §§ 2203, 2212 (4th ed. Rev. July 1981); MPEP §§ 2203, 2212 (9th ed. Rev. Jan. 2018).

98. *Return Mail, Inc. v. United States Postal Serv.*, 139 S. Ct. 1853, 1865–66 (2019).

99. *Id.* at 1866.

100. The Supreme Court has recognized this canon. *Lightfoot v. Cendant Mortg. Corp.*, 137 S. Ct. 553, 563 (2017) (using the term “prior construction canon”).

101. *Return Mail*, 139 S. Ct. at 1866 (quoting *Bragdon v. Abbott*, 524 U.S. 624, 645 (1998)) (alteration in original); *see also* *Silver v. Pueblo Del Sol Water Co.*, 244 Ariz. 553, 560 (2018) (“According to that canon, “[i]f a statute uses words or phrases that have already received . . . uniform construction by . . . a responsible administrative agency, they are to be understood according to that construction.”) (quoting ANTONIN SCALIA & BRYAN A.

it claimed that “there is no ‘settled’ meaning of the term ‘person’ with respect to the newly established AIA review proceedings.”¹⁰² On the contrary, the Court should have applied the above canon and relied on the ex parte reexamination rule from MPEP § 2212. Section 2212 settled the meaning of a “person” who could petition for ex parte reexamination. Based on the prior construction canon, this settled meaning of “person” for ex parte reexamination should have applied to the “closely related”¹⁰³ AIA review proceedings.

1. *The Settled Meaning of Eligible Petitioners for Ex Parte Reexamination*

To the extent that there is a question concerning the validity of the rules in the MPEP, MPEP § 2212 is entitled to deference.¹⁰⁴ As a result, MPEP § 2212 settled the meaning of eligible petitioners for ex parte reexamination.

Under 35 U.S.C. § 302 (2018), “[a]ny person at any time may file a request for reexamination by the Office . . .” MPEP § 2212, interpreting section 302, stated that “governmental entities are included within the scope of the term ‘any person.’” Although the “MPEP has no binding effect on [a] court,” “its provisions represent the PTO’s interpretation of the statutes and regulations, and an agency’s interpretation of statutes and regulations deserves some deference.”¹⁰⁵ Thus, the MPEP “is entitled to judicial notice as the agency’s official interpretation of statutes or regulations, provided that it is not in conflict with the statutes or regulations.”¹⁰⁶ For example, in *Bettcher Industries*,

GARNER, READING LAW: THE INTERPRETATION OF LEGAL TEXTS 322 (2012)) (alterations in original).

102. *Return Mail*, 139 S. Ct. at 1866.

103. *See SAS Inst., Inc. v. Iancu*, 138 S. Ct. 1348, 1355 (2018) (finding that the PTO Director could not initiate IPR because the IPR statute, unlike the closely related ex parte reexamination statute, did not explicitly give the PTO Director that authority).

104. Administrative interpretations do not need to be entitled to *Chevron* deference to settle the meaning of statutory language. *Bragdon v. Abbott*, 524 U.S. 624, 642, 645 (1998) (finding that administrative interpretations were relevant to determining whether the meaning of a term was settled even without looking to the *Chevron* deference steps). Although interpretations entitled to *Skidmore* deference are sufficient, the *Bragdon* Court did not comment on whether being entitled to *Skidmore* deference is necessary. *Id.* *Skidmore* deference refers to the principle that “reasonable agency interpretations carry ‘at least some added persuasive force’ where *Chevron* is inapplicable.” *United States v. Mead Corp.*, 533 U.S. 218, 235 (2001) (quoting *Metro. Stevedore Co. v. Rambo*, 521 U.S. 121, 136 (1997)).

105. *Molins PLC v. Quigg*, 837 F.2d 1064, 1067 (Fed. Cir. 1988) (citations omitted).

106. *Refac Int’l, Ltd. v. Lotus Dev. Corp.*, 81 F.3d 1576, 1584 (Fed. Cir. 1996) (citations omitted); *see also Enzo Biochem, Inc. v. Gen-Probe Inc.*, 323 F.3d 956, 964 (Fed. Cir. 2002) (stating that the MPEP is “not binding on this court, but may be given judicial notice to the extent [it does] not conflict with the statute”) (citations omitted); *Airbus S.A.S. v. Firepass Corp.*, 793 F.3d 1376, 1380 (Fed. Cir. 2015) (quoting *Litton Sys., Inc. v. Whirlpool Corp.*, 728 F.2d 1423, 1439 (Fed. Cir. 1984)) (“Although the MPEP does not have the force of law, it is

Inc. v. Bunzl USA, Inc., the Federal Circuit held that a MPEP rule clarifying the statutory language, “finally determined,” addressed a procedural issue and would have been “entitled to deference” if Congress’s intent was ambiguous.¹⁰⁷

Similarly, MPEP § 2212, interpreting “person” for a procedural matter, is entitled to deference. Because the Dictionary Act’s presumption against sovereigns is not a hard and fast rule of exclusion and is rebuttable by executive interpretation, the term “person” is ambiguous.¹⁰⁸ This was likely why, in *Return Mail*, the Court said it might have considered MPEP § 2212 if it was determining whether agencies could initiate *ex parte* reexamination.¹⁰⁹ Importantly, MPEP § 2212 is also a long-standing rule because it is one of the “original reexamination rules that were adopted to implement the reexamination statute”¹¹⁰ in July 1981. “As a contemporaneous agency construction of the statute, this interpretation is entitled to great weight.”¹¹¹

Also, MPEP § 2212 is not in conflict with the *ex parte* reexamination statute because it is consistent with Congress’s purposes behind it.¹¹² Specifically, the government, by using 35 U.S.C. §§ 301–302, can cite prior art to the PTO and petition for *ex parte* reexamination. By first introducing prior art in 35 U.S.C. § 301 and then explaining “the pertinency and manner of applying cited prior art to every claim for which reexamination is requested,”¹¹³ the government can test the patent’s quality and potentially lower the volume of district court litigation if the relevant patent claims are invalidated. In sum,

entitled to judicial notice ‘so far as it is an official interpretation of statutes or regulations with which it is not in conflict.’”).

107. *See Bettcher Indus., Inc. v. Bunzl USA, Inc.*, 661 F.3d 629, 644, 646 (Fed. Cir. 2011) (citing *Cooper Techs. Co. v. Dudas*, 536 F.3d 1330, 1336 (Fed. Cir. 2008) for the proposition that the rule at issue was not a substantive rule and that it clarified “ambiguous statutory language”). *But see Non VirnetX Inc. v. Apple Inc.*, 931 F.3d 1363, 1377 n.11 (Fed. Cir. 2019) (explaining that the discussion of the MPEP in *Bettcher* was dicta).

108. For a more in-depth discussion of this, see *infra* Section V.B. Section V.B discusses AIA review proceedings, but the reasoning is also applicable to *ex parte* reexamination.

109. *See Return Mail, Inc. v. United States Postal Serv.*, 139 S. Ct. 1853, 1865 (2019).

110. *In re Recreative Techs. Corp.*, 83 F.3d 1394, 1397 (Fed. Cir. 1996) (rejecting an MPEP rule and noting that it “was not among the original reexamination rules”).

111. *See Les v. Reilly*, 968 F.2d 985, 989 (9th Cir. 1992) (citing *Norwegian Nitrogen Products Co. v. United States*, 288 U.S. 294, 315, (1933)). The Supreme Court has, at times, cited “an agency interpretation’s longstanding pedigree without reference to any of the Supreme Court’s deference regimes or tests” when upholding the agency interpretation. Anita S. Krishnakumar, *Longstanding Agency Interpretations*, 83 FORDHAM L. REV. 1823, 1832 (2015) (citing numerous Supreme Court case examples, including *Smith v. City of Jackson*, 544 U.S. 228, 239–40 (2005); *W. Air Lines, Inc. v. Criswell*, 472 U.S. 400, 412 (1985); and *Tony & Susan Alamo Found. v. Sec’y of Labor*, 471 U.S. 290, 296–97 (1985)).

112. The reasoning in Section IV.A.2.b, which applies to AIA review proceedings, is also applicable to *ex parte* reexamination, since they share the same purposes.

113. 35 U.S.C. §§ 301–02 (2018).

MPEP § 2212's interpretation of eligible petitioners for ex parte reexamination is entitled to deference.

Importantly, even if MPEP § 2212 is found invalid in the future, that “does not affect the backdrop against which Congress legislated” the AIA in 2011.¹¹⁴ Thus, Congress could still have meant to incorporate MPEP § 2212 into the AIA, even if the rule's validity is later questioned in court.

2. *Incorporating the Settled Meaning of “Person” for Ex Parte Reexamination into AIA Review Proceedings*

The Court should have found that Congress had MPEP § 2212's settled meaning of “person” in mind when it created AIA review proceedings. After all, the ex parte reexamination statute is closely related to the AIA review proceedings statutes for two reasons: (a) reexamination and AIA review proceedings share the same purposes (to increase the quality of patents and provide a cost-effective alternative to litigation), and (b) permitting the government to petition for AIA review proceedings furthers these shared purposes.¹¹⁵ In other words, Congress repeated the term “person” in the AIA “against the background of . . . regulatory history”¹¹⁶

a) Shared Purposes

The shared purposes behind reexamination and AIA review proceedings demonstrate that a “person” for AIA review proceedings directly relates to a third-party “person” who could initiate ex parte reexamination. The Federal Circuit has also emphasized these shared goals and used the understandings of ex parte reexamination to inform discussions of AIA review proceedings.

First, inter partes reexamination and AIA review proceedings directly descended from ex parte reexamination, and all three proceedings share the

114. *See Johnson v. United States*, 529 U.S. 694, 710–11 n.11 (2000) (discussing statutes for parole following prison sentences and finding that an administrative interpretation of a statute for special parole was applicable to a related statute on supervised release, even though some courts had recently decided that the administrative interpretation was inconsistent with the original statute authorizing special parole); *see also id.* at 726 n.7 (Scalia, J. dissenting) (explaining that the Court was stretching the prior construction canon beyond its original boundaries).

115. The Supreme Court has found that an administrative interpretation of a prior similar statute, which arguably covered a different subject than the statute at issue, should be incorporated into the statute at issue. *See supra* note 114. The same logic of the *Johnson* majority is applicable to the closely related ex parte reexamination and AIA review proceedings.

116. *See Nat'l Cable & Telecommunications Ass'n v. Brand X Internet Servs.*, 545 U.S. 967, 992 (2005).

purpose of increasing patent quality and creating an efficient and effective alternative to long and expensive district court litigation.¹¹⁷

Congress enacted *ex parte* reexamination to “permit efficient resolution of questions about the validity of issued patents without recourse to expensive and lengthy infringement litigation”¹¹⁸ and to address potentially low quality patents and increase confidence in the patent system.¹¹⁹ However, Congress noted that third parties did not frequently initiate *ex parte* reexamination because they could not participate in the proceeding.¹²⁰ Due to “criticisms of the reexamination system,” Congress made “amendments to the process” by enacting *inter partes* reexamination.¹²¹

Congress similarly aimed to use *inter partes* reexamination to “reduce expensive patent litigation in U.S. district courts”¹²² and strike a better balance “toward the goal of improving patent quality and validity.”¹²³ With regard to improving patent quality, “[t]he participation by third parties is considered vital because in many circumstances they have the most relevant prior art available and incentive to seek to invalidate an allegedly defective patent.”¹²⁴ *Inter partes* reexamination was also “intended to reduce expensive patent litigation in U.S. district courts by giving third-party requesters, *in addition to the existing ex parte reexamination . . .*, the option of *inter partes* reexamination proceedings”¹²⁵ Thus, a third-party “person,” for purposes of *ex parte* reexamination, was the same third party who could petition for *inter partes* reexamination. In fact, the *inter partes* reexamination statute defined “third-party requestor” to mean “a person requesting *ex parte* reexamination under section 302 or *inter partes* reexamination under section 311 who is not the patent owner.”¹²⁶

However, since *inter partes* reexamination was rarely used,¹²⁷ Congress “convert[ed]” and “rename[d]” it to IPR¹²⁸ and created other AIA review proceedings.¹²⁹ Congress allowed “a person who is not the owner of a patent”

117. See H.R. REP. NO. 112-98, at 40 (2011).

118. H.R. REP. NO. 96-1307, pt.1, at 3 (1980).

119. See *id.* at 2, 4.

120. H.R. REP. NO. 106-464, at 133 (1999).

121. H.R. REP. NO. 112-98, at 46.

122. H.R. REP. NO. 106-464, at 133.

123. H.R. REP. NO. 107-120, at 4 (2001).

124. *Id.*

125. H.R. REP. NO. 106-464, at 133 (emphasis added).

126. PL 106-113, § 4603, 113 Stat. 1501, 1501A-567 (1999).

127. See H.R. REP. NO. 112-98, at 46 (2011).

128. *Id.* at 46–47.

129. *Return Mail, Inc. v. United States Postal Serv.*, 139 S. Ct. 1853, 1860 (2019) (“In 2011, Congress . . . phased out *inter partes* reexamination. In its stead, the AIA tasked the Board with three new types of post-issuance review proceedings.”) (citations omitted).

to petition for IPR and PGR of “the patent,”¹³⁰ and it allowed a “person” who was sued for or charged with infringement of a patent to petition for CBM of that patent.¹³¹ Congress, once again, repeated its dual goals to “establish a more efficient and streamlined patent system that will improve patent quality and limit unnecessary and counterproductive litigation costs.”¹³² For IPR, it referenced the goal of ex parte reexamination to “serve as an effective and efficient alternative to often costly and protracted district court litigation.”¹³³ It created PGR to “make the patent system more efficient and improve the quality of patents and the patent system,”¹³⁴ and designed CBM to address “the flood of poor quality business method patents”¹³⁵ Like with inter partes reexamination, third-party participation in AIA review proceedings was considered “vital”¹³⁶ for increasing patent quality, as AIA review “proceedings are designed to allow the USPTO to harness *third parties* for the agency to evaluate whether a prior grant of a public franchise was wrong”¹³⁷

Second, the Federal Circuit has looked to the shared purposes between ex parte reexamination and IPR, as well as the law surrounding ex parte reexamination, when deciding legal matters related to IPR. Three Federal Circuit cases demonstrate this comparison between ex parte reexamination and IPR.

In *MCM Portfolio LLC v. Hewlett-Packard Co.*, the Federal Circuit held that IPR does not contradict Article III of the Constitution.¹³⁸ The Federal Circuit compared the purposes of IPR and ex parte reexamination, which were to improve patent quality and provide an alternative to litigation.¹³⁹ It then looked to *Patlex Corp. v. Mossinghoff*,¹⁴⁰ in which the court held that ex parte reexamination did not contradict Article III.¹⁴¹ Seeing “no basis to distinguish the reexamination proceeding in *Patlex* from inter partes review,” the Federal Circuit held that IPR does not contradict Article III.

130. 35 §§ U.S.C. 311(a), 321(a).

131. See AIA § 18(a)(10)(B).

132. H.R. REP. NO. 112-98, at 40.

133. *Id.* at 46.

134. *Id.* at 48.

135. See 157 CONG. REC. S1363 (daily ed. Mar. 08, 2011) (statement of Senator Schumer).

136. See H.R. REP. NO. 107-120, at 4 (2001).

137. *Regents of the Univ. of Minnesota v. LSI Corp.*, 926 F.3d 1327, 1339 (Fed. Cir. 2019) (emphasis added).

138. *MCM Portfolio LLC v. Hewlett-Packard Co.*, 812 F.3d 1284, 1291 (Fed. Cir. 2015).

139. *Id.* at 1291–92.

140. 771 F.2d 480 (Fed. Cir. 1985).

141. *Id.* at 1291.

Three years later, in *Saint Regis Mohawk Tribe v. Mylan Pharmaceuticals, Inc.*,¹⁴² the Federal Circuit held that tribal sovereign immunity did not apply in IPR.¹⁴³ The majority noted that “[t]he Tribe acknowledged that sovereign immunity would not apply in ex parte or inter partes reexamination proceedings because of their inquisitorial nature.”¹⁴⁴ In the Federal Circuit’s view, “the mere existence of more inquisitorial proceedings in which immunity does not apply does not mean that immunity applies in a different type of proceeding before the same agency.”¹⁴⁵ This reasoning contradicts the notion that ex parte reexamination and AIA review proceedings should be treated differently merely because one proceeding is adversarial and adjudicatory while the other is not.

Judge Dyk’s concurrence in *Saint Regis Mohawk Tribe* then engaged in a thorough analysis of the similarities between ex parte reexamination and IPR. Judge Dyk similarly focused on the Tribe’s admission that tribal sovereign immunity was inapplicable in reexamination.¹⁴⁶ As in *MCM-Portfolio*, Judge Dyk retold the history of ex parte reexamination, inter partes reexamination, and IPR.¹⁴⁷ After recounting Congress’s goals of increasing confidence in patent quality and creating an efficient alternative to district court litigation, he concluded that, “at its core, [IPR] retains the purpose and many of the procedures of its reexamination ancestors, to which everybody agrees sovereign immunity does not apply.”¹⁴⁸ Because, like ex parte and inter partes reexamination, an IPR proceeding is “fundamentally agency reconsideration, assisted by third parties, rather than agency adjudication of a private dispute,” Judge Dyk concluded that sovereign immunity did not apply in IPR.¹⁴⁹

In *Celgene v. Peter*,¹⁵⁰ the Federal Circuit once more considered the similarities between ex parte reexamination and IPR proceedings. In *Celgene*, the PTAB invalidated a patent’s claims in an IPR.¹⁵¹ The patent owner argued that, because the patent was issued before IPR existed, invalidating the patent’s claims through an IPR was a taking.¹⁵² The Federal Circuit dismissed this reasoning because, given the existence of ex parte reexamination, “[f]or forty

142. *Saint Regis Mohawk Tribe v. Mylan Pharm. Inc.*, 896 F.3d 1322, 1323 (Fed. Cir. 2018), *cert. denied*, 139 S. Ct. 1547 (2019).

143. *Id.* at 1329.

144. *Id.*

145. *Id.*

146. *Id.* at 1330 (Dyk, J., concurring).

147. *Id.* at 1330–34 (Dyk, J., concurring).

148. *Id.* at 1332, 1334–35 (Dyk, J., concurring).

149. *Id.* at 1329–30 (Dyk, J., concurring).

150. *Celgene Corp. v. Peter*, 931 F.3d 1342, 1349 (Fed. Cir. 2019).

151. *See id.*

152. *Id.* at 1358.

years, patents owners have also had the expectation that the PTO could reconsider the validity of issued patents on particular grounds, applying a preponderance of the evidence standard.”¹⁵³ The Federal Circuit also noted that, in *Patlex*, the Court found that no taking occurred, although the PTO, through an ex parte reexamination, cancelled the claims of a patent that was issued before ex parte reexamination existed.¹⁵⁴ The Federal Circuit then reasoned that, although the patentee “identifie[d] a number of differences between reexaminations and IPRs, including that IPRs are adjudicative . . . these differences are not sufficiently substantive or significant to constitute a taking.”¹⁵⁵ The Federal Circuit additionally considered that “IPRs serve essentially the same purpose as their reexamination predecessors,”¹⁵⁶ such as “correcting prior agency error of issuing patents that should not have issued in the first place”¹⁵⁷

Given the shared goals of ex parte reexamination and AIA review proceedings, the Supreme Court in *Return Mail* should have treated them similarly.

b) Furthering the Shared Purposes

Next, it would have made sense for Congress to incorporate MPEP § 2212’s interpretation of persons in ex parte reexamination into the language for AIA review proceedings. Permitting the government to initiate AIA review proceedings furthers the goals behind reexamination and AIA review proceedings.

Involvement from federal agencies would bolster the first purpose, to improve patent quality and confidence in the patent system. AIA review “proceedings are designed to allow the USPTO to harness third parties for the agency to evaluate whether a prior grant of a public franchise was wrong”¹⁵⁸ “Because federal agencies have extensive experience in the patent system, and often have engaged in substantial research and development efforts, they are as well-positioned as private parties to assist the USPTO’s review by bringing to bear relevant information and expertise.”¹⁵⁹ In certain cases, “government agencies often possess . . . the precise prior art, that

153. *Id.* at 1362–63.

154. *Id.* at 1358 n.13.

155. *Id.* at 1360.

156. *Id.*

157. *Id.* at 1361.

158. *Regents of the Univ. of Minnesota v. LSI Corp.*, 926 F.3d 1327, 1339 (Fed. Cir. 2019).

159. Brief for the Respondents at 31, *Return Mail, Inc. v. United States Postal Service*, 139 S. Ct. 1853 (2019).

can help inform the Patent Office's 'second look' at an application."¹⁶⁰ By providing unique prior art and expertise, a government agency can effectively test a patent's validity and potentially remove patent claims that are of low quality.¹⁶¹

The following examples demonstrate that a government agency's expertise can be relevant to an AIA review proceeding, as patented inventions might be necessary to comply with the law. In one example, the Federal Communications Commission (FCC) issued "the E 911 mandate, requiring a specific percentage of wireless phone handsets to have the capability to be located within a certain range of accuracy."¹⁶² The mandate implicated a patent owned by Zoltar, who sued LG and other parties for infringement.¹⁶³ One provider of E911 services claimed it was being sued for patent infringement merely because it was "in compliance with the Commission's E911 regulations"¹⁶⁴ Since the FCC promulgated the E911 mandate, it would likely have possessed expertise concerning such patented technology. Thus, if the FCC participated in an AIA review proceeding against patentees like Zoltar, the FCC could have presented unique arguments that tested the quality of their patents.

As another example, "Section 308 of the Clean Air Act . . . explicitly authorizes the grant of a compulsory license for patents deemed necessary to comply with an EPA standard."¹⁶⁵ Even if the EPA ordered a compulsory license for such a patent, it might still challenge the patent, as some parties that the regulation affects might not be able to afford the compulsory license. Given the EPA's expertise in the matters of the Clean Air Act and related

160. Brief of Amicus Curiae Professor Tejas N. Narechania, *supra* note 49, at 13 (citing *Oil States Energy Servs., LLC v. Greene's Energy Grp., LLC*, 138 S. Ct. 1365, 1374 (2018)).

161. Although the government could litigate the validity of a low quality patent in the Court of Federal Claims, in contrast to a court, the "USPTO is a particularly appropriate venue for making validity determinations in a cost-effective and technically sophisticated environment." Joe Matal, *A Guide to the Legislative History of the America Invents Act: Part II of II*, 21 FED. CIRCUIT B. J. 539, 601 (2012) (citation omitted). Moreover, parties who do not have judicial standing to litigate patent validity in court can still petition for IPR and PGR. *See Cuozzo Speed Techs., LLC v. Lee*, 136 S. Ct. 2131, 2143–44 (2016).

162. *Zoltar Satellite Sys., Inc. v. LG Elecs. Mobile Comm'ns Co.*, 402 F. Supp. 2d 731, 734 n.2 (E.D. Tex. 2005).

163. *See id.* at 733–34, 734 n.2; Tejas N. Narechania, *Patent Conflicts*, 103 GEO. L.J. 1483, 1498–99, 1499 n.80 (2015) (citations omitted).

164. Public Safety and Homeland Security Bureau Seeks Comment on Petition for Declaratory Ruling and/or Rulemaking Filed by Telecommunication Systems, Inc., 28 FCC Rcd. 1435 (2013) (citations omitted).

165. Narechania, *supra* note 163, at 1501.

regulations, it would be able to effectively argue for the invalidation of the patent's relevant claims.

The Department of Justice's (DOJ) decision to initiate an IPR proceeding is yet another example of when a patent might be necessary for legal compliance.¹⁶⁶ Before the IPR, IRIS Corporation Berhad (IRIS), an owner of a patent "covering a method of manufacturing an electronic passport," filed suit against the United States.¹⁶⁷ IRIS alleged that compliance with the Visa Waiver Program and the Enhanced Border Security Act equated to patent infringement.¹⁶⁸ The Secretary of Homeland Security and the Secretary of State were authorized to implement the Visa Waiver Program,¹⁶⁹ and the Department of Homeland Security (DHS) and Department of the Treasury were authorized to implement the Enhanced Border Security Act.¹⁷⁰ After receiving the complaint, the DOJ petitioned for IPR against IRIS's patents.¹⁷¹ Here, multiple agencies possessed expertise relating to the patent since they were responsible for implementing regulations related to the statutes at issue. Therefore, they probably possessed unique prior art and expertise that could have helped the DOJ, which would have helped the PTO better reevaluate the patent's claims.

Allowing the government to initiate AIA review proceedings also furthers the second goal of reducing district court patent litigation. If the government invalidates a patent claim in an AIA review proceeding, district court litigation for that claim generally would disappear.¹⁷² This would allow private parties,

166. *IRIS Corp. v. Japan Airlines Corp.*, 769 F.3d 1359, 1362 (Fed. Cir. 2014) ("JAL cannot comply with its legal obligations without engaging in the allegedly infringing activities.") However, this proceeding was eventually terminated due to a time bar from 35 U.S.C. § 315 (b). *Department of Justice v. IRIS Corporation Berhad*, IPR2016-00497 1, 2 (2017).

167. *Id.* at 1.

168. *See IRIS*, 769 F.3d at 1361.

169. *See* 8 U.S.C. § 1187(a) (2018).

170. 19 C.F.R. § 122.75a(d) implemented the Enhanced Border Security Act. Complaint at 2, *IRIS Corporation Berhad v. United States*, No. 15-175C (Fed. Cl. 2015). This regulation, which can be found in an official copy of the 2017 version of the Code of Federal Regulations, falls under Title 19 Chapter I, which is titled "U.S. Customs and Border Protection, Department of Homeland Security; Department of the Treasury."

171. *IRIS*, IPR2016-00497 at 1.

172. *See Using the US Patent and Trademark Office to Neutralize Invalid Patents*, COOLEYGO, <https://www.cooleygo.com/us-patent-and-trademark-office-neutralize-invalid-patents-ptab-digest/> (last visited Mar. 3, 2020) ("If the USPTO finds the patent to be invalid, then the threat of a patent infringement lawsuit may diminish or even disappear altogether."); *Fresenius USA, Inc. v. Baxter Int'l, Inc.*, 721 F.3d 1330, 1340 (Fed. Cir. 2013) ("[I]n general, when a claim is cancelled, the patentee loses any cause of action based on that claim, and any pending litigation in which the claims are asserted becomes moot."). If the PTAB finds that the patent claim is

who are being sued or who might be later sued for alleged infringement of those specific claims, to avoid “costly and protracted district court litigation.”¹⁷³

The ’548 patent in *Return Mail* shows that private parties have the potential to use the same patented inventions that the government allegedly uses. In *Return Mail*, Return Mail alleged that USPS infringed the ’548 patent.¹⁷⁴ In 2005, Dantom Systems, a private party, requested a declaratory judgment of invalidity for the ’548 patent, to protect its system for processing undeliverable mail.¹⁷⁵ Thus, it is possible that Return Mail might later assert its patent against other private parties. If the Supreme Court had upheld the government-initiated CBM’s holding, that the relevant claims of the ’548 patent were invalid,¹⁷⁶ it could have ended potential district court litigation for these patent claims.

The aforementioned scenario, in which usage of a patented invention might be necessary to comply with the law, directly implicates the second purpose behind AIA review proceedings as well. When a patented invention is potentially required for legal compliance, the government may be in a unique position to best argue against the validity of that patent’s claims. To encourage compliance, the government may attempt to invalidate the relevant claims in an AIA review proceeding. As explained previously, if these patent claims were invalidated in an AIA review proceeding, pending and subsequent district court litigation for those claims generally would disappear.

To further elaborate, if a private party carries out an allegedly infringing activity that is merely helpful for compliance with the law, the patentee can likely sue the private party in district court.¹⁷⁷ When the allegedly infringing

not invalid, then the patent would have withstood scrutiny, thus ensuring its quality and increasing confidence in the patent system.

Although an invalidity defense or request for declaratory judgement in court might also further the second purpose, PGR and IPR allow parties to initiate review proceedings even if they do not have constitutional standing to litigate patent validity in court. *See* *Cuozzo Speed Techs., LLC v. Lee*, 136 S. Ct. 2131, 2143–44 (2016). In addition, a patent’s claims will more likely be invalidated in an AIA review proceeding than in court because of the lower preponderance of the evidence standard. *See* 35 U.S.C. §§ 316(e), 326(e) (2018).

173. *See* H.R. REP. NO. 112-98, at 45 (2011).

174. *Return Mail, Inc. v. United States Postal Serv.*, 139 S. Ct. 1853, 1866 (2019).

175. *See* Complaint at 2–3, *Dantom Systems, Inc. v. Return Mail, Inc.*, No. 05-CV-70956 (E.D.M.I. 2005).

176. *See* *United States Postal Service and United States v. Return Mail, Inc.*, CBM2014-00116 1, 3 (2014).

177. To explain, one requirement for a private party to be immunized from a patent infringement suit under 28 U.S.C. § 1498 is that the government provides authorization and consent to carry out the allegedly infringing activity. *IRIS Corp. v. Japan Airlines Corp.*, 769 F.3d 1359, 1361–62 (Fed. Cir. 2014). The government does not provide authorization and

activity is required for legal compliance, as long as the activity was not performed “for the benefit of the government,”¹⁷⁸ then the patent owners can also potentially sue the private parties who are complying with the law in district court.¹⁷⁹

In sum, the prior construction canon dictates that Congress intended to incorporate MPEP § 2212 into the statutory language governing AIA review proceedings. Still, the Court’s other main arguments must be addressed.

B. IMPROPERLY PENALIZING THE GOVERNMENT FOR ITS SPECIAL STATUS UNDER 28 U.S.C. § 1498

The Court seemed especially concerned with how the government can only be sued for patent infringement in the Court of Federal Claims, where there is no jury trial, and that the government does not face a threat of injunction or punitive damages. According to the Court, Congress could have intended to foreclose the government from petitioning for AIA review proceedings because the government already received special treatment under 35 U.S.C. § 1498.¹⁸⁰ But this view is problematic for two reasons: (1) section 1498 is justified under sovereign immunity, eminent domain, and national security concerns, and (2) permitting government agencies to petition for AIA review proceedings would not chill inventors from obtaining patents.

1. *Theories That Justify 28 U.S.C. § 1498*

The Court appeared to be concerned about fairness when it stated that, “because federal agencies face lower risks, it is reasonable for Congress to have treated them differently.”¹⁸¹ But the government is entitled to special treatment under 28 U.S.C. § 1498 based on sovereign immunity, eminent domain, and national security concerns.

consent, expressly or impliedly, when the allegedly infringing activity is not necessary for compliance with the government’s requirements. *See* Severson Envtl. Servs., Inc. v. Shaw Envtl., Inc., 477 F.3d 1361, 1366–67 (Fed. Cir. 2007); *Madey v. Duke Univ.*, 413 F. Supp. 2d 601, 609 (M.D.N.C. 2006) (citations omitted). *But see* TVI Energy Corp. v. Blane, 806 F.2d 1057, 1060 (Fed. Cir. 1986) (explaining that the “mere fact that the Government specifications . . . did not absolutely require [infringement] does not extinguish the Government’s consent”).

178. *IRIS Corp. v. Japan Airlines Corp.*, 769 F.3d 1359, 1362 (Fed. Cir. 2014) (quoting *Advanced Software Design Corp. v. Fed. Reserve Bank of St. Louis*, 583 F.3d 1371, 1378 (Fed. Cir. 2009)). If the parties engaged in the allegedly infringing activity for the benefit of the government, then 28 U.S.C. § 1498(a) would dictate that the parties were immune from the suit, and only the government would be able to be sued for patent infringement. *Id.*

179. *See id.* at 1361–62 (citations omitted).

180. *Return Mail, Inc. v. United States Postal Serv.*, 139 S. Ct. 1853, 1867 (2019).

181. *Id.*

Notably, “[s]ection 1498 waives the United States’ sovereign immunity from suit.”¹⁸² Without a formal waiver of sovereign immunity, claiming patent infringement against the government would be difficult, as the relevant law prior to the formal waiver was filled with inconsistencies and inequities.¹⁸³ Before Congress created section 1498’s precursor¹⁸⁴ in the Act of 1910,¹⁸⁵ the only method to deal with governmental patent infringement was congressional reference or implied-in-fact contract actions in court.¹⁸⁶ Generally, “the Court of Claims possessed no jurisdiction over patent infringement actions before the statutory authorization of such jurisdiction.”¹⁸⁷

Furthermore, section 1498 is founded on an eminent domain theory and is thus consistent with the Fifth Amendment.¹⁸⁸ As the Federal Circuit has acknowledged, “[s]ection 1498(a) ‘is an eminent domain statute,’ wherein the government ‘has consented thereunder only to be sued for its taking of a patent license.’”¹⁸⁹ Thus, under the theory of eminent domain, the government should still be allowed to use property, as long as it pays, since private property “shall [not] be taken for public use, without just compensation.”¹⁹⁰ The idea makes sense practically, as the government may need the patent for important reasons like preserving national security.

Indeed, if the government or its contractors faced the threat of injunction, then the government would be unable to provide for national security effectively in the case that a patented invention was necessary to further that goal. As an example, the Act of 1910, an ancestor of section 1498, stated that the government could only be sued in the Court of Claims for patent infringement.¹⁹¹ The Supreme Court held that this did not cover contractors for the government, some of which were “in the course of producing warships

182. *Zoltek Corp. v. U.S.*, 672 F.3d 1309, 1318 (Fed. Cir. 2012).

183. Lionel M. Lavenue, *Patent Infringement Against the United States and Government Contractors Under 28 U.S.C. § 1498 in the United States Court of Federal Claims*, 2 GA. LAW REV. 389, 411 (1995).

184. *Zoltek*, 672 F.3d at 1315.

185. Lavenue, *supra* note 183, at 411.

186. *Id.* at 408.

187. *Id.* at 411 (citing Louis H. Le Mieux, *Patent Jurisdiction of the Court of Claims*, 41 J. PAT. OFF. SOC’Y 112, 114 (1959)).

188. *Return Mail, Inc. v. United States Postal Serv.*, 868 F.3d 1350, 1361 (Fed. Cir. 2017), (“It is settled that recovery of reasonable compensation under § 1498 is premised on a theory of an eminent domain taking under the Fifth Amendment.”) (quoting *Tektronix, Inc. v. United States*, 552 F.2d 343, 346 (Ct. Cl. 1977)).

189. *Id.* (quoting *Decca Ltd. v. United States*, 640 F.2d 1156, 1167 (Ct. Cl. 1980)).

190. U.S. CONST. amend. V.

191. *Zoltek Corp. v. United States*, 672 F.3d 1309, 1315 (Fed. Cir. 2012).

during World War I.”¹⁹² This was problematic because contractors were “exposed to expensive litigation, involving the *possibilities of prohibitive injunction* [sic] payment of royalties, rendering of accounts, and payment of punitive damages, and they [were] reluctant to take contracts that may bring such severe consequences.”¹⁹³ As a result, Congress amended the statute so that government manufacturers and contractors could be immunized from patent infringement suits.¹⁹⁴ In this way, the government was not deprived of a necessary patented invention, and the public ultimately benefited from this.

Because the government’s special status is justified, it should not be penalized for its treatment under section 1498.

2. *Allaying Potential Chilling Effect Concerns*

The Court could also have been concerned that by providing the government with multiple advantageous avenues, it might chill innovators from creating inventions that would mostly be used by the government. After all, not only would the government have the advantages that come with being sued in the Court of Federal Claims, but it also would have access to AIA review proceedings that do not involve a presumption of validity and use the lower preponderance of the evidence standard.¹⁹⁵

But a chilling effect already exists for such inventors. Since the government cannot be forced to pay punitive damages or be enjoined from using a patented invention under section 1498, these inventors might already be discouraged from innovation. Also, the existence of secrecy orders, which qualifying defense agencies can request on patent applications,¹⁹⁶ adds to the potential effect of chilling innovation, and yet they are allowed for purposes of national security. Indeed, some patent applications have been placed under a secrecy order long enough to diminish their value,¹⁹⁷ and obtaining compensation is not an easy task.¹⁹⁸ Like secrecy orders, which are permitted for national security reasons, the government should also be permitted to initiate AIA

192. *Id.* (citing *William Cramp & Sons Ship & Engine Bldg. Co. v. Int’l Curtis Marine Turbine Co.*, 246 U.S. 28 (1918)).

193. *Id.* at 1316 (quoting *Wood v. Atl. Gulf & Pac. Co.*, 296 F. 718, 720–21 (S.D. Ala. 1924)) (emphasis added).

194. *Id.*

195. *See* 35 U.S.C. §§ 316(e), 326(e). Innovators that create products that the government *and* others will use would not cause a chilling effect. The innovator could still enforce injunctions against and obtain punitive damages from those nongovernmental users.

196. 35 U.S.C. § 181 (2018).

197. G.W. Schulz, *Government Secrecy Orders on Patents Have Stifled More Than 5,000 Inventions*, WIRED (April 16, 2013), <https://www.wired.com/2013/04/gov-secrecy-orders-on-patents/>.

198. *See* Sabling H. Lee, *Protecting the Private Inventor Under the Peacetime Provisions of the Invention Secrecy Act*, 12 BERKELEY TECH. L.J. 345, 375 (1997).

review proceedings, since providing the government with that ability specifically furthers the purposes of the AIA.¹⁹⁹

Furthermore, government agencies would suffer the most from a lack of innovation if they contributed to the chilling effect. An overwhelming chilling effect would prevent inventors from innovating products that would benefit the government, which would not want to lose access to necessary inventions.²⁰⁰ Therefore, it is likely that the government would only petition for AIA review proceedings at a reasonable rate. Indeed, the government has only petitioned for AIA review proceedings about twenty times.²⁰¹

C. PERMITTING GOVERNMENTAL PETITIONS IS NOT AWKWARD

In *Return Mail*, the Court was also concerned that it would be awkward for a civilian patent owner to defend its patent in a proceeding initiated by one government agency and overseen by another government agency.²⁰² The Court's underlying concerns amount to the following: (1) the President should be resolving disputes between governmental agencies, and (2) there is a danger that the PTO will be biased towards a fellow government agency. Neither of these arguments withstand scrutiny, especially because similar “awkward” situations occur in other agency proceedings.

1. *Nonexistent Executive Agency Conflicts*

The first argument (made in an amicus brief), that the President should be resolving conflicts between executive agencies,²⁰³ is not persuasive. This argument assumes that permitting government agencies to initiate AIA review proceedings would lead to a dispute between two government agencies—the PTO and the government agency petitioner. But conflict would only exist between either the government agency and the civilian patent owner, or between the PTO and the patentee. Regarding the former, the patent owner

199. *See supra* Section IV.A.2.b. The government's ability to petition for AIA review proceedings also plays a similar role in preserving national security. *Infra* Section IV.D.2.

200. During the hearings for amending 28 U.S.C. § 1498, the following was stated: “This bill . . . ultimately serves the interests of the U.S. Government. Without this bill, companies have little incentive to spend their intellectual resources to help the Government solve its technical problems. As a member of the National Security Committee, I am well aware of some of the circumstances where companies can help us solve technical problems and thus add to our military capabilities, and this bill will be of great help in that regard.” 141 CONG. REC. H14319 (daily ed. Dec. 12, 1995) (statement of Senator Schroeder) (emphasis added).

201. *See* Transcript of Oral Argument at 61, *Return Mail, Inc. v. United States Postal Serv.*, 139 S. Ct. 1853 (2019) (No. 17-1594) [hereinafter Transcript of Oral Argument, *Return Mail*].

202. *See Return Mail, Inc. v. United States Postal Serv.*, 139 S. Ct. 1853, 1867 (2019).

203. *See* Brief for Amicus Curiae Pharmaceutical Research and Manufacturers of America in Support of Petitioner at 13–14, *Return Mail, Inc. v. United States Postal Service*, 139 S. Ct. 1853 (2019) (No. 17-1594).

would be defending its patent from the government agency petitioner.²⁰⁴ Consistent with precedent, one could also describe the proceeding as taking place between the PTO and the patent owner.²⁰⁵ Either way, no conflict exists between the PTO and the government agency petitioner.

There is the counterargument that, if the PTAB rules against the government agency, the PTAB has created a conflict or dispute. Even assuming this is true, as of now, government agencies that own patents can still be forced into AIA review proceedings.²⁰⁶ When a private party petitions for an AIA review proceeding against the government, and the PTAB holds that the government's patent claims are invalid, a conflict between the two agencies would also arguably exist (at least using the above logic). In order to remain consistent, the Court would need to prevent civilian patent owners from petitioning for AIA review proceedings against the government as well.

2. *Bias Does Not Necessarily Exist*

The second potential argument, that one government agency may be inclined to give more deference to a fellow government agency, also fails. The PTO has demonstrated that it will not offer an agency special treatment merely because it belongs to the same branch of government.

For instance, the PTO and the Federal Trade Commission (FTC) disagreed about whether a patentee had “permission to use her patent to engage in anticompetitive behavior, so long as such behavior is within the patent’s scope.”²⁰⁷ The PTO believed that “‘suspicion’ of patent rights by ‘competition regulators’ would ‘interfer[e] with these market-based incentives to innovation.’”²⁰⁸ By contrast, the FTC thought the PTO would “‘benefit from much greater consideration and incorporation of economic insights’ in their attempts ‘to find the proper balance between patent and competition law.’”²⁰⁹ Notably, the agencies did not come to a consensus; it was the Court’s decision in *Federal Trade Commission v. Actavis, Inc.* that put this matter to rest.²¹⁰

204. This argument might be more persuasive if the Supreme Court held that government agencies could not petition for AIA review proceedings against one another.

205. *See Regents of the Univ. of Minnesota v. LSI Corp.*, 926 F.3d 1327, 1340, 1340 n.19 (Fed. Cir. 2019) (quoting *Oil States Energy Servs., LLC v. Greene’s Energy Grp., LLC*, 138 S. Ct. 1365, 1373 (2018)) (explaining that “IPR is a proceeding between the United States and the patent owner, rather than adjudication between two private parties”).

206. *Return Mail, Inc. v. United States Postal Serv.*, 139 S. Ct. 1853, 1871–72 (2019) (Breyer, J., dissenting).

207. Narechania, *supra* note 163, at 1489–90.

208. *Id.* at 1493 (alteration in original) (citations omitted).

209. *Id.* (citations omitted).

210. *Id.* at 1495–96.

The National Institute of Health (NIH) and the PTO also disagreed about the patentability of expressed sequence tags (ESTs), a type of gene fragments.²¹¹ The NIH was concerned that patenting ESTs would stifle innovation by tying up a significant portion of upstream genomic research.²¹² The PTO disagreed and, instead, lowered the utility threshold for patent applications involving ESTs.²¹³ Although the PTO eventually amended its policy after dialogue with the NIH, the PTO refused to fully concede to the NIH's requests.²¹⁴

3. *Similar Proceedings*

Finally, it is not uncommon to have one agency adjudicate a proceeding that another government agency initiates against a civilian.²¹⁵ For example, in immigration courts, the DHS initiates removal proceedings against a person by filing a notice to appear.²¹⁶ The DOJ, through its Executive Office for Immigration Review, oversees the proceeding.²¹⁷

Congress has permitted similar adversarial proceedings in the patent system. Under 35 U.S.C. § 181, at the request of a qualifying defense agency, the Commissioner of Patents can place a secrecy order on a patent application and seal it from public disclosure.²¹⁸ When a secrecy order is imposed, an application owner can petition to rescind the secrecy order.²¹⁹ If the petition is denied, the application owner can then appeal the decision to maintain the secrecy order.²²⁰ In the appeal, the Secretary of Commerce, who is the head of the Department of Commerce, oversees an adversarial-like proceeding

211. *See id.* at 1503.

212. *Id.*

213. *See id.*

214. *See id.*

215. Brief of Amicus Curiae Professor Tejas N. Narechania, *supra* note 49, at 21.

216. *See* 5 CHARLES GORDON, STANLEY MAILMAN, STEPHEN YALE-LOEHR & RONALD Y. WADA, IMMIGRATION LAW AND PROCEDURE § 64.02 (2019).

217. *See Immigration Benefits in EOIR Removal Proceedings*, U.S. CITIZENSHIP AND IMMIGRATION SERVICES (Dec. 6, 2019), <https://www.uscis.gov/laws/immigration-benefits-eoir-removal-proceedings>.

218. 35 U.S.C. § 181 (2018). Under 35 U.S.C. § 181, for patent applications that the government does not have a property interest in, if the Commissioner of Patents believes the invention might implicate national security, the Commissioner “shall” notify a potentially interested defense agency about the patent. If the head of the defense agency believes that the public disclosure of the application could be detrimental to national interests, the Commissioner of Patents “shall” place a secrecy order on the application. *Id.*

219. *Id.*; 37 C.F.R. §§ 5.4(a)–(c) (2019).

220. 35 U.S.C. § 181; 37 C.F.R. § 5.4(d) (2019).

involving the application owner and the defense agency that requested the secrecy order.²²¹

Proceedings at the ITC raise even more questions about the reasoning in *Return Mail*.²²² The ITC oversees arguments brought by three parties; like AIA review proceedings, two parties are in an adversarial relationship.²²³ But there is also a third party that is represented by an Investigative Attorney, who belongs to the ITC's Office of Unfair Import Investigation and represents the public interest.²²⁴ The opposing parties usually attempt to convince the Investigative Attorney to support their positions.²²⁵ Here, one government agency oversees a proceeding in which the same government agency represents one of the three parties. This creates a greater risk of bias and awkwardness than in *Return Mail*, as the ITC might find it difficult not to defer to its Investigative Attorney, at least based on the Court's reasoning. To be consistent, the Court would need to hold that the above situations are just as awkward, if not even more so, than the situation in *Return Mail*.

D. DELIBERATELY OMITTING ACTIONS IN THE COURT OF FEDERAL CLAIMS FROM STATUTORY ESTOPPEL

Judge Newman's dissent at the Federal Circuit argued that, because Congress did not mention 28 U.S.C. § 1498 in the statutory estoppel provisions for AIA review proceedings, Congress never intended for government entities to be able to petition for AIA review proceedings.²²⁶ Court of Federal Claims actions from 28 U.S.C. § 1498 are the only method a party can use to sue the government for patent infringement. But the statutory estoppel provisions only mentioned actions that occur in district court and before the ITC.²²⁷ Judge Newman's position is admittedly more persuasive than the Court's reasoning, but it still fails for two reasons: (1) common law estoppel would likely still

221. See 37 C.F.R. § 5.4(d) (2019).

222. The ITC can oversee certain patent infringement issues under 19 U.S.C. § 1337 (a)(1)(B) (2018).

223. William P. Atkins & Justin A. Pan, *An Updated Primer on Procedures and Rules in 337 Investigations at the U.S. International Trade Commission*, 18 U. BALT. INTEL. PROP. L.J. 105, 112 (2010) (explaining that ITC proceedings are "typically based on a complaint filed by a private party" and "resemble[] private litigation").

224. *Id.* at 116.

225. *An Introduction to Section 337 Proceedings at the ITC*, MORRISON & FOERSTER 1, 2 n.4, <https://mofoaitc.mofo.com/wp-content/uploads/2015/08/Intro337ProceedingsITC.pdf> (last visited Mar. 3, 2020).

226. See *Return Mail, Inc. v. U.S. Postal Serv.*, 868 F.3d 1350, 1373–74 (Fed. Cir. 2017) (Newman, J., dissenting). The Supreme Court did not embrace Judge Newman's argument. *Return Mail, Inc. v. United States Postal Service*, 139 S. Ct. 1853, 1876 n.10 (2019).

227. 35 U.S.C. §§ 315(a), 325(e) (2018); AIA § 18(a)(1)(D).

apply in section 1498 suits, and (2) Congress had good reason to subject the government only to common law estoppel.

As a reminder, the statutory estoppel provisions for IPR and PGR preclude a petitioner from raising grounds in district court or ITC proceedings that it “raised or reasonably could have raised” in a prior IPR or PGR.²²⁸ The provision for CBM only precluded raising arguments that were raised in a CBM, as opposed to arguments that reasonably could have been raised.²²⁹

1. *Common Law Estoppel Is Applicable to the Government*

As the Court pointed out in *Return Mail*, common law estoppel in the form of mutual collateral estoppel could apply in section 1498 suits after an AIA review proceeding.²³⁰ Mutual collateral estoppel, or issue preclusion, prevents a party from relitigating issues “[w]hen an issue of fact or law is *actually litigated and determined* by a valid and final judgment, and the determination is essential to the judgment”²³¹ Under mutual collateral estoppel, even “the government may be estopped under certain circumstances from relitigating a question when the parties to the two lawsuits are the same.”²³²

For arguments litigated and decided in administrative proceedings, precedent requires that courts presume Congress intended collateral estoppel to preclude those arguments from being made again in litigation. In *Astoria Federal Savings & Loan Association v. Solimino*, the Supreme Court stated that “[w]e have long favored application of the common-law doctrines of collateral estoppel (as to issues) . . . to those determinations of administrative bodies that have attained finality.”²³³ Thus, a presumption exists that “Congress has legislated with an expectation that the principle will apply except ‘when a statutory purpose to the contrary is evident.’”²³⁴

For the situation in *Return Mail*, no evidence exists that the purposes of increasing patent quality and creating an effective and efficient alternative to litigation contradict applying collateral estoppel to government agencies in the Court of Federal Claims, subsequent to an AIA review proceeding. Indeed, applying collateral estoppel would be consistent with creating a “streamlined”

228. *Id.* §§ 315(a), 325(e).

229. AIA § 18(a)(1)(D).

230. *See Return Mail*, 139 S. Ct. at 1867 n.10 (noting that “this Court has not decided whether common-law estoppel applies in § 1498 suits”).

231. *B & B Hardware, Inc. v. Hargis Indus., Inc.*, 135 S. Ct. 1293, 1303 (2015) (quoting Restatement (Second) of Judgments § 27, p. 250 (1980)) (emphasis added).

232. *United States v. Mendoza*, 464 U.S. 154, 163 (1984) (citations omitted); *see also* Brief for the Respondents, *Return Mail*, *supra* note 159, at 41 (quoting *Mendoza*, 464 U.S. at 163).

233. *Astoria Fed. Sav. & Loan Ass’n v. Solimino*, 501 U.S. 104, 107 (1991).

234. *Id.* at 108 (1991) (quoting *Isbrandtsen Co. v. Johnson*, 343 U.S. 779, 783 (1952)).

patent system.²³⁵ Moreover, the elements of issue preclusion would be met for arguments made in AIA proceedings. AIA proceedings result in “final written decision[s]” that determine issues of validity that were actually litigated and essential to the decision.²³⁶ Thus, the collateral estoppel presumption, if applied in *Return Mail*, would have likely withstood rebuttal.

In fact, the Supreme Court, when deciding the relation of estoppel to another proceeding at the PTO that lacked statutory estoppel provisions, applied this presumption.²³⁷ In *B & B Hardware, Inc. v. Hargis Industries Inc.*, the Court held that, for PTO trademark proceedings overseen by the Trademark Trial and Appeal Board (TTAB), “a court should give preclusive effect to TTAB decisions if the ordinary elements of issue preclusion are met.”²³⁸

If collateral estoppel applied in the Court of Federal Claims, it would likely apply only to invalidity doctrines that were litigated and determined.²³⁹ For instance, if the petitioner in an AIA review proceeding argued that there was invalidity based on obviousness²⁴⁰ but did not raise any other invalidity contentions (such as novelty) in that proceeding, the petitioner would be able to assert the unraised invalidity contentions in subsequent litigation.

There still remains the question of why Congress would want the government to only be subject to the more lenient mutual collateral estoppel instead of the AIA’s stricter statutory estoppel provisions.

2. *Good Reasons for Excluding the Government from Statutory Estoppel*

The statutory estoppel provisions are stricter than collateral estoppel.²⁴¹ Unlike statutory estoppel for IPR and PGR, which applies to “grounds” that

235. See H.R. REP. NO. 112-98, at 40.

236. 35 U.S.C. §§ 318(a), 328(a) (2018), AIA § 18(D).

237. See Brief for the Respondents, *Return Mail*, *supra* note 159, at 41 (citing *B & B Hardware, Inc. v. Hargis Indus., Inc.*, 135 S. Ct. 1293 (2015)) (explaining that “decisions of the USPTO’s Trademark Trial and Appellate Board have such preclusive effect”).

238. *B & B Hardware*, 135 S. Ct. at 1299.

239. See *What’s Next for Issue Preclusion and Patent Invalidity*, KIRKLAND & ELLIS (June 28, 2018), <https://www.kirkland.com/publications/article/2018/06/whats-next-for-issue-preclusion-and-patent-invalid> (“[I]nvalidity is not a single ‘issue’ for purposes of collateral estoppel. Under the court’s rationale, a § 101 theory is at least distinct from a § 102 theory or a § 112 theory (for example).”).

240. Collateral estoppel likely would not preclude obviousness arguments based on prior art that had not been previously raised in an AIA review proceeding. See *id.* (arguing that collateral estoppel should only apply to obviousness or novelty arguments in relation to the specific prior art that was brought up in the prior proceeding).

241. See Reply Brief for Petitioner at 21, *Return Mail, Inc. v. United States Postal Service*, 139 S. Ct. 1853 (2019) (stating that collateral estoppel is “a different substantive standard” than the standards in the AIA’s statutory estoppel provisions).

were or could reasonably have been raised in the previous proceedings,²⁴² collateral estoppel only prevents relitigating issues that were “actually litigated and determined” in a previous decision.²⁴³ The statutory estoppel provision for CBM, which only applied to grounds that were “raised,” is also a stricter standard than “actually litigated and determined.” For instance, “[a]n issue that [is] raised but abandoned [is] not actually litigated for purposes of issue preclusion” and can thus be brought up in a later proceeding.²⁴⁴ Nevertheless, Congress had good reasons for intending *only* mutual collateral estoppel to apply to the government. Two points demonstrate this.

First, Congress probably did not believe the government would use proceedings and suits to harass patentees. Congress enacted statutory estoppel so that AIA review proceedings would not “be used as tools for harassment or a means to prevent market entry through repeated litigation and administrative attacks on the validity of a patent.”²⁴⁵ However, the government differs from private parties, who may be more prone to abusing proceedings and suits.²⁴⁶ The government has especially “limited resources” devoted to suits and similar proceedings, and it spends them on issues “frequently involving legal questions of substantial importance.”²⁴⁷ “It is not open to serious dispute that the Government is a party to a far greater number of cases on a nationwide basis than even the most litigious private entity . . .”²⁴⁸ With its limited resources, much of which the government spends on important public policy issues, Congress could have reasonably concluded that the government “did not require additional deterrence to avoid wasting resources on duplicative validity proceedings.”²⁴⁹ Indeed, the government has only petitioned for AIA review proceedings about twenty times.²⁵⁰

242. See 35 U.S.C. §§ 315(e)(2), 325(e)(2).

243. Brief for the Respondents, *Return Mail*, 139 S. Ct. 1853, at 42 (quoting Restatement (Second) of Judgments Section 27 (1982)).

244. 18 JAMES WILLIAM MOORE, MOORE’S FEDERAL PRACTICE—CIVIL § 132.03(2)(e) (2019) (alteration in original, except for the first alteration).

245. H.R. REP. NO. 112-98, at 48 (2011); see also Motion for Leave to File Brief of Amicus Curiae Electronic Frontier Foundation and Brief of Amicus Curiae in Support of Respondent at 13, *Return Mail*, 139 S. Ct. 1853 (No. 17-1594); Brief of the R Street Institute as Amicus Curiae in Support of Respondents at 9, *Return Mail*, 139 S. Ct. 1853 (No. 17-1594).

246. See Motion for Leave to File Brief of Amicus Curiae Electronic Frontier Foundation, *supra* note 245, at 13 (discussing “potential harassment of patent owners by *private* litigants”) (emphasis added).

247. See *United States v. Mendoza*, 464 U.S. 154, 155 (1984).

248. *Id.* at 159.

249. Motion for Leave to File Brief of Amicus Curiae Electronic Frontier Foundation, *supra* note 245, at 13.

250. See Transcript of Oral Argument, *Return Mail*, *supra* note 201, at 61.

Second, Congress had good reasons to subject the government to less restrictive estoppel, as patents can implicate national security. One patent, which had some of its claims cancelled in an IPR proceeding initiated by the DHS,²⁵¹ “describes detector units housed within ‘products’ to ‘prevent[] terrorist activity’ ”²⁵² In addition, the patent in *IRIS* implicated border concerns, as *IRIS* alleged that using electronic passports equated to patent infringement.²⁵³

With these types of patents in mind, Congress might have imagined that the following undesirable scenario could occur if statutory estoppel applied to the government. In this hypothetical, patent claims that implicate national security should have been invalidated in an AIA review proceeding, but they still exist because the government inadvertently omitted the best arguments in a prior AIA review proceeding. The Court of Federal Claims subsequently holds that the government infringed the patent and forces the government to pay license fees for a problematic patent. In this case, the government would find it more difficult to preserve national security, as it would need to pay license fees for patent claims that should have been invalidated.²⁵⁴

Furthermore, treating the government differently in the patent system due to national security concerns is not novel. Section IV.B.2 discussed Congress’s decision to allow government agencies to request secrecy orders on patents that implicate national security. Section IV.B.1 also explained how Congress immunized government contractors from patent infringement suits; in this way, the government had access to warships during WWI.

In sum, Judge Newman’s argument is problematic. There could be multiple explanations for why the statutory estoppel provisions did not mention the Court of Federal Claims. Judge Newman’s explanation is that Congress did not intend for government agencies to petition for AIA review proceedings. But it is also plausible that Congress intentionally omitted the government from the statutory estoppel provisions, or that Congress merely omitted the government in its provision through an inadvertent drafting error. One possible explanation should not overcome Congress’s intent to

251. *See* United States Department of Homeland Security v. Larry Golden, IPR2014-00714 1, 2–3 (2015).

252. Petition for Inter Partes Review of U.S. Patent No. RE43,990 at 4, United States Department of Homeland Security v. Larry Golden, IPR2014-00714 (2015) (citing the ’990 patent col. 3 l. 17–22, col. 4 l. 32–38).

253. *IRIS Corp. v. Japan Airlines Corp.*, 769 F.3d 1359, 1361–62 (Fed. Cir. 2014).

254. A similar hypothetical could apply to explain the difference between the “raised” standard for CBM and the “actually litigated and determined” standard for mutual collateral estoppel. The CBM hypothetical would replace the government’s inadvertent omission with a decision to abandon an argument that, at the time, might not have seemed to be helpful.

incorporate MPEP § 2212's interpretation of "person" for ex parte reexamination into the meaning of "person" for AIA review proceedings.

V. SOLUTION

The holding in *Return Mail* requires change because the Court's decision was incorrect and conflicts with the AIA's goals. Through its procedural rulemaking authority, the PTO can and should create a rule stating that a government agency qualifies as a "person" who can petition for AIA review proceedings. Specifically, the PTO could adopt the following language: "governmental entities are included within the scope of the term '[a] person.'" ²⁵⁵

If the PTO promulgates this rule, it will likely be challenged in light of *Return Mail*. However, under *Brand X*, the PTO's interpretation could become controlling, as an agency interpretation can trump a court's statutory construction. ²⁵⁶ Courts look to the *Chevron* framework when applying *Brand X*. ²⁵⁷ *Chevron* step zero asks whether Congress intended to give the agency the relevant rulemaking authority and whether the agency properly promulgated the rule. ²⁵⁸ If step zero is met, *Chevron* step one then asks "whether Congress has directly spoken to the precise question at issue. If the intent of Congress is clear, that is the end of the matter; for the court, as well as the agency, must give effect to the unambiguously expressed intent of Congress." ²⁵⁹ If Congress's intent is ambiguous, *Chevron* step two then asks "whether the agency's answer is based on a permissible construction of the statute." ²⁶⁰ Here, the proposed rule would likely satisfy the *Chevron* framework.

A. DELEGATED RULEMAKING AUTHORITY

Chevron deference is applicable when "Congress delegated authority to the agency generally to make rules carrying the force of law, and . . . the agency interpretation claiming deference was promulgated in the exercise of that authority." ²⁶¹ The proposed rule would likely meet both requirements.

255. See MPEP § 2212 (9th ed. Rev. June 2020).

256. *Nat'l Cable & Telecommunications Ass'n v. Brand X Internet Servs.*, 545 U.S. 967, 982 (2005).

257. *Id.* at 980, 982 ("This principle follows from *Chevron* itself.")

258. See *United States v. Mead Corp.*, 533 U.S. 218, 226–27 (2001).

259. *Chevron, U.S.A., Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837, 842–43 (1984).

260. *Id.* at 843.

261. *United States v. Mead Corp.*, 533 U.S. 218, 226–27 (2001).

1. *Procedural Rulemaking Authority*

Congress delegated to the PTO the requisite rulemaking authority to promulgate the proposed rule. Under 35 U.S.C. § 2(2)(A)(2018), the PTO “may establish regulations, not inconsistent with law, which shall govern the conduct of proceedings in the Office.” The Federal Circuit held that this provides the PTO with procedural rulemaking authority for its proceedings,²⁶² as long as it is also interpretative, or “a prospective clarification of ambiguous statutory language regarding a matter of procedure.”²⁶³ Importantly, the Federal Circuit has applied *Chevron* deference to procedural rules from the PTO that clarify ambiguous statutory language.²⁶⁴ The Supreme Court also found that the AIA provides additional rulemaking authority to the PTO for IPR under 35 U.S.C. § 316(a)(4) (2018),²⁶⁵ which states that “the Director shall prescribe regulations establishing and governing inter partes review under this chapter and the relationship of such review to other proceedings under this title.”²⁶⁶

The proposed rule, which would state that federal agencies can petition for AIA review proceedings, qualifies as a procedural rule that clarifies ambiguous statutory language. In *Tafas v. Doll*,²⁶⁷ the Federal Circuit looked to a D.C.

262. *See* *Cooper Techs. Co. v. Dudas*, 536 F.3d 1330, 1335 (Fed. Cir. 2008) (“To comply with section 2(b)(2)(A), a Patent Office rule must be ‘procedural’—i.e., it must ‘govern the conduct of proceedings in the Office.’”) (quoting 35 U.S.C. § 2(b)(2)(A) (2006)); Michaels, Andrew, *How Much Deference Courts Owe to USPTO Guidance* (June 20, 2019), Law 360, <https://www.law360.com/articles/1171217/how-much-deference-courts-owe-to-uspto-guidance> (writing that “the USPTO has been delegated the authority to speak with the force of law (and thus does qualify for *Chevron* deference) on certain procedural matters before the office”).

263. *Cooper Techs. Co. v. Dudas*, 536 F.3d 1330, 1336 (Fed. Cir. 2008).

264. *Id.* at 1337 (“[B]ecause the Patent Office is specifically charged with administering statutory provisions relating to ‘the conduct of proceedings in the Office,’ we give *Chevron* deference to its interpretations of those provisions.”)(quoting 35 U.S.C. § 2(a)(2)(A)). The Federal Circuit reaffirmed this when explaining that it “review[s] the PTO’s regulations and statutory interpretation pursuant to *Chevron*” *Aqua Prod., Inc. v. Matal*, 872 F.3d 1290, 1302 (Fed. Cir. 2017) (plurality opinion).

265. *Cuozzo Speed Techs., LLC v. Lee*, 136 S. Ct. 2131, 2142 (2016). A similar provision exists for PGR: “The Director shall prescribe regulations establishing and governing a post-grant review under this chapter and the relationship of such review to other proceedings under this title.” 35 U.S.C. § 326(a)(4) (2018).

266. The Supreme Court has hinted that the AIA’s additional grant of authority gives the PTO rulemaking authority that extends beyond promulgating procedural rules, at least for AIA review proceedings. *See* *Cuozzo Speed Techs., LLC v. Lee*, 136 S. Ct. 2131, 2135 (2016).

267. *Tafas v. Doll*, 559 F.3d 1345 (Fed. Cir. 2009), *vacated*, 328 F. App’x 658 (Fed. Cir. 2009). This appeal was vacated. Before the rehearing occurred, the PTO resolved the case by removing the rules at issue. Sarah Tran, *Administrative Law, Patents, and Distorted Rules*, 80 GEO. WASH. L. REV. 831, 854 (2012) (citing *Tafas v. Kappos*, 586 F.3d 1369, 1371 (Fed. Cir. 2009) (en banc)). Still, the case is at least persuasive authority.

Circuit case, *JEM Broadcasting Co. v. FCC*, to distinguish between procedural and substantive rules.²⁶⁸ In *JEM*, the D.C. Circuit held that certain rules were procedural because they did not “change the *substantive standards* by which the FCC evaluates license applications.”²⁶⁹ As an example of a procedural matter, in *Fund Democracy, LLC v. S.E.C.*, the D.C. Circuit held that providing a person with the ability to request a hearing before the SEC was a “grant of a procedural right”²⁷⁰

Here, like in *JEM*, participation in AIA review proceedings only affects procedural matters and does not affect the proceedings’ substantive standards, such as preponderance of the evidence. Based on the reasoning in *Fund Democracy*, the ability to initiate PTO AIA review proceedings is a procedural right and thus falls under the PTO’s rulemaking authority. Moreover, the proposed rule would clarify the ambiguity surrounding the statutory language of “person,”²⁷¹ just as the procedural interpretative rule in *Cooper* clarified the ambiguity surrounding the term “original application.”²⁷²

2. *Promulgating the Rule*

To promulgate the rule properly, the PTO should subject the proposed rule to notice-and-comment²⁷³ and publish it in the Code of Federal Regulations.²⁷⁴ With Chevron step zero satisfied, the next step is to apply step one.

268. *See Tafas*, 559 F.3d at 1355.

269. *JEM Broad. Co. v. F.C.C.*, 22 F.3d 320, 327 (D.C. Cir. 1994) (emphasis in original).

270. *Fund Democracy, LLC v. S.E.C.*, 278 F.3d 21, 28 (D.C. Cir. 2002).

271. *See infra* Section V.B.

272. *See Cooper Techs. Co. v. Dudas*, 536 F.3d 1330, 1336 (Fed. Cir. 2008).

273. The notice-and-comment requirements are set out in 5 U.S.C. § 553 (2018). The PTO already promulgates some procedural rules through notice-and-comment. *See, e.g.*, 77 Fed. Reg. 48710 (Aug. 14, 2012); 77 Fed. Reg. 56080 (Sep. 11, 2012). A deviation from this practice might raise concerns, as the proposed rule will contradict a Supreme Court decision. Furthermore, the process of notice-and-comment would demonstrate a detailed and well-reasoned analysis that could be useful for Chevron step two. *See infra* Section V.C.

274. Publishing proposed rules in the Federal Register is one requirement under notice-and-comment. 5 U.S.C. § 553(b). The Code of Federal Regulations then codifies these rules. *About the Code of Federal Regulations*, GOVINFO, <https://www.govinfo.gov/help/cfr> (last visited Feb. 5, 2020). The Federal Circuit has discussed whether a rule that is not published in the Code of Federal Regulations can be entitled to deference. *Cooper Techs. Co. v. Dudas*, 536 F.3d 1330, 1337 (Fed. Cir. 2008). That the Federal Circuit addressed this means it was a point of concern that should be avoided.

Even though these steps are not required under 5 U.S.C. § 553(b)(3)(A) for rules that are procedural and interpretative, this Note recommends following them. *See Cooper*, 536 F.3d at 1336–37. Indeed, according to the Federal Circuit, when Congress only grants an agency with the authority to establish regulations, the agency should follow the proper procedures for

B. AMBIGUOUS STATUTORY TERMS

The proposed rule would likely satisfy *Chevron* step one. “Only a judicial precedent holding that the statute unambiguously forecloses the agency’s interpretation, and therefore contains no gap for the agency to fill, displaces a conflicting agency construction.”²⁷⁵ The Supreme Court did not hold that the statutes in *Return Mail* unambiguously required the Court’s construction. Indeed, when discussing usages of “person” throughout the Patent Act and the AIA, the Court stated that “[s]ometimes ‘person’ plainly includes the Government, sometimes it plainly excludes the Government, and sometimes—as *here*—it might be read *either way*.”²⁷⁶ The nature of the Dictionary Act’s presumption also demonstrates that the word “person” “does not unambiguously require [the] result” that sovereigns are excluded,²⁷⁷ as the presumption is “no hard and fast rule of exclusion”²⁷⁸ and can be rebutted by the purpose, subject matter, context, legislative history, or executive interpretation.²⁷⁹

Moreover, although the Court held that no directly relevant executive interpretation currently exists, the Court did not foreclose the possibility of considering a future executive interpretation. The Court’s decision in *Brand X* illustrates this point. In *Brand X*, the Court held that the FCC’s interpretation of a statute controlled, despite a prior conflicting Ninth Circuit decision.²⁸⁰ In the conflicting Ninth Circuit opinion, the Ninth Circuit “noted that it was ‘not presented with a case involving potential deference to an administrative agency’s statutory construction pursuant to the *Chevron* doctrine’ ”²⁸¹ and that the relevant agency “declined . . . to address the issue before us.”²⁸² Therefore, because the Ninth Circuit explicitly recognized that an agency interpretation (which did not yet exist) might have been relevant, the Supreme Court in *Brand*

promulgating regulations in order to receive *Chevron* deference. See *Facebook, Inc. v. Windy City Innovations, LLC*, 973 F.3d 1321, 1349–50, 1353 (Fed. Cir. 2020) (Additional Views).

275. *Nat’l Cable & Telecommunications Ass’n v. Brand X Internet Servs.*, 545 U.S. 967, 982 (2005).

276. *Return Mail, Inc. v. United States Postal Serv.*, 139 S. Ct. 1853, 1863 (2019) (emphases added).

277. *Brand X*, 545 U.S. at 989.

278. *Return Mail*, 139 S. Ct. at 1868 (Breyer, J., dissenting) (quoting *United States v. Cooper Corp.*, 312 U.S. 600, 604–05 (1941)).

279. *Id.* (citing *Int’l Primate Prot. League v. Administrators of Tulane Educ. Fund*, 500 U.S. 72, 83 (1991)).

280. *Brand X*, 545 U.S. at 974, 984.

281. *Id.* at 985 (2005) (quoting *AT&T Corp. v. City of Portland*, 216 F.3d 871, 876 (9th Cir. 2000)).

282. *AT&T*, 216 F.3d at 876.

X concluded that the Ninth Circuit did not hold that the statute unambiguously required the Ninth Circuit's construction.²⁸³

In a similar manner, the Supreme Court in *Return Mail* implied that it might have deferred to the PTO's interpretation of a "person" who could initiate AIA review proceedings if the PTO had promulgated a directly relevant rule in the MPEP. In MPEP § 2212, the PTO stated that governmental entities qualified as persons who could petition for ex parte reexamination. With this in mind, the Court reasoned the following:

We might take account of this "executive interpretation" [from MPEP § 2212] if we were determining whether Congress meant to include the Government as a "person" for purposes of the ex parte reexamination procedures themselves. Here, however, the Patent Office's statement in the 1981 MPEP has no direct relevance.²⁸⁴

Thus, in *Return Mail*, it was specifically because an explicit agency interpretation was absent (among other things) that the Court held that a "person" did not include sovereigns. After the Court's holding, if the PTO promulgated the proposed rule stating that a person includes federal agencies, then the reasoning behind the Court's opinion—that a directly relevant agency interpretation was lacking—would fall apart.

In addition, the Court left open the question of whether common law estoppel applied to section 1498 suits,²⁸⁵ thus further supporting that the Court believed the term "person" was ambiguous. Multiple potential interpretations of the statutory estoppel provisions exist. One is that Congress omitted section 1498 claims from the provision because it never meant for the government to petition for AIA review proceedings. Congress could also have made a drafting error. Or, as argued in Section IV.D, Congress intentionally omitted section 1498 suits and was satisfied that mutual collateral estoppel would apply to 1498 suits instead.

As a result, the meaning of the term "person" is still in flux; as long as the PTO's later executive interpretation fulfills the other steps of *Chevron* deference, it should be sufficient to trump the Supreme Court's interpretation.

C. PERMISSIBLE CONSTRUCTION

The proposed rule would also fulfill *Chevron* step two. A court is "obliged 'to accept the agency's construction of the statute, even if the agency's reading

283. See *Brand X*, 545 U.S. at 985.

284. *Return Mail, Inc. v. United States Postal Serv.*, 139 S. Ct. 1853, 1865 (2019) (citing *United States v. Cerecedo Hermanos y Compania*, 209 U.S. 337, 339 (1908)).

285. *Id.* at 1867 n.10.

differs from what the court believes is the best statutory interpretation,’²⁸⁶ so long as the construction is “a reasonable interpretation.”²⁸⁷ An agency can demonstrate that its interpretation is reasonable by “consider[ing] the matter in a detailed and reasoned fashion”²⁸⁸ and by showing that the interpretation furthers “the primary aim of the statute.”²⁸⁹

The proposed rule would further the AIA’s purposes. Section IV.A.2.b demonstrated that permitting government agencies to initiate AIA review proceedings can increase patent quality and decrease the volume of, or shorten the length of, district court litigation. Also, as argued in Section IV.A.2.a, under the prior construction canon, eligible petitioners of *ex parte* reexamination, including governmental entities, should be eligible to initiate AIA review proceedings. After all, the two types of proceedings share the same purposes. Parts III.B–III.D also rebutted relevant counterarguments from the Supreme Court and Judge Newman. If the PTO publishes the proposed rule in the Federal Register and explains these points thoroughly, it would demonstrate a well-reasoned analysis.²⁹⁰

VI. CONCLUSION

Government agencies should be permitted to initiate AIA review proceedings. MPEP § 2212 interpreted “person,” for purposes of *ex parte* reexamination, to include governmental entities. Under the prior construction canon, MPEP § 2212, which is entitled to deference, should have applied to AIA review proceedings. Like *ex parte* reexamination, AIA review proceedings were meant to increase patent quality and decrease patent litigation, and permitting agencies to initiate AIA review proceedings furthers these purposes.

The Court and Judge Newman’s counterarguments are not persuasive. Even though 28 U.S.C. § 1498 provides the government with special treatment, it is not unfair—it is justified by sovereign immunity, eminent domain, and national security concerns. Moreover, letting agencies initiate AIA

286. *Alabama Educ. Ass’n v. Chao*, 455 F.3d 386, 396 (D.C. Cir. 2006) (quoting *Nat’l Cable & Telecommunications Ass’n v. Brand X Internet Servs.*, 545 U.S. 967, 980 (2005)).

287. *Nat’l Mining Ass’n v. Kempthorne*, 512 F.3d 702, 709 (D.C. Cir. 2008).

288. *Id.* at 709–10 (quoting *Kennecott Utah Copper Corp. v. U.S. Dep’t of Interior*, 88 F.3d 1191, 1206 (D.C. Cir. 1996)) (alteration in original); *see also* *Chevron, U.S.A., Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837, 865 (1984).

289. *Kempthorne*, 512 F.3d at 710.

290. *See* *Cooper Techs. Co. v. Dudas*, 536 F.3d 1330, 1342 (considering the PTO’s remarks in the Federal Register and holding that the procedural rule satisfied *Chevron* step two).

review proceedings would not chill inventors from making governmentally beneficial inventions, as such a chilling effect would also hurt the government.

Next, it is not awkward for government agencies to petition for AIA review proceedings. Allowing governmental petitioners does not create a conflict between executive agencies that the President should resolve. It could only create conflicts between the governmental petitioner and the patentee, or the PTO and the patentee. The concern that the PTO would be deferential to a fellow agency is also unsupported given that the PTO has disagreed with other agencies. Most importantly, proceedings already exist, even in the patent system, that permit an agency to exist in an adversarial relationship with a civilian while another government agency oversees the proceeding.

Furthermore, although the AIA statutory estoppel provisions do not mention 28 U.S.C. § 1498, Congress could still have meant to include the government as a “person.” A presumption exists that collateral estoppel prevents relitigating issues that were actually litigated and determined in administrative proceedings. Congress had good reasons to subject the government to less stringent estoppel requirements based on national security concerns and the belief that the government would less likely harass patentees.

These issues are not confined to *Return Mail*. The Federal Reserve banks, in *Bozeman Financial LLC v. Federal Reserve Bank of Atlanta et al.*,²⁹¹ successfully argued before the Federal Circuit that they should be able to initiate AIA review proceedings because they are more like corporations, not government agencies.²⁹² Whether or not this is correct, the PTO can resolve such issues by using its procedural rulemaking authority to promulgate a rule allowing governmental entities to initiate AIA review proceedings. The Court implied in *Return Mail* that an executive interpretation could have affected its decision, and the rule would further the goals of increasing patent quality and decreasing patent litigation. It is, therefore, up to the PTO to fix the Court’s mistake and to prevent more cases from being litigated based on erroneous reasoning.

291. 955 F.3d 971, 974, 974-76 (Fed. Cir. 2020), *cert. denied*, 141 S. Ct. 1053 (2021).

292. Britain Eakin, *Justices Won't Consider If Fed Banks Can Challenge Patents*, LAW360 (Jan. 11, 2021), <https://www.law360.com/articles/1343388/justices-won-t-consider-if-fed-banks-can-challenge-patents>; Tiffany Hu, *Fed Banks Have No Place At PTAB*, *Fed. Circ. Told*, LAW360 (Aug. 7, 2019), <https://www.law360.com/articles/1185999/fed-banks-have-no-place-at-ptab-fed-circ-told>.

FOURTH ESTATE AS A VEHICLE TO IMPEL REGISTRATION AND LIMIT ORPHAN WORKS?

Kevin Han Yang[†]

I. INTRODUCTION

Whenever a work is fixed in a tangible medium of expression for the first time, a copyright in that work automatically comes into existence.¹ But if the copyright holder wants to sue an alleged infringer, they must comply with the provisions of Title 17, the part of the U.S. Code that deals with copyrights. One such provision is 17 U.S.C. § 411(a), which states that, for U.S. works, a complaint cannot be filed until “registration . . . has been made in accordance with this title.”²

Section 411(a) is written in the passive voice, a grammatical faux pas which led to a question that split the circuits: by whom must registration be made? On one side, the Fifth and Ninth Circuits adopted the “application approach,” where the copyright holder satisfies the registration requirement by sending in a complete application.³ On the other side, the Tenth and Eleventh Circuits championed the “registration approach,” which views registration as made by the Copyright Office (“Office”) when it either approves or denies the registration application.⁴ In *Fourth Estate Public Benefit Corp. v. Wall-Street.com, LLC*, the Supreme Court affirmed the Eleventh Circuit and adopted the registration approach nationwide.⁵ Justice Ginsburg wrote a unanimous opinion for the court, and no separate opinions were filed.⁶

DOI: <https://doi.org/10.15779/Z38MG7FW7F>

© 2020 Kevin Han Yang.

† J.D., 2021, University of California, Berkeley, School of Law. Many thanks to Professor Talha Syed, Erin Delaney, Julea Lipiz, and Miranda Rutherford for their advice and guidance throughout the process of writing and revising this Note.

1. 17 U.S.C. § 102(a) (2018).

2. 17 U.S.C. § 411(a) (2018).

3. *See* *Apple Barrel Prods. v. Beard*, 730 F.2d 384, 386–87 (5th Cir. 1984), *abrogated in part* by *Fourth Estate Pub. Benefit Corp. v. Wall-Street.com, LLC*, 139 S. Ct. 881 (2019); *Cosmetic Ideas, Inc. v. IAC/InteractiveCorp*, 606 F.3d 612, 621–22 (9th Cir. 2010), *abrogated by Fourth Estate*, 139 S. Ct. 881.

4. *See* *La Resolana Architects v. Clay Realtors Angel Fire*, 416 F.3d 1195, 1205 (10th Cir. 2005), *abrogated in part* by *Reed Elsevier, Inc. v. Muchnick*, 559 U.S. 154 (2010); *Fourth Estate Pub. Benefit Corp. v. Wall-Street.com, LLC*, 856 F.3d 1339, 1341 (11th Cir. 2017), *aff’d*, 139 S. Ct. 881 (2019).

5. 139 S. Ct. 881 (2019).

6. *Id.*

Fourth Estate was correctly decided, both as a matter of statutory analysis and as a matter of policy. Textually, the Court was correct to focus specifically on the context of § 411(a) itself and decline Fourth Estate's invitation to go on a broad-ranging survey of other sections. More importantly for this Note, the Court also correctly handled the policy dimensions. The application approach courts were sympathetic to plaintiffs, seeking to protect their works from infringement, who may have accidentally run afoul of the statute of limitations. But they neglected the general public's interest in a complete, accurate, and current registry. And they also did not mention that Congress, the people's representatives, had already balanced these interests, and had over multiple decades consistently struck the balance in favor of maintaining the registry. The Supreme Court recognized all of this, albeit somewhat obliquely, and therefore, rightly rejected the application approach attempt to write policy preferences for plaintiffs into the copyright law.

Section 411(a) and the registration formality have also played a central role in the vigorous, long-raging debate in legal academia over whether copyright formalities are a good thing.⁷ This debate has generated several proposed formalities to help tailor the American copyright system to better suit the author's end goals.⁸ In the meantime, the Office has noted a growing problem with so-called "orphan works."⁹ Orphan works, which are works for which the copyright owner cannot be found, can create significant legal uncertainty, which may prevent an otherwise-interested potential licensee from productively using the orphaned work.¹⁰ Fourth Estate, copyright formalities, orphan works: these three seemingly-disparate strands actually tie together. Congress's insistence on maintaining a complete and accurate registry through formalities helped guard against works being orphaned.¹¹ In contrast, the modern, international trend of loosening formalities has worsened the orphan

7. See, e.g., Christopher Sprigman, *Reform(aliz)ing Copyright*, 57 STAN. L. REV. 485, 489–90 (2004) (promoting formalities as a way to adapt copyright law to the internet age and allow easy reuse of works with limited commercial value); Jane C. Ginsburg, "With Untired Spirits and Formal Constancy": *Berne Compatibility of Formal Declaratory Measures to Enhance Copyright Title-Searching*, 28 BERKELEY TECH. L.J. 1583, 1584–86 (2013) (criticizing formalities proponents as either actively seeking to expropriate authors for non-compliance or hiding their true confiscatory motives behind a mask of concern for the completeness of the public record).

8. See, e.g., Ginsburg, *supra* note 7, at 1588 (proposing conditioning the validity of a transfer of copyright ownership on recordation of a written document with the Copyright Office).

9. See U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS 1–14 (2006).

10. *Id.* at 1.

11. See *id.* at 43.

works problem.¹² But there may be a solution: Fourth Estate incentivizes timely registration of works. When combined with new proposed formalities, ones that even anti-formalities scholars approve of, Fourth Estate can help Office records stay valid and complete, thereby mitigating the creation of future orphan works.

This Note proceeds in four parts. Part II addresses § 411(a) and other relevant provisions of the Copyright Act, including their development and the legislative history and intent that shaped their current forms. Part III discusses Fourth Estate and the circuit split it resolved. Part IV surveys the academic debate surrounding copyright formalities, as well as the literature discussing the orphan works problem. Lastly, Part V argues that Fourth Estate, when combined with other proposed formalities in the literature, will help prevent future orphan works, thus vindicating the policy rationales that have preserved the registration formality in American copyright law.

II. A BRIEF HISTORY OF THE REGISTRATION REQUIREMENT

To contextualize § 411(a) and the *Fourth Estate* decision, this Note briefly discusses the history of the registration requirement in American copyright law. The current version of 17 U.S.C. § 411(a) has two main components. First, § 411(a) sets the general rule that a copyright infringement lawsuit for a U.S. work cannot be filed “until preregistration or registration of the copyright claim has been made in accordance with this title.”¹³ Second, an applicant who is denied registration by the Office is given an entitlement to file suit anyway, provided that they serve notice and a copy of the complaint on the Register of Copyrights.¹⁴ The Register then has a statutory right to intervene in the lawsuit on the issue of registrability, although refusal to intervene does not deprive the court of jurisdiction over that issue.¹⁵

12. *See id.* As the trend is international, the orphan works problem is also international. *See, e.g.*, MARCO RICOLFI, LYNNE BRINDLEY, CLAUDIA DILLMAN, TARJA KOSKINEN-OLSSON, TOBY BAINTON, ANNE BERGMAN-TAHON, JEAN-FRANÇOIS DEBARNOT, MYRIAM DIOCARETZ & OLAV STOKKMO, COPYRIGHT SUBGROUP, HIGH LEVEL EXPERT GRP. ON DIG. LIBRARIES, REPORT ON DIGITAL PRESERVATION, ORPHAN WORKS, AND OUT-OF-PRINT WORKS 4–9 (2007).

13. 17 U.S.C. § 411(a) (2018). Authors asserting moral rights in visual works are exempt from this requirement. *Id.*

14. *Id.*

15. *Id.*

A. BEFORE THE 1976 ACT

Harsh, mandatory formalities were a part of American copyright law from its very beginnings.¹⁶ When the newly-independent United States enacted its first copyright law, the Copyright Act of 1790, it required several formalities as preconditions to copyright protection.¹⁷ Among these formalities was a requirement that an author register their work with the local district court.¹⁸ Since a copyright under the 1790 Act could not arise without registration, it necessarily followed that a suit for copyright infringement could not be filed until the copyrighted work had actually been registered by the government. The rest of the world gradually loosened their formalities over time, foreshadowing the eventually successful calls for the United States to follow suit.

The non-mandatory registration requirement, as we know it today, was first enacted by the Copyright Act of 1909.¹⁹ The 1909 Act was a compromise—it made copyright attach to a work upon publication with notice of copyright,²⁰ thereby removing registration as a prerequisite to copyright for the first time in U.S. history.²¹ To counterbalance this, § 12 of the Act instituted formalities as a prerequisite to filing a copyright infringement suit.²² As originally enacted, § 12 stated that a copyright infringement suit could not be maintained until the plaintiff complied with the deposit and registration formalities.²³ With this, the United States reached a general balance between rewarding authors and promoting registration, which it maintains to this day (at least domestically). But even this was not enough for some.

The 1909 Act registration requirement did not answer a major question—can an author who applies for registration but is rejected nevertheless sue?²⁴ In the influential 1958 case, *Vacheron & Constantin-Le Coultre Watches, Inc. v. Benrus Watch Co.*, the Second Circuit held that such an applicant cannot sue unless they first succeed in a mandamus action compelling the Register to accept their

16. See 5 WILLIAM F. PATRY, PATRY ON COPYRIGHT § 17:83 (2020).

17. Copyright Act of 1790, §§ 3–4, 1 Stat. 124, 125 (repealed 1831).

18. *Id.* § 3.

19. Pub. L. No. 60-349, § 12, 35 Stat. 1075, 1078 (1909) (repealed 1976).

20. *Id.* § 9.

21. See Daniel Gervais & Dashiell Renaud, *The Future of United States Copyright Formalities: Why We Should Prioritize Recordation, and How To Do It*, 28 BERKELEY TECH. L.J. 1459, 1467–68 (2013).

22. Copyright Act of 1909, Pub. L. No. 60-349, § 12.

23. *Id.*

24. See *id.*

application.²⁵ The plaintiff had twice tried to register a copyright in a watch, and each time the Office refused to issue a certificate of registration, on the grounds that the watch was not copyrightable as a work of art.²⁶ Judge Learned Hand held that the plaintiff failed to comply with the registration requirement, and therefore could not sue.²⁷

The majority made one argument from precedent and one from text.²⁸ Judge Hand emphasized the D.C. Circuit's decision in *Bouve v. Twentieth Century-Fox Film Corp.*, affirming a grant of mandamus against the Register.²⁹ Judge Hand reasoned that since mandamus is only available when no other adequate remedy exists, that must mean a copyright lawsuit cannot be filed while the question of registration has not been decided.³⁰ And because the Register's rejection is reviewable by the courts, plaintiff Vacheron was obligated to first seek mandamus.³¹ The opinion concluded its analysis with the text of the registration requirement: the applicant was required to comply with both the deposit and registration formalities.³² But submission of an application for registration was required to fulfill the deposit formality, so the only remaining condition for registration as a separate formality had to be the Register's acceptance.³³ The majority, therefore, embraced a souped-up version of the registration approach, forcing rejected applicants to start a separate lawsuit against the Register—and presumably leaving those who lost that lawsuit out of luck.

Dissenting, Chief Judge Clark expressed the same policy views and party sympathies as modern day application approach advocates.³⁴ In Clark's view, the plaintiff complied with the deposit and registration requirements by submitting the application for registration, even though they were rejected.³⁵ Plaintiff should not have to “assume the risk of waiting[,] perhaps until its right is lost[,]” for the Register to perform their statutory duty.³⁶ Therefore, as an act of “simple justice” and equity, the lower court should have held the case

25. 260 F.2d 637, 639–41 (2d Cir. 1958), *superseded by statute*, Copyright Act of 1976, Pub. L. No. 94-553, 90 Stat. 2541, *as recognized in* *Reed Elsevier, Inc. v. Muchnick*, 559 U.S. 154 (2010).

26. *Id.* at 644 (Clark, C.J., dissenting).

27. *Id.* at 639 (majority opinion).

28. *See id.* at 639–41.

29. *See id.* at 640.

30. *Id.*

31. *See id.*

32. *Id.* at 640–41.

33. *Id.*

34. *See id.* at 645 (Clark, C.J., dissenting).

35. *Id.* at 644–45.

36. *Id.* at 645.

until the Register granted or was compelled to grant registration, after which plaintiff could file a supplemental complaint.³⁷ Eventually, Chief Judge Clark would win the specific battle, but lose the overall war. Congress abrogated *Vacheron* by statute in 1976, but—ironically—the details of Congress’s enacted provision would later prove critical to the registration approach’s nationwide victory.³⁸

B. THE COPYRIGHT ACT OF 1976 AND BEYOND

The Copyright Act of 1976 (the 1976 Act) is the basis of the modern American copyright regime.³⁹ The 1976 Act represented a sea change.⁴⁰ For the first time, copyright in a work became automatic upon fixation, without any prerequisite formalities.⁴¹ More importantly for this Note, the 1976 Act codified the registration requirement as 17 U.S.C. § 411(a). The new wording of the registration requirement was as follows: “Subject to the provisions of subsection (b), no action for infringement of the copyright in any work shall be instituted until registration of the copyright claim has been made in accordance with this title.”⁴²

The 1976 Act’s major change to the registration requirement was the addition of a provision explicitly repudiating *Vacheron*.⁴³ The new § 411(a) granted rejected applicants the right to sue upon denial, without the need to first obtain mandamus against the Register.⁴⁴ If the Register refused a properly made application, the applicant would still be allowed to file suit for copyright infringement, provided that they serve a notice and a copy of the complaint on the Register.⁴⁵ The Register would then have the statutory right to intervene within sixty days, thereby becoming a party to the action with respect to the issue of registrability.⁴⁶

The 1976 Act brought American copyright law significantly closer to the international, anti-formality vision of copyright.⁴⁷ By the time the 1976 Act was enacted, the international community had coalesced around the anti-formality

37. *Id.*

38. *See* Copyright Act of 1976, Pub. L. No. 94-553, § 411(a), 90 Stat. 2541, 2583; *Fourth Estate Pub. Benefit Corp. v. Wall-Street.com, LLC*, 139 S. Ct. 881, 889, 890–91 (2019).

39. *See* 1 PATRY, *supra* note 16, § 1:71.

40. *See id.* § 1:72.

41. *Compare* Copyright Act of 1976 § 102, *with* Copyright Act of 1909, Pub. L. No. 60-349, § 9, 35 Stat. 1075, 1078 (1909).

42. Copyright Act of 1976 § 411(a).

43. H.R. REP. NO. 94-1476, at 157 (1976).

44. *See* Copyright Act of 1976 § 411(a).

45. *Id.*

46. *Id.*

47. *See* 1 PATRY, *supra* note 16, § 1:71.

regime of the Berne Convention for the Protection of Literary and Artistic Works (“Berne”).⁴⁸ Berne, the major international copyright treaty, allows for international copyright protection by obligating its member states to respect the copyrights in works originating from other member states.⁴⁹ Berne’s view of formalities is made clear in Article 5(2), which explicitly states that “[t]he enjoyment and the exercise” of copyright “shall not be subject to any formality.”⁵⁰ The Berne formality ban does have an important exception: Berne member countries are allowed to impose formalities on domestic authors.⁵¹ This exception explains both § 411(a)’s continued existence and its limited application to United States works only.

The United States brought its copyright law into compliance with Berne through the Berne Convention Implementation Act of 1988.⁵² Joining Berne was a pragmatic decision born of economic necessity.⁵³ But, just because the United States signed on to the letter of Berne, did not mean that it had signed on to its spirit. In some areas, including formalities and copyright registration, Congress took a grudging, minimalist approach to Berne compliance.⁵⁴ The House of Representatives, in its official report, made it clear that it had decided to “revise United States law only to the extent absolutely necessary to make it compatible with Berne.”⁵⁵ In keeping with this philosophy, the House elected to keep § 411 on the grounds that it was a procedural requirement working no loss of copyright.⁵⁶ Since a copyright owner could sue regardless of whether their application was accepted or rejected, the House thought that registration was not a formality.⁵⁷ Importantly, the House report stated that § 411(a) was in the public interest because it would promote litigation efficiency, deter suits

48. See S. REP. NO. 100-352, at 2 (1988) (“For more than 100 years, the Berne Convention has been the major multilateral agreement governing international copyright relations.”).

49. Berne Convention for the Protection of Literary and Artistic Works art. 5(1), July 24, 1971, S. Treaty Doc. No. 99-27, 828 U.N.T.S. 221 [hereinafter Berne Convention].

50. See *id.* art. 5(2).

51. See *id.* art. 5(3).

52. Berne Convention Implementation Act of 1988, Pub. L. No. 100-568, § 2, 102 Stat. 2853, 2853.

53. See 7 PATRY, *supra* note 16, § 23:1 (characterizing the U.S. decision to join Berne as “driven by a serious deficit in the international balance of payments and an increased importance of intellectual property as a net export industry”).

54. See H.R. REP. NO. 100-609, at 40 (1988) (making clear that there was a “conceptual decision to revise United States law only to the extent absolutely necessary to make it compatible with Berne”).

55. *Id.*

56. *Id.* at 41.

57. *Id.*

over uncopyrightable material, and ensure a central public repository of copyright ownership claims.⁵⁸

Even after joining the anti-formality Berne regime, the United States has nevertheless rejected calls to take the last step and eliminate the registration before suit requirement. The proposed Copyright Reform Act of 1993 would have eliminated § 411(a) entirely.⁵⁹ The 1993 Act explicitly sought to bring American copyright law into harmony with international norms.⁶⁰ In language befitting the strongest Berne advocates, the House report on the 1993 Act decried § 411(a) as “no longer represent[ing] good copyright policy.”⁶¹ The report condemned the registration requirement for discriminating against U.S. authors in favor of foreigners, failing as a mechanism of incentivizing deposits, and serving as the Office’s justification for turning examination of registration applications into a more restrictive process that served as a gatekeeper to courts.⁶² The bill passed the House, but never made it through the Senate.⁶³

Instead, Congress addressed dissatisfaction with the registration requirement by enacting certain carveouts. First, in 1990 Congress exempted lawsuits invoking moral rights, under 17 U.S.C. § 106A, from § 411(a)’s scope.⁶⁴ Next, in 2005, Congress addressed the problem of pre-publication infringement by creating a copyright preregistration option.⁶⁵ As the Supreme Court later suggested in *Fourth Estate*, from this history a pattern emerges.⁶⁶ Congress has occasionally enacted new provisions to alleviate the concerns of copyright holders, but the public interest in the copyright registry is so strong that registration as a prerequisite to suit has been maintained since its 1909 inception.⁶⁷

58. *Id.* at 41–42.

59. Copyright Reform Act of 1993, H.R. 897 (103rd Cong. § 6(a)(B) (1993)).

60. H.R. REP. NO. 103-388, at 9 (1993).

61. *Id.*

62. *Id.* at 9–11.

63. John Tehranian, *The Emperor Has No Copyright: Registration, Cultural Hierarchy, and the Myth of American Copyright Militancy*, 24 BERKELEY TECH. L.J. 1399, 1442 (2009).

64. Judicial Improvements Act of 1990, Pub. L. No. 101-650, § 606(c)(1), 104 Stat. 5089, 5131.

65. Family Entertainment and Copyright Act of 2005, Pub. L. No. 109-9, § 104(1), 119 Stat. 218, 221–22.

66. *See Fourth Estate Pub. Benefit Corp. v. Wall-Street.com, LLC*, 139 S. Ct. 881, 891 (2019).

67. *See id.*

III. A CIRCUIT SPLIT AND ITS RESOLUTION

Because of Congress's failure to specify in § 411(a) who makes registration, and when, circuits were forced to make the decision themselves. A circuit's reasoning and ultimate decision tended to reflect its views, not just of registration before suit, but of formalities in general. Along this axis, a split emerged. Circuits adopting the "application approach" tended to focus on the subset of public policy concerns weighing against formalities, with special concern for authors who might lose the right to sue because of the statute of limitations expiring during Office processing of the registration application.⁶⁸ Circuits following the "registration approach" instead emphasized formalities as critical incentives in an overall statutory structure, noting the sheer number of provisions that would be meaningless if registration did not require approval or denial by the Office.⁶⁹ The Supreme Court essentially adopted the registration approach reasoning wholesale, with no concession for public policy.⁷⁰ This Note will now survey both sides of the circuit split, as well as the Supreme Court's eventual resolution.

A. THE APPLICATION APPROACH

The Ninth Circuit's *Cosmetic Ideas* decision clearly illustrates application approach courts' textual analysis of copyright law, as well as their anti-formalist policy concerns.⁷¹ *Cosmetic Ideas* involved copyright in a necklace design.⁷² The plaintiff filed suit after sending an application to the Office and receiving confirmation of receipt, but before the Register of Copyrights approved or denied the application.⁷³ The district court granted defendant's motion to dismiss for lack of subject-matter jurisdiction.⁷⁴

The majority acknowledged that it needed to start with the plain language of the statute.⁷⁵ But since neither § 411(a) nor the definition of "registration"

68. See *Cosmetic Ideas, Inc. v. IAC/InteractiveCorp*, 606 F.3d 612, 618–22 (9th Cir. 2010) (choosing the application approach because it better fulfilled Congress's purpose in passing § 411(a)), *abrogated by Fourth Estate*, 139 S. Ct. 881.

69. See *La Resolana Architects v. Clay Realtors Angel Fire*, 416 F.3d 1195, 1203–05 (10th Cir. 2005) (rejecting the application approach, despite its policy appeal, because the plain meaning of § 411(a) commanded otherwise), *abrogated in part by Reed Elsevier, Inc. v. Muchnick*, 559 U.S. 154 (2010).

70. See *Fourth Estate*, 139 S. Ct. at 888–92.

71. See *Cosmetic Ideas*, 606 F.3d at 616–21. These policy concerns persisted even after the U.S. eliminated many of its formalities to join Berne. See H.R. REP. NO. 103-388, at 9–11 (1993).

72. *Cosmetic Ideas*, 606 F.3d at 614.

73. *Id.*

74. *Id.*

75. *Id.* at 616.

in 17 U.S.C. § 101 stated who makes registration or when it occurs, the court turned to the statute as a whole and found it ambiguous.⁷⁶ The court observed that the second part of § 411(a) discusses “registration ha[ving] been refused,” suggesting that registration is an act by the Register that is separate from mere delivery.⁷⁷ But instead of letting that settle the textual inquiry, the majority noted two other provisions of the Copyright Act, §§ 408 and 410(d), as creating doubt as to who obtains registration and whether the applicant’s or Register’s action is more important.⁷⁸ The majority thus decided to resolve the conflict through policy, by choosing the approach it felt best served Congress’s purpose for enacting, not the specific § 411(a) formality, but the entire, generally anti-formality 1976 Act.⁷⁹

Turning to policy, the court found four reasons why the application approach better suited the purpose of “providing broad copyright protection while maintaining a robust federal register.”⁸⁰ Two of the arguments were about the application approach’s advantages.⁸¹ First, the application approach avoided unnecessary delay, since the Office’s ultimate decision does not impact plaintiff’s right to sue.⁸² In contrast, the registration approach would lead to a period of “legal limbo” that is strange and inefficient, one that academics and even registration approach courts had lamented.⁸³ Second, Office delay could result in a plaintiff running afoul of the three year statute of limitations for copyright infringement claims, thereby losing the right to sue entirely.⁸⁴ The court cited 17 U.S.C § 410(d) as proving that Congress had a goal of protecting applicants from the consequences of Office delay.⁸⁵ And when it came to fulfilling that goal, only the application approach would suffice.⁸⁶

The other two arguments played defense by arguing that the application approach could match the registration approach’s purported advantages.⁸⁷ Since the application and registration approaches both require the copyright

76. *Id.* at 616–17.

77. *Id.* (quoting 17 U.S.C. § 411(a) (1976)) (emphasis omitted).

78. *Id.* at 617–18.

79. *See Cosmetic Ideas*, 606 F.3d at 618.

80. *Id.* at 619–21.

81. *See id.*

82. *Id.* at 619–20.

83. *Id.* (quoting 2 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 7.16[B][1][a][i] (2008)).

84. *Id.* at 620–21. The court noted that the registration approach would therefore harm plaintiffs “who applied for registration towards the end of the three-year [statute of limitations]” *Id.* at 620.

85. *Cosmetic Ideas*, 606 F.3d at 620.

86. *Id.* at 620–21.

87. *Id.*

holder to submit an application for registration, complete with all the necessary information, both approaches equally served the goal of maintaining a robust national registry.⁸⁸ The court ended by rejecting arguments that the registration approach would give courts the benefit of the Register's opinion on copyrightability at the outset of litigation.⁸⁹ The majority noted that Office approval/rejection will typically occur before a lawsuit is resolved, giving the Register the opportunity to intervene.⁹⁰ Because the Register's decision is both perfunctory and reviewable by the courts, the better approach would be to let the Office and the courts work simultaneously, rather than waiting for the Register.⁹¹ In conclusion, the majority adopted the application approach to lessen the impact of the registration formality on authors and to allow plaintiffs to be heard on the merits, despite the lack of an Office decision on registrability.⁹²

Overall, the main error of the *Cosmetic Ideas* majority, both textually and on policy, was that it looked at the whole forest while missing the specific tree of interest. For its textual analysis, the court acknowledged that the second part of § 411(a) supported the registration approach, yet it went on a broad hunt for references to registration in other sections to give meaning to § 411(a)—even if that meant making § 411(a) internally inconsistent.⁹³ And even the court's holistic view of copyright's statutory scheme was incorrect: it did not mention—which the Supreme Court picked up on—the fact that the application approach would render useless several provisions of the Copyright Act.⁹⁴

On policy, the majority telegraphed its preferences when it found that Congress's purpose was “providing broad copyright protection while maintaining a robust federal register.”⁹⁵ The majority correctly noted that the 1976 Act *generally* reduced and eliminated formalities.⁹⁶ But the retention of registration before suit in particular—as an *exception* to the trend—should have alerted the majority as to just how important Congress thought it was to have a complete, accurate, and timely copyright registry. When viewed in that light,

88. *Id.* at 620.

89. *Id.* at 621.

90. *Id.*

91. *Id.*

92. *See Cosmetic Ideas*, 606 F.3d at 620–22.

93. *See id.* at 617–18; *see also* Brief for the Respondents at 22, *Fourth Estate Pub. Benefit Corp. v. Wall-Street.com, LLC*, 139 S. Ct. 881 (2019) (No. 17-571).

94. *Compare Cosmetic Ideas*, 606 F.3d at 617–18, *with Fourth Estate*, 139 S. Ct. at 888–90.

95. *See Cosmetic Ideas*, 606 F.3d at 619. It is quite revealing that the interest in maintaining a robust federal register, which § 411(a) was designed to further, is relegated to a dependent clause.

96. *Id.* at 618–19.

the court should have concluded that the registration approach was the correct interpretation of 17 U.S.C. § 411(a) to adopt.

B. THE SUPREME COURT AND THE REGISTRATION APPROACH

In contrast to *Cosmetic Ideas*, Justice Ginsburg's unanimous affirmance in *Fourth Estate Public Benefit Corp. v. Wall-Street.com, LLC*, wholly endorsed both registration approach reasoning and § 411(a) as part of Congress's vision of copyright formalities.⁹⁷ The very facts of *Fourth Estate*, which involved the copyrights in online news articles, are a stark demonstration of the importance of actually waiting for the Register's decision.⁹⁸ Plaintiff's application for registration was rejected multiple times, both because its check for fees was rejected by the bank and, more importantly, because it had incorrectly submitted its news articles as a group database.⁹⁹ Nevertheless, plaintiff filed suit, with a complaint alleging that an application for registration had been filed.¹⁰⁰ The district court dismissed the complaint, and the Eleventh Circuit affirmed.¹⁰¹ The Eleventh Circuit confirmed its adoption of the registration approach using mostly textual arguments, but it did briefly acknowledge that § 411(a) was a part of Congress's statutory scheme to encourage early registration without formally compelling it.¹⁰²

Justice Ginsburg began the Court's opinion with the statutory text, mainly focusing on how many of the Copyright Act's statutory provisions would be superfluous under the Ninth Circuit's application approach.¹⁰³ Starting with § 411(a) itself, there would be no need to give an applicant the right to sue upon rejection if the application approach were adopted.¹⁰⁴ Furthermore, the last sentence of § 411(a), giving the Register the right to intervene, would be useless if a suit could be filed and resolved (perhaps through settlement) before the Register could finish examination.¹⁰⁵ The Court rejected plaintiff's argument that the purpose of the second sentence of § 411(a) was to require plaintiffs to give notice to the Office of suits based on rejected registrations.¹⁰⁶

97. See *Fourth Estate*, 139 S. Ct. at 886, 888–91.

98. See *id.* at 887.

99. Brief for the United States as Amicus Curiae Supporting Respondents at App. 1a–9a, *Fourth Estate Pub. Benefit Corp. v. Wall-Street.com, LLC*, 139 S. Ct. 881 (2019) (No. 17–571).

100. *Fourth Estate*, 139 S. Ct. at 887.

101. *Id.*

102. See *Fourth Estate Pub. Benefit Corp. v. Wall-Street.com, LLC*, 856 F.3d 1338, 134–42 (11th Cir. 2017), *aff'd*, 139 S. Ct. 881 (2019).

103. See *Fourth Estate*, 139 S. Ct. at 888–90.

104. *Id.* at 889.

105. *Id.*

106. *Id.*

Such a conclusion would require the word “application” to have two different meanings in two consecutive sentences.¹⁰⁷ The Court found other parts of Title 17 further confirmed the registration approach’s correctness: §§ 410(a)–(b) showed that application and registration are separate, and there would be no reason to enact §§ 410(d) and 408(f) under the application approach.¹⁰⁸

Ultimately, Justice Ginsburg, unlike the *Cosmetic Ideas* court, correctly recognized that it was necessary to “look to the specific context” of the word “registration.”¹⁰⁹ And in the context of § 411(a) in particular, registration would have to refer to the Office’s approval of an application.¹¹⁰ The Court persuasively articulated why the statutory structure supported the registration approach. But ultimately all that was needed was the fact that the application approach forces § 411(a) to contradict itself, a point which Wall-Street.com’s brief made more clearly and directly than the Court’s opinion.¹¹¹

Next, the Court refuted Fourth Estate’s historical arguments.¹¹² § 411(a)’s enactment was not an endorsement of Chief Judge Clark’s *Vacheron* dissent and the application approach, but rather an endorsement of the Hand majority opinion and a reaffirmation of the registration requirement, with an added exception in the case of rejection.¹¹³ The exception, as discussed earlier, would be unnecessary if Office action was no prerequisite to filing suit.¹¹⁴ In addition, Congress had rejected multiple attempts across multiple decades to revoke the registration requirement.¹¹⁵ Instead, Congress balanced the policy concerns by keeping the general rule, while restricting its reach to domestic authors only.¹¹⁶ The Court’s historical analysis is confirmed by the text of the House Report on the 1976 Act, which the Court alluded to but did not directly quote.¹¹⁷ *Vacheron*, as construed by the report, covered “an applicant[] who has sought registration and has been refused”—which is inconsistent with Fourth Estate’s argument that such an applicant had already made registration by submitting their ill-fated application.¹¹⁸ Besides, if Congress had intended to embrace the

107. *Id.*

108. *Fourth Estate*, 139 S. Ct. at 889–90.

109. *See id.* at 890 (quoting Brief for Petitioner at 29, *Fourth Estate*, 139 S. Ct. 881 (No. 17-571)).

110. *Id.*

111. *See id.* at 889; Brief for the Respondents at 22–23, *Fourth Estate*, 139 S. Ct. 881 (No. 17-571).

112. *Fourth Estate*, 139 S. Ct. at 890–92.

113. *Id.* at 890–91.

114. *Id.* at 891.

115. *Id.*

116. *Id.*

117. *See id.* at 890–91.

118. *See* H.R. REP. NO. 94-1476, at 157 (1976).

application approach wholesale, would it not have chosen the easier and clearer route of saying that a lawsuit may not be filed unless an *application for registration* has been made in accordance with Title 17?

Fourth Estate's policy-based arguments fared no better. Fourth Estate argued that since registration is not required for copyright to exist, the application approach was needed to protect authors of works not yet fully registered.¹¹⁹ But copyright owners are protected by damages even before registration, as long as they make sure to get the Register's decision on their application before filing suit.¹²⁰ The Court also dismissed copyright plaintiffs' biggest concern, that copyright holders might be barred from relief due to the statute of limitations expiring before Office action.¹²¹ The majority noted that the average processing time was seven months.¹²² Responsibility for fixing administrative delays ultimately rested with Congress, who could fix any issues with budget or staff.¹²³ The Eleventh Circuit had therefore properly concluded that § 411(a) requires the Register to issue a decision on an application before a lawsuit can be filed, and the judgment was affirmed.¹²⁴

On policy, the Court correctly noted that the interests of copyright plaintiffs can be, *and in the past have been*, accommodated by Congress, without the need to reinterpret § 411(a).¹²⁵ Copyright policy, as well as history and text, thus confirm that Congress, regardless of what it thought about other formalities, highly valued registration before suit in particular. When faced with a plaintiff prejudiced by § 411(a), Congress chose, and would later choose again and again, to pass a specific accommodation, not to eliminate the registration formality entirely. For all these reasons, the Court was correct in choosing the registration approach.

IV. OF ORPHANS AND COPYRIGHT SCHOLARS

A. ORPHAN WORKS

The United States has a problem with orphan works, i.e., works for which the copyright owner cannot be determined or found—and the registration approach may help mitigate it.¹²⁶ Having a complete, accurate, and timely

119. *Fourth Estate*, 139 S. Ct. at 891.

120. *Id.*

121. *Id.* at 892.

122. *Id.*

123. *Id.*

124. *Id.*

125. *See id.* at 891–92.

126. *See* U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS, *supra* note 9, at 1; Brief of Pub. Knowledge & The R St. Inst. As Amici Curiae in Support of Respondents at 9–11,

registry serves an important public purpose by making it easier for prospective users to know that a work is copyrighted and contact the author, thus facilitating rights clearance and productive transactions.¹²⁷ As some have noted, Congress kept the registration formality to ensure such a comprehensive registry—and the registration approach adopted by the Supreme Court in *Fourth Estate* both vindicates this purpose and mitigates the orphan works problem about to be described.¹²⁸ Conversely, an incomplete or incorrect registry will leave prospective users in doubt as to whether they can make use of a work without the threat of future infringement liability.¹²⁹ A registry's usefulness is thus influenced by the level of formalities in the copyright law, and Congress's deformatizing in the 1976 Act therefore worsened the orphan works problem.¹³⁰ And even scholars who are generally opposed to formalities have acknowledged both that orphan works are a problem and that formalities may be part of the solution.¹³¹

While there are many reasons why a work could end up orphaned, one substantial reason is that the work was never registered with the Office.¹³² When this happens, potential users are deprived of a valuable source of information and a lead in the search process.¹³³ The inability to secure permission from an identifiable copyright owner makes the use of orphan works risky.¹³⁴ There is a possibility that the owner could resurface after extensive investment in and use of the orphan work has already been made, possibly resulting in massive liability.¹³⁵ The risk and uncertainty surrounding an orphan work is often enough to deter others, especially organizations such as libraries and nonprofits, from making productive use of it.¹³⁶ In this manner, a significant part of American culture and its works can end up behind a legal wall.¹³⁷

Fourth Estate, 139 S. Ct. 881 (No. 17-571), 2018 WL 5617891 [hereinafter Public Knowledge Amicus Brief].

127. See Ginsburg, *supra* note 7, at 1585.

128. See Public Knowledge Amicus Brief, *supra* note 126, at 9–11, 14–16.

129. See *id.* at 7–11.

130. See *id.* at 3, 5; see also U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS, *supra* note 9, at 43–44.

131. See Ginsburg, *supra* note 7, at 1613.

132. See U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS, *supra* note 9, at 3–4.

133. See *id.*

134. *Id.* at 1.

135. *Id.*

136. U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS, *supra* note 9, at 1.

137. See *id.*

In 2006, at the request of several senators and representatives, the Office released a report on the orphan works problem.¹³⁸ Despite the lack of prior systematic, empirical research, the Office estimated as a proxy that about fifty percent of the over 850 comments it received described orphan works situations.¹³⁹ It also drew statistics from some of the public comments it received—for example, a Carnegie Mellon University book digitization project had been unable to locate copyright owners for about twenty-two percent of its books.¹⁴⁰ Since then, other studies have confirmed the existence and substantial scope of the orphan works problem.¹⁴¹ U.S. studies of the proportion of orphan works in library collections have put forward estimates ranging from seventeen to more than fifty percent.¹⁴²

The Office's ultimate recommendation was to amend Title 17 to add a section limiting the remedies available to orphan work copyright owners.¹⁴³ The provision would be triggered if the user conducted a reasonably diligent search for the copyright owner without success and then proceeded to provide attribution if possible and reasonably appropriate.¹⁴⁴ If those conditions were met, both monetary and injunctive relief for the copyright holder would be limited.¹⁴⁵ Damages would not be assessed against a noncommercial user who promptly ceased use of the work upon notice, and in all other cases damages would be capped to reasonable compensation.¹⁴⁶ As for injunctions, a user turning the orphan work into a derivative work incorporating a significant amount of the user's expression could not be enjoined from continued use and preparation of the new work, provided that reasonable compensation was paid.¹⁴⁷ In all other circumstances, courts should account for harm to the user's reliance interest in continued use of the orphan work.¹⁴⁸ Lastly, the Office recommended that the provision sunset in ten years, so that Congress could evaluate if any changes were necessary.¹⁴⁹

138. *Id.* at 1–2.

139. *Id.* at 21.

140. *Id.* at 92 n.345.

141. See David R. Hansen, Gwen Hinze, Kathryn Hashimoto, Pamela Samuelson & Jennifer M. Urban, *Solving the Orphan Works Problem for the United States*, 37 COLUM. J.L. & ARTS 1, 4–11 (2013) (discussing several studies, both domestic and foreign, estimating the size of the orphan works problem).

142. *Id.* at 7–8.

143. U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS, *supra* note 9, at 93, 95.

144. *Id.* at 95–96.

145. *Id.* at 115.

146. *Id.*

147. *Id.*

148. *Id.*

149. *Id.* at 121.

This proposal, although never passed by Congress, may be a balanced method of allowing use of existing orphan works, but by its own terms it only deals with works that have already been orphaned.¹⁵⁰ The proposed statutory language does nothing to prevent works from becoming orphaned in the first place.¹⁵¹ The Report acknowledged that the orphan works problem was a result of the dramatic changes of the 1976 Act, even specifically pointing to the elimination of renewal registrations as a factor aggravating the orphan works problem.¹⁵² This suggests that formalities, especially registration, play an important role in providing the information needed to prevent orphan works. Vigorous enforcement of the registration formality through the registration approach thus may counteract the growing number of orphan works.¹⁵³

Ultimately, preventing the orphan works problem from worsening, rather than just finding ways to allow reasonable use of already-orphaned works, will require an effective method of keeping track of copyright ownership over long periods of time. The registration formality plays a crucial part in enabling this tracking. Even those suspicious of formalities have acknowledged the value that registration information provides to the public and to authors.¹⁵⁴ But registration, even as enforced by the registration approach, is only one piece of the puzzle. Prompt registration can track a work's creation, but subsequent events, such as sale or inheritance of the copyright, can render the registry out-of-date.¹⁵⁵ Might there be a way to ensure that the registry is updated whenever a copyright is transferred, such that the current owner is always known?

B. THE ACADEMIC LITERATURE ON COPYRIGHT FORMALITIES

Despite their somewhat technical and dry nature, copyright formalities have spawned considerable academic criticisms and proposed reforms.¹⁵⁶ Copyright academics writing about formalities have largely split into two sides. One side is in favor of increased formalities. Professor Chris Sprigman has written in defense of this side's views.¹⁵⁷ The other side, spearheaded by Professor Jane Ginsburg, is skeptical of the harsh penalties involved in new

150. *See id.* at 127.

151. *See* U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS, *supra* note 9, at 127.

152. *Id.* at 41–44.

153. *See* Public Knowledge Amicus Brief, *supra* note 126, at 2–3, 9–11.

154. *See, e.g.*, Ginsburg, *supra* note 7, at 1585–87.

155. *See id.* at 1613.

156. *See, e.g.*, Sprigman, *supra* note 7, at 486–91; Ginsburg, *supra* note 7, at 1584–88.

157. *See, e.g.*, Sprigman, *supra* note 7, at 554–68 (advocating the enactment of an inexpensive, mandatory license for works out of compliance with formalities, on the grounds that this would allow others to easily use works which the author economically judged to be worth less than the cost of formalities compliance).

formalities proposed by those like Sprigman.¹⁵⁸ While both sides agree that registration is generally desirable, they differ in the means they would use to encourage it.¹⁵⁹

In Sprigman's view, mandatory formalities are desirable in the internet age, so as to prevent copyright law from being the main barrier to creative reuse of works.¹⁶⁰ The current, voluntary incentives to register fall short because they cannot incentivize copyright holders who do not expect to make significant revenue from their works.¹⁶¹ Before the 1976 Act, these copyright holders would let their works fall into the public domain.¹⁶² Then anyone could copy and make use of the work to their heart's content, without need for payment or fear of a lawsuit.¹⁶³ But in the current regime, the costs of finding the copyright owner and negotiating with them in a limited-information context will be so high that the work will not be used.¹⁶⁴ Of course, Sprigman recognized that mandatory formalities affecting both domestic and international authors would violate Berne.¹⁶⁵ But Sprigman believes that re-imposing the full, traditional set of mandatory formalities on U.S. authors only, as permitted by Berne, would be "a significant improvement on the status quo."¹⁶⁶

Sprigman outlined two main proposals to reintroduce formalities into copyright.¹⁶⁷ One would require amending Berne, and the other would not.¹⁶⁸ First, Sprigman suggested revising Berne to replace the Article 5(2) formalities prohibition with a provision allowing certain "interoperable" formalities.¹⁶⁹ While Berne members would not be required to re-impose formalities, any countries that chose to do so would be required to permit foreign authors to instead comply with the formalities of either their home country or the country of first publication or registration.¹⁷⁰ The Berne signatory countries would also establish an international central copyright registry run by the World Intellectual Property Organization, and they would enter into a side agreement

158. See Ginsburg, *supra* note 7, at 1584–87.

159. See *id.* at 1587–88; Sprigman, *supra* note 7, at 488–91.

160. See Sprigman, *supra* note 7, at 489–90.

161. *Id.* at 495.

162. *Id.* at 496.

163. See 3 PATRY, *supra* note 16, § 9:6.

164. Sprigman, *supra* note 7, at 496–97.

165. See *id.* at 541–42.

166. *Id.* at 551–52.

167. *Id.* at 545–68.

168. *Id.* at 546, 551.

169. *Id.* at 546–47.

170. Sprigman, *supra* note 7, at 547.

standardizing the requirements and formatting of registration data.¹⁷¹ While this is an interesting proposal, it has very little chance of becoming reality—Berne requires unanimous approval of any amendments.¹⁷² One holdout nation can prevent any changes to the status quo.¹⁷³

If Berne could not be amended, Sprigman proposed an alternative set of “new-style formalities.”¹⁷⁴ Sprigman’s main unilateral proposal was to limit the remedies of copyright owners, including authors of foreign works not subject to the registration requirement, who did not comply with copyright formalities.¹⁷⁵ Instead, these owners would only be allowed to recover a nominal fee, equivalent to the cost of complying with formalities, as part of a statutory compulsory licensing scheme.¹⁷⁶ Sprigman viewed this proposal through a law and economics lens: an author who does not comply with formalities has sent a price signal, stating that they value the copyright in their work at less than the cost of compliance with formalities.¹⁷⁷ A default license set at the cost of compliance would thus promote efficiency by eliminating transaction costs and giving the author a good price, one that is not less than the author’s own expected revenue.¹⁷⁸

While Sprigman’s compulsory license proposal is interesting, it is ultimately unlikely to be effective or become law. Sprigman’s proposal presumes authors are rational actors who have set internal dollar values for their works.¹⁷⁹ But noncompliance with formalities does not necessarily signal any conscious opinions that the author might have about their work’s value. Instead, it could signal a general ignorance about copyright law. And actually finding and setting the correct default license fee to accomplish Sprigman’s aims is much easier said than done. Anti-formalities scholars also dispute whether this scheme would comply with Berne, as well as possible additional international treaty obligations to ensure available, effective injunctive relief.¹⁸⁰ As such, a proposal like Sprigman’s is unlikely to end up solving copyright’s ills.

The anti-formality academics, such as Professor Ginsburg, are skeptical of both pro-formality proposals and motivations.¹⁸¹ Ginsburg seems to place pro-

171. *Id.*

172. Berne Convention, *supra* note 49, art. 27(3).

173. *See id.*

174. Sprigman, *supra* note 7, at 554–55.

175. *Id.* at 555–56.

176. *Id.* at 555.

177. *See id.* at 556.

178. *Id.*

179. *See, e.g., id.* at 513, 555.

180. *See* Ginsburg, *supra* note 7, at 1593–97.

181. *See id.* at 1584–85.

formalities advocates into two subgroups.¹⁸² The first group seeks to use formalities as an excuse to neuter copyrights they dislike and erect hurdles for litigious authors.¹⁸³ Accordingly, she believes this group should be condemned.¹⁸⁴ The second group wants increased formalities to maintain a better public record of copyright claims, thus facilitating rights searching and clearance.¹⁸⁵ While Ginsburg sympathizes with this goal, she expresses concern over arguments by some in this group that a robust public record can only be built through the threat of copyright confiscation.¹⁸⁶ To this end, she proposes her own, voluntary measure to promote a complete public registry—a measure which may provide unexpected synergy with *Fourth Estate* and the registration approach.¹⁸⁷

Ginsburg argues that, regardless of intent, the pro-formality approach of limiting remedies to incentivize registration is likely to violate Berne.¹⁸⁸ Conditioning injunctive relief on registration would ignore the historical widespread availability of injunctive relief, which Berne presumably incorporated in its command to provide means of redress.¹⁸⁹ Furthermore, it could also violate the United States' obligations under the Agreement on Trade-Related Aspects of Intellectual Property Rights, which may require injunctions to, not just be available, but also be used and effective.¹⁹⁰ As for default licensing schemes like Sprigman's, Ginsburg contends that they would run afoul of Berne's three-step test for exceptions to exclusive rights, since they would not be a "special case[] that do[es] not conflict with a normal exploitation of the work."¹⁹¹

Despite Professor Ginsburg's general skepticism of new-style formalities, she was not categorically opposed to copyright formalities. In particular, she proposed her own formality to facilitate rights clearance: a requirement that transfers of copyright ownership be recorded with the Office to be given legal effect.¹⁹² In Ginsburg's view, this requirement does not violate Berne, since it merely controls who can enjoy and exercise the full panoply of Berne rights.¹⁹³

182. See Ginsburg, *supra* note 7, at 1584–88.

183. *Id.* at 1584–85.

184. *Id.*

185. *Id.* at 1585.

186. See *id.* at 1585–86.

187. See *id.* at 1613–21.

188. See *id.* at 1593–97.

189. See *id.* at 1593–94.

190. Ginsburg, *supra* note 7, at 1595–97.

191. See *id.* at 1594–95 (quoting Agreement on Trade-Related Aspects of Intellectual Property art. 13, Apr. 15, 1994, 1869 U.N.T.S. 299) (internal quotations omitted).

192. *Id.* at 1612–13.

193. *Id.* at 1611–13.

A transfer recordation requirement would be akin to the commonplace requirement that transfers of a copyright's exclusive rights be memorialized in a signed writing, which is generally believed to be compatible with Berne.¹⁹⁴ In fact, failure to comply could actually *benefit* the author-transferor by returning the rights to them.¹⁹⁵

Professor Ginsburg did acknowledge that her proposal, meant to ensure a complete and accurate chain of title, would have difficulty filling two major types of gaps.¹⁹⁶ First, under a voluntary registration regime, recording transfers of ownership means little without an initial starting point, a first link in the chain that can be confirmed valid.¹⁹⁷ Second, as the Office noted, oftentimes the ownership of a work can become muddled if the author dies, or if a corporate owner merges or goes bankrupt, thus leading to transfer by operation of law.¹⁹⁸ Ginsburg's main response to both points is to impose on the transferee of copyright an obligation to record both the transfer and the registration.¹⁹⁹

This response is not fully convincing. Transfers by operation of law can make many people joint co-owners, which could lead to a confusing state of affairs where nobody is sure who has to record. In addition, Ginsburg's proposal casts all the risks and obligations regarding chain of title on the transferee. Forcing otherwise willing buyers to shoulder the risk may scare them off, especially when there is even a hint of a cloud over the copyright's title, thereby depriving the author of revenue and the public of a useful work.

There is another, more fundamental flaw with Ginsburg's proposal. As a Congressional study recognized, the mere recordation of transfers, even if compelled by law, is of little value "[i]n the absence of a basic [copyright] registry system."²⁰⁰ Such a registry is vital to identifying the work covered by the transfer and deriving the full chain of title for the copyright.²⁰¹ Ginsburg's proposal to force grantees to record missing initial registrations will do nothing for copyrights that are never transferred in the first place. But perhaps *Fourth Estate* and the registration requirement can supply the registry with the information needed for Ginsburg's proposal to take full effect, thereby also

194. *See id.* at 1612–13.

195. *See* Ginsburg, *supra* note 7, at 1613.

196. *See id.* at 1616–19.

197. *See id.* at 1616.

198. U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS, *supra* note 9, at 28–29.

199. *See* Ginsburg, *supra* note 7, at 1616–18.

200. STAFF OF S. SUBCOMM. ON PATENTS, TRADEMARKS, & COPYRIGHTS, 86TH CONG., STUDY NO. 19: THE RECORDATION OF COPYRIGHT ASSIGNMENTS AND LICENSES 124–25 (Comm. Print 1960) (authored by Alan Latman).

201. *Id.*

lessening the burden on transferees to do something the author should have done in the first place.

Ultimately, Ginsburg's recordation requirement has significant gaps. But if even Ginsburg, no fan of formalities, finds this proposal agreeable, it will likely be approved by the pro-formality camp as well. The transfer recordation requirement can also have international effect,²⁰² unlike § 411(a), which explicitly only applies to U.S. works.²⁰³ The recordation of transfer requirement, together with the registration requirement, could thus serve as a consensus first step towards a registry that keeps perfect track of a copyright's chain of title, at least for U.S. domestic works.

V. **COULD *FOURTH ESTATE* ADDRESS ORPHAN WORKS BY ENCOURAGING PROMPT REGISTRATION?**

The ideal copyright registry would contain ownership and contact information for every copyright at every moment in time, from creation to expiry. There are two major ways a registry could fail to meet this standard. First, a copyright may be registered a significant time after its creation, or even never at all. Second, the registry might lose track of the ownership claims to an initially registered copyright due to subsequent transfers of ownership.

The strength or weakness of formalities plays a key role in determining whether the registry meets the standard or falls into the two traps mentioned above. Indeed, as part of its decision to keep § 411(a) on the books, Congress has recognized how formalities benefit the public and authors, both in general and with regards to the registry.²⁰⁴ Professor Ginsburg's proposal would address the second problem, and the registration approach to § 411(a) will help mitigate the first. Of course, little can be done about already-orphaned works. But even by itself, the *Fourth Estate* decision should help curtail future orphan works, at least for works created in the United States.²⁰⁵

Fourth Estate provides an answer to the problem of initial registration. The registration approach to § 411(a) will incentivize the author, the very first link in the copyright chain of title, to register their claim early. As for the issue of losing track of copyright ownership due to transfers, Ginsburg's proposal will address that issue by ensuring that valid transfers of ownership must be recorded. Ginsburg identifies several concerns with her proposal, of which the cost of recordation and the problem of gaps in the record are most relevant to

202. See Ginsburg, *supra* note 7, at 1620.

203. 17 U.S.C. § 411(a) (2018).

204. See H.R. REP. NO. 100-609, at 40-44 (1988).

205. § 411(a) only applies to U.S. works, not to foreign ones. 17 U.S.C. § 411(a).

this Note.²⁰⁶ The cost of recordation, much like the delay in registration processing, can be dealt with by taking the Supreme Court's suggestion in *Fourth Estate* and appealing to Congress.²⁰⁷ And initial gaps in the record can be dealt with through a combination of forcing transferees to record, and through *Fourth Estate's* incentives to register early. With those suggestions implemented, Ginsburg's proposal will ensure that the Office's records of transfer of ownership will correctly reflect subsequent owners of a copyright. And the registration approach will help the Office know about the first link in the chain of title. When put together, the result is a copyright registry that best facilitates rights-clearing and will minimize the chance of future orphan works.

One major objection to the effectiveness of *Fourth Estate* (and therefore to the effectiveness of the registration requirement combined with Professor Ginsburg's proposal) comes out of the *Cosmetic Ideas* opinion. Since the litigious copyright owner is obligated under either approach to deliver a complete application to the Office, the registration approach should be no more effective than the application approach at ensuring a complete registry.²⁰⁸

What this objection misses is the value of *timely* registration. Every day an existing copyright is not registered is a day where a potential licensee might search the registry and not find the copyright and its owner. An out-of-date registry puts the burden on every prospective user of a copyrighted work to search and re-search the records, in case the situation has changed, leading to unnecessary and redundant costs and perhaps even dissuading the prospective user from making beneficial use of the work. The Office itself recognized the importance of prompt registration long ago.²⁰⁹ The Office's report also demonstrates the value of registration in maintaining an accurate registry, noting that the elimination of the renewal registration formality, which required an author to re-register their copyright after twenty-eight years,

206. See Ginsburg, *supra* note 7, at 1613–14, 1617.

207. See *Fourth Estate Pub. Benefit Corp. v. Wall-Street.com, LLC*, 139 S. Ct. 881, 892 (2019).

208. *Cosmetic Ideas, Inc. v. IAC/InteractiveCorp*, 606 F.3d 612, 620 (9th Cir. 2010), *abrogated by Fourth Estate*, 139 S. Ct. 881.

209. See REGISTER OF COPYRIGHTS, 87TH CONGRESS, REP. OF THE REGISTER OF COPYRIGHTS ON THE GENERAL REVISION OF THE U.S. COPYRIGHT LAW 74 (Comm. Print 1961) (“To be most useful and reliable as a source of information, registration should be made shortly after the first public dissemination of the work.”). The Register recommended a three-month grace period after U.S. dissemination and a six-month one for dissemination abroad. All remedies would be available if registration was made within the grace period. But, if registration was delayed beyond that, then pre-registration infringement would only have limited remedies. *Id.* This recommendation impacts American copyright law to this day. See 17 U.S.C. § 412 (2018).

exacerbated the orphan works problem.²¹⁰ Of course, Congress had weighed this interest against the benefits of an automatic life-plus-fifty year copyright term, ultimately adopting the latter.²¹¹ But Congress has made the exact opposite determination about § 411(a)—it felt that the requirement of registration before suit would improve the copyright registry, and that § 411(a) should therefore be kept.²¹² The application approach may have resulted in copyrights eventually being registered, but *Fourth Estate* and the registration approach will encourage registration shortly after creation. And for that reason, *Fourth Estate* will help improve the copyright registry.

Another objection is based on technological progress. Yes, it says there was a problem with the paper registry of the past, and as a result we have an orphan works problem. But now that we have modern technologies, like computers, the risk of future orphan works should be much lower, since it should be easier than ever before to both find works and track their ownership.²¹³ Because of that, there is no need to make life harder for copyright holders, either by forcing them to wait for the Register before suing or by requiring transfers of ownership to be recorded with the Office. This critique has some force.

But even the most diligent tracker of copyright ownership claims cannot track claims whose existence it does not know. Better technology can address half of the problem by improving the Office's tracking of existing copyrights, but it cannot omnisciently scan existing works, especially works with no information in them tending to identify the author. Better technology is no silver bullet for an author's failure to register their copyright. For that, the Office needs incentives for authors to register their works, and *Fourth Estate's* adoption of the registration approach will provide a powerful incentive to do so.

Perhaps the strongest objection is related to non-sophisticated authors. Many smaller authors will probably not keep up with the latest in copyright law or Supreme Court decisions. The registration approach will just harm them when they show up at the courthouse without first complying with § 411(a). This could even create a two-tier copyright system, where sophisticated corporate users can enjoy the powerful remedies conditioned on prompt registration, such as injunctions and maximum statutory damages, while being

210. U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS, *supra* note 9, at 42–44.

211. *Id.* at 43.

212. *See* 5 PATRY, *supra* note 16, § 17:83.

213. *See* Michael W. Carroll, *A Realist Approach to Copyright Law's Formalities*, 28 BERKELEY TECH. L.J. 1511, 1534 (2013); Olive Huang, Note, *U.S. Copyright Office Orphan Works Inquiry: Finding Homes for the Orphans*, 21 BERKELEY TECH. L.J. 265, 278 (2006).

impervious to economically infeasible lawsuits filed by unsophisticated authors who did not register in time.²¹⁴

This objection can be broken up into two different subcomponents. First is an argument that small-time authors do not know about the registration requirement and should therefore not be held to it. But ultimately, getting a lawsuit dismissed for § 411(a) grounds (possibly in combination with a statute of limitations defense) is the kind of thing that should only happen at most once. Once the author realizes that their suit was dismissed, they should be cognizant of the need to promptly register any future copyrights. Knowledge can also be increased through efforts by the Office or other parts of the government to educate the public at large about copyright. Second, even if the small-time author knows that they should register promptly, the expense of registering each work they create, with an uncertain likelihood that the registration will even be worth it, is too burdensome. But fees, like delay, are a real-life concern that should be addressed by Congress, not through judicial rewriting of § 411(a). As such, the *Fourth Estate* Court has already provided the mechanism to address both arguments: appeal to Congress to increase the staffing and funding of the Library of Congress.

As a closing reminder, it should be emphasized that § 411(a)'s reach is strictly limited to U.S. works in order to comply with Berne. That means that the registration requirement, as well as *Fourth Estate*, will have very little to no impact on foreign orphan works. And the Office did note that foreign orphan works form “a large class of works for which locating the copyright owner is often very difficult.”²¹⁵ Unfortunately, motivating foreign copyright owners to register promptly will be more difficult than motivating domestic copyright owners, since foreign authors have stronger Berne protections against formalities. Perhaps work could be done here by the new-style formalities proposed by commentators such as Professor Sprigman. But the scope of this Note is limited to *Fourth Estate* and § 411(a), and neither of them have the power to mitigate the foreign orphan works problem.

VI. CONCLUSION

The Supreme Court has finally resolved a long-standing circuit split regarding when copyright registration is made. The immediate impact may be limited to a flurry of motions to dismiss in copyright cases that were filed in application approach circuits. But the reasoning in *Fourth Estate* makes it clear that the Supreme Court places a high, maybe even controlling, value on

214. See Tehranian, *supra* note 63, at 1403.

215. U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS, *supra* note 9, at 59.

statutory arguments over public policy arguments as they apply to copyright registration. Those unhappy with the current state of affairs or the current copyright law must direct their complaints to Congress, which can enact measures to alleviate their concerns.

In its ruling, the Supreme Court vindicated the registration requirement as a tool to facilitate a copyright registry useful to both authors and users. By doing so, the Court properly recognized and deferred to the rationale expressly cited by past Congresses as a reason to keep § 411(a) around. This stands in marked contrast to the application approach courts, who tended to specifically favor plaintiffs in their public policy analysis, with less attention paid to the public policy concerns favoring a strong registry and prompt registration.

The *Fourth Estate* decision will likely encourage not just registration, but *prompt* registration, which is an important first step in maintaining a copyright registry. To help keep better track of registered copyrights, Congress could enact a proposal akin to Ginsburg's, which would require copyright transfers to be recorded in the registry. This formality was previously a part of American copyright law and is one of few formalities that might be able to garner support from copyright scholars on different sides of the formalities debate. With a registry that both knows of claims earlier and keeps better track of ownership changes, the orphan works problem can be prevented from getting worse. American copyright law will have moved one step closer to a world where the formalities Congress has chosen to retain can have their full intended effect, a world hopefully without orphan copyrights.

KEEPING THE DMCA AWAY FROM FUNCTIONAL USE

Madison Bower[†]

I. INTRODUCTION

The Digital Millennium Copyright Act (DMCA) protects copyrighted works available in digital formats. It prohibits circumvention of digital locks that protect these works. This allows creators to make their works available for purchase on the internet without having to fear that file sharing will prevent them from profiting. The DMCA had the purpose of protecting copyright with the expansion of digital access, and it did not extend copyright protection to any works or aspects of a work that would not have been copyrightable under prior copyright. Unfortunately, the anticircumvention provisions of the DMCA have caused confusion over what copyright protects in functional works. This law has allowed makers of functional products, which would not otherwise warrant copyright protection, to sue for copyright liability in these products, when they have technological protection measures. As a result, the DMCA has inappropriately increased copyright protection beyond the scope of protecting an author's investment in a creative and original work.¹ In particular, the anticircumvention provisions have allowed makers of functional products, like calculators, appliances, and cars, to sue consumers who repair or modify those products.

Imagine you purchased a smart-home system, and you set it up to automate your thermostat and your outdoor lights; it opens your garage door when you get home from work, and you can now change your lighting, or lock your doors, with your smartphone. The price is \$299, but you expect it to be worth the investment due to the time and energy you will save. Then, the unexpected happens. The company that sold these systems announces it will sunset this product and completely shut down the background software service that makes the device function. You now have a useless \$299 plastic disc. This unfortunate scenario occurred for consumers that purchased a Revolv home system: Nest sunset the device just two years after acquiring the Revolv

DOI: <https://doi.org/10.15779/Z38KP7TS1R>

© 2020 Madison Bower.

[†] J.D. Candidate, 2021, University of California, Berkeley School of Law

1. *See, e.g.*, AARON PERZANOWSKI & JASON SCHULTZ, THE END OF OWNERSHIP 144 (2016); *Unintended Consequences: Sixteen Years under the DMCA*, EFF (Sept. 2014), <https://www.eff.org/files/2014/09/16/unintendedconsequences2014.pdf>.

company.² To make matters worse, even if you wrote your own operating system for that device, so that it could work after the product line was shut down, the company that made it could hold you liable for bypassing any digital locks to replace the original programming with your own, due to section 1201 of the Digital Millennium Copyright Act.³

To address this issue, this Note recommends two possible solutions. First, courts could analyze whether the work—in this case the software—was accessed to infringe copyright or if the work was accessed for a non-infringing use, such as accessing the work to make changes or to fix functionality in the product. Second, Congress should enact a permanent statutory exemption for non-infringing uses that would prevent owners from being liable for mere circumvention of a digital lock on items they have already purchased. This statutory revision would provide clarity in a way that the temporary exemptions for repair have been unable, and it would keep the law up to date when new technologies emerge that the temporary exemptions do not cover.

This Note begins by discussing the background of the DMCA anticircumvention provisions in Part II. Section III.A shows that prior to the DMCA, functionality in copyrightable works did not have copyright protection. Section III.B proposes reasoning courts could use when determining when circumvention violates section 1201. Then, as a more feasible solution, in Section III.C, this Note proposes a more expansive permanent exemption to allow non-infringing and fair uses under the DMCA anticircumvention provisions so that courts can reach consistent outcomes when adjudicating section 1201 violations. Part IV then discusses some additional reasons why the DMCA anticircumvention provisions should not be interpreted to provide copyright protection for functionality.

II. BACKGROUND ON DMCA SECTION 1201 AND COPYRIGHT LAW

A. DMCA SECTION 1201

The Digital Millennium Copyright Act (DMCA) addresses how copyrighted materials can lawfully be used in digital media and contains provisions which address liability for copyright infringement in the digital age.⁴

2. Kyle Wiens, *Nest to Brick Revolt, Pull the Plug on Smarthomes*, iFIXIT (June 17, 2016), <https://www.ifixit.com/News/nest-bricks-revolv>.

3. *Id.*

4. U.S. COPYRIGHT OFFICE, SECTION 1201 OF TITLE 17: A REPORT OF THE REGISTER OF COPYRIGHTS 1 (2017), *available at* <https://www.copyright.gov/policy/1201/section-1201-full-report.pdf> [hereinafter SECTION 1201]. The DMCA also was enacted to fulfill

Section 1201 of this act prohibits “[c]ircumvention of [t]echnological [m]easures” that protect copyrighted works.⁵ For example, a website or piece of software may be protected by a password, or encrypted, so that only those who have paid for access may use it. Breaking that encryption or password as a non-paying customer would violate section 1201.⁶

However, not all actions taken to get around such technological protection measures (TPMs) are prohibited by law. In fact, section 1201 lists several permanent exemptions, such as allowances for law enforcement and specific types of encryption research.⁷ These are part of the statute itself and were established by Congress “out of recognition of the importance of these activities.”⁸ Additional permanent exemptions can be passed by Congress, and unlike the temporary exemptions discussed below, permanent statutory exemptions are not limited to “specific classes of works” (such as sound recordings or computer programs for smart TVs).⁹ As a result, permanent exemptions provide better clarity to users engaging in an exemption activity, for devices that do not properly fit into one of the temporary exemption classes of works.¹⁰

The temporary exemptions are updated triennially; the statute itself does not set out the exact process for updating them, but the Copyright Office has established a procedure that allows interested parties to give input. Currently, the Office classifies the proposed exemptions roughly based on whether the copyrighted works are distributed similarly and whether the works have similar uses to which the exemption may apply.¹¹ The Copyright Office then administers three rounds of public comment submissions.¹² Parties in support

international treaty obligations: the WIPO Copyright Treaty and the WIPO Performances and Phonograms Treaty. *Id.*

5. 17 U.S.C. § 1201(a) (2012).

6. *Id.* § 1201(a)(1)(A).

7. *Id.* §§ 1201(e), (g). Other permanent exemptions include uses by non-profit libraries and educational institutions, uses by minors, and to protect personally identifiable information. *Id.* §§ 1201(d), (f), (h)–(i).

8. SECTION 1201, *supra* note 4, at 14.

9. *Id.* at 92–93.

10. *Id.*

11. KARYN A. TEMPLE, SECTION 1201 RULEMAKING: SEVENTH TRIENNIAL PROCEEDING TO DETERMINE EXEMPTIONS TO THE PROHIBITION ON CIRCUMVENTION, RECOMMENDATION OF THE ACTING REGISTER OF COPYRIGHTS, at 14 (2018), *available at* https://cdn.loc.gov/copyright/1201/2018/2018_Section_1201_Acting_Registers_Recommendation.pdf [hereinafter ACTING REGISTER'S RECOMMENDATION].

12. *See* SECTION 1201, *supra* note 4, at 24. For example, in the Seventh Triennial Rulemaking Process, the Copyright Office considered twelve classes including, “[l]iterary works – compilations of data generated by implanted medical devices – to access personal data,” “[c]omputer programs – ‘jailbreaking’ of smartphones, smart TVs, tablets, or other all-

of the exemption submit comments in the first and third rounds while parties which oppose the exemption submit comments during the second round.¹³ Those supporting the exemption have to show that the alleged harm from not granting the exemption is supported by a preponderance of the evidence.¹⁴ They must show that “users of a copyrighted work” will be “adversely affected . . . in their ability to make non-infringing uses . . . of a particular class of copyrighted works.”¹⁵ Non-infringing uses include fair use and other limitations on the scope of copyright protection afforded under the Copyright Act.¹⁶ In order to show that there will be adverse effects, the interested party must “demonstrate ‘distinct, verifiable, and measurable impacts’ occurring in the marketplace.”¹⁷ This burden of proof falls on the party proposing an exemption; opposed parties must then show that this demonstration of adverse effects was insufficient.¹⁸ The National Telecommunications and Information Administration (NTIA) and Acting Register of Copyrights both analyze the comments submitted and issue their own recommendations to the Copyright Office.¹⁹

Under section 1201, fair use and other limitations can influence which temporary exemptions are enacted in the Code of Federal Regulations. However, prior to the DMCA, non-infringing or fair uses did not need exemptions because it was understood that they did not violate the copyright holders’ rights. Prior to the DMCA, most courts allowed non-infringing uses of functionality within copyrightable works. In contrast, under the DMCA, some courts, as well as the Copyright Office, have held that non-infringing or fair uses still violate the DMCA if they do not have an exemption, and as a result, exemptions are required in order to decidedly avoid liability.

This issue has been addressed to some extent in the granted temporary exemptions, but the temporary exemptions alone are not enough. The current temporary exemptions do allow the repair of certain classes of items, but repair

purpose mobile computing devices,” and “[a]udiovisual works – educational and derivative uses,” to name a few. ACTING REGISTER’S RECOMMENDATION, *supra* note 11, at 23–27.

13. *See* SECTION 1201, *supra* note 4, at 24–25.

14. *See id.* at 26.

15. 17 U.S.C. § 1201(a)(1)(C).

16. ACTING REGISTER’S RECOMMENDATION, *supra* note 11, at 15.

17. *Id.* at 17 (citing Commerce Comm. Report).

18. *Id.* at 16–17.

19. NAT’L TELECOMM. & INFO. ADMIN., SEVENTH TRIENNIAL SECTION 1201 RULEMAKING: RECOMMENDATIONS OF THE NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION TO THE REGISTER OF COPYRIGHTS 1 (2018), *available at* https://www.ntia.doc.gov/files/ntia/publications/ntia_dmca_consultation_09252018.pdf [hereinafter NTIA RECOMMENDATIONS]; ACTING REGISTER’S RECOMMENDATION, *supra* note 11, at 1.

is still not allowed for all types of goods. The current repair exemptions only concern bypassing TPMs to diagnose or repair software that affects the functioning of what were, historically, “dumb items.”²⁰ As more devices contain software, it becomes more likely that one will need to access the software to make repairs when the device stops working, or malfunctions. Temporary exemptions can only apply to specific classes of works, so each Triennial Rulemaking Process’s repair exemptions are generally underinclusive.²¹ Additionally, the temporary exemption review process itself is burdensome for advocacy organizations to establish temporary exemptions that generally benefit the public.²²

In the 2018 Triennial Rulemaking Process, temporary exemptions for repair were an important focus. The Electronic Frontier Foundation (EFF) and several other organizations proposed a broader repair exemption which would include more Internet of Things devices, as well as video game consoles, because these devices often have some of the same repair issues.²³ The NTIA found that accessing software in phones and appliances for repairs would likely qualify as a non-infringing use of copyrighted work because it would be fair use.²⁴ Additionally, the NTIA found an adverse effect to consumers, due to the need to circumvent security measures in order to make repairs to devices one already owns.²⁵ The Acting Register agreed in recommending that smartphones and home appliances be included in the exemption to circumvention for repair and diagnostics.²⁶ However, video game consoles were not recommended to be exempt because widespread warranties and low-cost repair options from the manufacturers made it less likely users would experience adverse effects.²⁷ This conclusion and the recommendations

20. 37 C.F.R. § 201.40 (2016). Prior to the Internet of Things, home appliances, vehicles, and home systems did not have software, and, as a result, these types of items could not have had TPM protection nor fall under § 1201. PERZANOWSKI & SCHULTZ, *supra* note 1, at 140; Derek Russell Chipman, Note, *More Breaking, Less Rulemaking: Why Congress Should Go Beyond the Copyright Office’s Section 1201 Report and Amend the DMCA to Require a Nexus to Infringement*, 33 BERKELEY TECH. L.J. 1067, 1079 (2018).

21. *See* SECTION 1201, *supra* note 4, at 92–93.

22. *Id.* at 128–29 (“The commitment was keenly felt by individuals . . . [whose] participation in the rulemaking process competed with demands of their occupations, and by other communities, including the blind, visually impaired, and print-disabled, who have come to rely upon an exemption but must go through the process again each rulemaking.”).

23. ACTING REGISTER’S RECOMMENDATION, *supra* note 11, at 184–85.

24. NTIA RECOMMENDATIONS, *supra* note 19, at 51.

25. *Id.*

26. ACTING REGISTER’S RECOMMENDATION, *supra* note 11, at 216–17.

27. *Id.* at 220.

overall, though, ignore that bypassing TPMs for repair, which only involves the functionality of the device, should not be protected by copyright.²⁸

In the final ruling on October 26, 2018, the exemptions closely matched the recommendations from the Acting Register, instead of broader repair exemptions, which would have recognized the overreach of the DMCA into non-copyrightable items. A previous exemption for land-use vehicles was granted again, and an exemption was also granted to bypass TPMs to diagnose or repair smartphones, smart appliances, and smart home systems.²⁹ The exemptions do not mention, or extend to, other devices, such as video game consoles or computers.³⁰ EFF lamented that the current temporary exemptions still do not extend protections to all right-to-repair uses.³¹

Finally, because this is a temporary exemption, public interest organizations in support of this exemption will have to participate in the subsequent rulemaking processes triennially for as long as they want to advocate for this statutory exemption. The Copyright Office also recognized this burden, as well as the fact that the temporary exemptions only apply to specific classes of works, so temporary exemptions regarding repair will likely not adequately cover all types of devices, which might warrant circumvention in order to repair.³² Overall, the temporary exemption process failed to adequately curtail copyright protection of functional devices.³³

28. See *infra* discussion in Section III.B.

29. 37 C.F.R. §§ 201.40(b)(9)–(10) (2019).

30. See 37 C.F.R. § 201.40(b).

31. See, e.g., Mitch Stoltz, *New Exemptions to DMCA Section 1201 are Welcome, but Don't Go Far Enough*, EFF (Oct. 26, 2018), <https://www.eff.org/deeplinks/2018/10/new-exemptions-dmca-section-1201-are-welcome-dont-go-far-enough>.

32. SECTION 1201, *supra* note 4, at 92–93.

33. An additional concern in the anticircumvention law, with its current exemptions, is that the law violates the First Amendment. An ongoing case brought by the EFF on behalf of the plaintiffs alleges that the “threat of enforcement” of section 1201 “chills protected and non-infringing speech that relies on copyrighted works,” making the statute unconstitutional. Complaint at 2, *Green v. U.S. Dep’t of Justice*, 1:16-cv-01492-EGS (D.D.C. 2016), *available at* <https://www.eff.org/document/1201-complaint>. The EFF’s argument for how section 1201 burdens speech is that the granted exemptions have been too narrow to adequately address many non-infringing uses and therefore chills speech. *Id.* at 9–10. So far, the case has been allowed to continue past summary judgment on the issue of whether section 1201 substantially burdens more speech than necessary. *Green v. U.S. Dep’t of Justice*, 1:16-cv-01492-EGS, at 46 (D.D.C. June 27, 2019), *available at* <https://www.eff.org/document/green-v-doj-memorandum-opinion>. Notably, the plaintiffs in the case represent uses similar to modification and repair; Dr. Green publishes works on how to fix security flaws, and Dr. Huang develops technology for modifying digital video streaming. *Green v. U.S. Department of Justice*, EFF, <https://www.eff.org/cases/green-v-us-department-justice> (last visited Feb. 17, 2020).

The Copyright Office's current interpretation and exemption recommendations could lead to problematic results. The temporary exemptions for repair do not fully allow non-infringing uses, because the accessed work is entirely functional. However, court interpretations of the current DMCA law, or further revision of section 1201, could prevent such an issue and allow section 1201 to better align with existing copyright protections.

B. COPYRIGHT LAW GENERALLY

Copyright generally does not protect facts, functionality, or other technical operations. Instead, copyright law protects original works of authorship, which means that works must meet “a minimal requirement of creativity,” in addition to being independently created.³⁴ However, when a copyrightable work contains non-copyrightable functional or factual material, the copyright protection does not extend to the facts or functionality themselves.³⁵ The purpose of copyright protection is to promote progress and innovation by allowing creators to recover the time and cost invested in making a creative work.³⁶ Copyright protection of software and other similar works does not extend to any “idea, procedure, process, system, method of operation, concept, principle, or discovery.”³⁷ However, the expression and explanation of a “method of operation,” a process, or a system is protectable, while the general use of those abstract principles is not.³⁸

III. COPYRIGHT PROTECTION OF FUNCTIONAL WORKS IN CASELAW

A. CASE LAW PRIOR TO THE DMCA

The following two cases illustrate how underlying functionality is not protected by copyright, and they demonstrate how copyright protection has changed with the DMCA to include protection for non-copyrightable elements in copyrighted works.

1. *Baker v. Selden*

In *Baker v. Selden*, Selden, the author of a book on a particular book-keeping method, sued Baker for copyright infringement because Baker used the forms

34. 1 NIMMER ON COPYRIGHT § 2.01(B)(1).

35. *Id.* § 2.11(A).

36. Nicolo Zingales, *Digital Copyright, “Fair Access” and the Problem of DRM Misuse*, B.C. INTEL. PROP. & TECH. F. (2012), at 4, <http://bciptf.org/wp-content/uploads/2012/08/DRM-final.pdf>.

37. 17 U.S.C. § 102(b).

38. 17 U.S.C. § 102(b); 1 NIMMER ON COPYRIGHT § 2.11(A).

from the book, as well as the book-keeping system disclosed in Selden's account-books.³⁹ Even though Baker copied the forms from Selden's book, the Court did not find that Baker infringed on Selden's copyright because Baker did not copy them to put into his own publishable work but rather to use in his own book-keeping records.⁴⁰ The Court explained how far copyright protection extends in a work that explains facts or explains a functional system. The Court used the example of a book on mathematics: the copyright "cannot give to the author an exclusive right to the methods of operation which he propounds," but instead, the copyright only protects the author's book itself from being copied by others without permission.⁴¹ To have protection in the functional innovation and its use, the author of any such book would need to secure a patent instead (although here the book-keeping forms would likely not qualify for patent protection either).⁴²

This case draws a clear line in the amount of copyright protection available to books describing something functional or factual.⁴³ For copyright protection, functional aspects of a work can be treated in the same way as facts. The expression of the functions or facts in a creative way can have copyright protection, but the facts or functions themselves are not copyrightable. This reasoning can be extrapolated to any copyrightable work which expresses functionality.⁴⁴ The takeaway of the case is that Baker did not infringe "because he used the forms as *tools* and not as *works*."⁴⁵ In other words, "he copied the plaintiff's *invention*, not the plaintiff's *work*."⁴⁶

For example, the drawing in a patent file is, as an original work, copyrightable.⁴⁷ However, after the patent term expires, the patent owner cannot then sue for copyright infringement those who re-create the patented invention to exactly match the copyrightable drawings. Allowing the patent owner to sue for infringement via the drawings would provide a "back-door" extension of the patentee's monopoly.⁴⁸ Basically, one cannot use copyright to extend protection for innovation that does not qualify for intellectual

39. Baker v. Selden, 101 U.S. 99, 100–01 (1879).

40. *Id.* at 107.

41. *Id.* at 103.

42. *Id.*

43. ABRAHAM DRASSINOWER, WHAT'S WRONG WITH COPYING? 91 (2015) ("[T]he copyright is as operative and as ordinary as would a copyright in the text of a recipe or a work on mathematical science. . . . [However], the copyright in his work is not an exclusive right to the content expressed in that work.").

44. *See id.*

45. *Id.* at 93.

46. *Id.*

47. *Id.* at 105.

48. *Id.*

property protection (such as a mathematical function) or that could be protected by patent.

2. *Lotus v. Borland*

Building on this idea, the court in *Lotus v. Borland* found that using the same menu command hierarchy from a competitor's copyrighted software did not infringe Lotus's software copyright.⁴⁹ In this case, Borland released a spreadsheet program that used the same menu commands as Lotus's spreadsheet program, so that users of the Lotus program could switch to Borland's without having to re-do their macros or learn new commands to do the same functions.⁵⁰ The court held that Lotus's command hierarchy was a "method of operation" because one had to use the commands in order to make use of the spreadsheet program.⁵¹ As a method of operation, the command hierarchy could not be copyrighted.⁵²

The reasoning in this case demonstrates how even some of the creative aspects of a functional work, such as the words used to represent various computer functions, may not be copyrightable.⁵³ Although Borland could have used synonyms of the commands from Lotus, the benefit to users from standardization mattered more than protecting copyright of the program commands.⁵⁴ The expressive components of the method of operation do not change it from being a method or process, which cannot be copyrighted.⁵⁵

Copyright prior to DMCA section 1201 did not protect computer programs and other expressions of functionality or facts. Prior to the anticircumvention provisions, modifying or repairing software code usually would not have been found to violate copyright law because the modifications would be making use of the functionality or, even when copied to build on the existing program, because of the practicality of using standard options.

49. *Lotus Dev. Corp. v. Borland Int'l, Inc.*, 49 F.3d 807, 817, *aff'd*, 516 U.S. 233 (1996).

50. *Id.* at 810.

51. *Id.* at 815.

52. *Id.*

53. *Id.* ("The initial inquiry should not be whether individual components of a menu command hierarchy are expressive, but rather whether the menu command hierarchy as a whole can be copyrighted.")

54. *Id.* at 817–18 (finding that requiring users "to learn not just one method of operating the computer such that it prints, but many different methods . . . [is] absurd"); see Dan L. Burk, *Method and Madness in Copyright Law*, 2007 UTAH L. REV. 587, 591 (2007) ("Having learned one command structure, users would become confused or would incur the personal expense of relearning new commands each time they tried to change products. The benefits of standardization created a functional merger, if not an expressive merger.")

55. *Id.* at 592.

B. CASELAW ANALYZING SECTION 1201: REQUIRING COPYRIGHT INFRINGEMENT

Requiring copyright infringement in addition to circumvention of a digital lock would allow the DMCA's anti-circumvention provisions to further the progress of innovation and creativity. Such a requirement could either come from an established interpretation from the Supreme Court or from revision of the current statute. To allow repair or modification, courts could take a closer look at whether the accessed work was copyrightable, or if only functionality was accessed or used. If only functionality was accessed, then the courts could hold that the DMCA does not apply to the circumvention at issue. Alternatively, Congress could pass a permanent exemption for non-infringing uses, which would allow modification and repair, and this exemption could clear up confusion amongst circuits and result in more consistent decisions.

1. *The Circuit Split: Do Courts Require Infringement?*

Currently, interpretation of violation of the DMCA lacks consistency across circuits. Some courts have interpreted section 1201 to require copyright infringement, in addition to circumvention, in order for violation to have occurred, while others have determined that section 1201 only requires circumvention of a TPM protecting a copyrighted work in establishing liability.

a) *Chamberlain*: Requiring Copyright Infringement

In *Chamberlain v. Skylink*, the Federal Circuit held that the anticircumvention provisions “establish causes of action for liability” rather than a “new property right.”⁵⁶ Because the DMCA did not create a new property right, the plaintiff would need to show that the defendant lacked authorized access, meaning that the plaintiff would need to show that the defendant infringed on the plaintiff's copyright.⁵⁷ In this case, the plaintiff, Chamberlain, which sold electronically secure garage door openers, sued Skylink, a company which made a remote transmitter that could operate with their garage door opener system.⁵⁸ The transmitter “circumvented” the digital lock on the garage door system by broadcasting three signals which forced the system to operate in response to Skylink's transmitter.⁵⁹ From this circumvention alone, Chamberlain sued Skylink for violation of the DMCA; notably Chamberlain did not also sue for copyright infringement or

56. *Chamberlain Group, Inc. v. Skylink Techs., Inc.*, 381 F.3d 1178, 1192 (Fed. Cir. 2004).

57. *Id.* at 1193.

58. *Id.* at 1183.

59. *Id.* at 1185.

contributory copyright infringement, even though one would expect such claims to accompany a lawsuit concerning copyright law.⁶⁰

The Federal Circuit held that Chamberlain interpreted the scope of the DMCA's protections beyond what Congress intended.⁶¹ The court found such an interpretation conflicted with "the DMCA's statutory prescription that 'nothing in this section shall affect rights, remedies, limitations, or defenses to copyright infringement, including fair use, under this title.'" ⁶² The court further expressed concern that Chamberlain's interpretation of the anticircumvention provision would lead to aftermarket monopolies: "any manufacturer of any product [could] add a single copyrighted sentence or software fragment to its product, wrap the copyrighted material in a trivial 'encryption' scheme, and thereby gain the right to restrict consumers' rights to use its products in conjunction with competing products."⁶³ The court also worried that establishing liability through circumvention alone would substantially limit non-infringing uses as well as established fair uses.⁶⁴

b) *MDY Industries*: Section 1201 Creates a New Property Right

The Ninth Circuit, however, declined to follow the Federal Circuit's interpretation of the anticircumvention law. In *MDY Industries v. Blizzard*, the court found that MDY Industries violated section 1201(a)(2), (which prohibits producing a device designed primarily for circumventing a TPM) even though it did not infringe on Blizzard's copyrights.⁶⁵ In this case, MDY released a bot that automated play through early levels of World of Warcraft, a videogame owned by Blizzard.⁶⁶ Blizzard, aware of such bots and trying to prevent them, launched Warden, which prevented players that used bots, like the one MDY released, from connecting to the game's servers.⁶⁷ MDY then updated the bot to avoid detection from Warden, effectively "bypassing" this digital protection for purposes of section 1201.⁶⁸ Because MDY's bot did not "alter or copy WoW software," both MDY and users of the bot did not infringe Blizzard's copyrights.⁶⁹ As a result, the court did not find MDY to be liable for direct or contributory copyright infringement.⁷⁰ Even though no copyright

60. *Id.*

61. *Id.* at 1197.

62. *Id.* at 1200 (quoting 17 U.S.C. § 1201(c)(1)).

63. *Id.* at 1201.

64. *Id.*

65. *MDY Indus., LLC v. Blizzard Ent., Inc.*, 629 F.3d 928 (9th Cir. 2010).

66. *Id.* at 935.

67. *Id.* at 936.

68. *Id.*

69. *Id.* at 941.

70. *Id.* at 941–42.

infringement was found, the court still needed to decide whether MDY violated section 1201 of the DMCA, which relied on whether they chose to follow the Federal Circuit's interpretation.

The court in *MDY* declined to adopt the Federal Circuit's interpretation because subsections 1201(a)(1) and 1201(a)(2) fail to mention traditional copyright infringement.⁷¹ Additionally, in the Ninth Circuit's view, creating a new property right under section 1201 is consistent with the legislative history of the DMCA.⁷² The court addressed the Federal Circuit's interpretation, but found the "plain language" of the statute, as well as the policy rationale of needing to protect copyright owners with digitally stored work, outweighed the language of section 1201(c)(1) and the antitrust concerns brought up by the Federal Circuit.⁷³ Under this interpretation, the court found that MDY violated section 1201(a)(2) by selling this bot which circumvented Warden.⁷⁴ It is worth noting that here: Blizzard did not have to rely solely on the DMCA's anticircumvention provisions as a legal solution; MDY's actions also violated Blizzard's terms of service, so Blizzard was also able to bring a tortious interference claim.⁷⁵

2. *Doctrinal Recommendations*

The circuit split highlights the current tension in section 1201: on one side there is the language of section 1201(c)(1) and prior copyright doctrine, which prohibits copyright protection of functionality. Section 1201(c)(1) states that "[n]othing in this section shall affect rights, remedies, limitations, or defenses to copyright infringement, including fair use, under this title," which appears to say that section 1201 does not create new property rights and that copyright infringement is necessary for liability.⁷⁶ On the other side, the legislative history understanding of section 1201(a), as well as its plain language, indicate that one could be liable for circumvention alone. The analysis in the House Report states that "(a)(1) establishes a general prohibition against gaining unauthorized access to a work by circumventing a technological protection measure put in place by the copyright owner,"⁷⁷ and section 1201(a)'s language supports that view: "[n]o person shall circumvent a technological measure that effectively controls access to a work protected under this title."⁷⁸ However, the

71. *Id.* at 945.

72. *Id.* at 947.

73. *Id.* at 949–50.

74. *Id.* at 953.

75. *Id.* at 955.

76. 17 U.S.C. § 1201(c)(1).

77. H. REP. NO. 105-551, pt. 1, at 17–18 (1998).

78. 17 U.S.C. § 1201(a)(1)(A).

legislative history also indicates that TPMs in a lawfully acquired item may not be protected by section 1201, further supporting that repairing or modifying items should not violate the DMCA.⁷⁹

Both these cases failed to adequately analyze whether unprotected functionality or copyrightable material were accessed, to determine whether section 1201 should actually apply. But, the court in *Chamberlain* still reached the correct outcome by interpreting the law to require infringement. In some ways, the facts in each case may have influenced the outcome in each circuit and explain why the Federal Circuit took on such an interpretation. The policy concerns, which led to the Federal Circuit's requirement of circumvention being used for infringement, were readily apparent in *Chamberlain*. Anticircumvention and copyright protection in garage door openers looks much more like a misuse of copyright law, while the work accessed in *MDY*, a highly creative work and a video game, fits squarely into the type of work copyright law is meant to protect. Additionally, a competitor trying to interoperate and enter the market brings up antitrust concerns that do not exist in the context of a bot helping videogame players cheat in the game. Still, the Ninth Circuit should have required some form of copyright infringement. While the Ninth Circuit likely felt the need to punish a relatively bad actor, it did not need to use the DMCA's anticircumvention provisions to do so. Blizzard also had a tortious interference of contract claim, which more appropriately addressed the bad action at hand. Using copyright law to replace other forms of liability that are not related to intellectual property undermines other areas of law, as well as the legitimacy of intellectual property protection.

Legal scholars studying both cases have noticed that the Federal Circuit largely ignored both Congress' intent and the language of section 1201(a) when writing the *Chamberlain* opinion.⁸⁰ However, the court saw that strict adherence would result in substantial policy concerns, as well as a potential

79. H. Rep. 105-551, pt. 1, at 18. The legislative history states:

Paragraph (a)(1) does not apply to the subsequent actions of a person once he or she has obtained authorized access to a copy of a work protected under Title 17, even if such actions involve circumvention of additional forms of technological protection measures. In a fact situation where the access is authorized, the traditional defenses to copyright infringement, including fair use, would be fully applicable. So, an individual would not be able to circumvent in order to gain unauthorized access to a work, but would be able to do so in order to make fair use of a work which he or she has acquired lawfully.

Id.

80. See Zingales, *supra* note 36, at 12–13.

constitutionality issue.⁸¹ Because providing a new right of access would greatly change the balance between consumers and authors, the court instead found itself “resorting to good sense” rather than following the exact language of the statute.⁸² There is, however, another way to resolve the repair/modification prohibition issue without going against the language of the statute.

Instead, courts should go back to the jurisprudence of *Baker* and *Lotus*, which both hold that making use of functionality in a copyrighted work is not infringement because the functionality itself is *not copyrightable*; thus, the DMCA does not apply to those aspects of the work.⁸³ Even for courts which hold that section 1201 violations do not require copyright infringement, this analysis will lead to an appropriate outcome, because the DMCA, like the rest of the copyright statute, can only apply to copyrighted *works*, not the functionality.⁸⁴ As such, modifications or repairs of software that are functional in nature would not trigger section 1201, even if one bypassed a digital lock to get to the code at issue.

This type of ruling is best illustrated by *Lexmark v. Static Control Components*. In *Lexmark*, a printer company, Lexmark, had an authentication sequence which allows its printers to function with its printer cartridges.⁸⁵ Static Control Components (SCC) marketed a microchip that could mimic this sequence so that consumers could use cheaper, recycled cartridges instead.⁸⁶ Lexmark then sued SCC for copyright infringement under section 106 and for violation of the anticircumvention provision of the DMCA.⁸⁷ The appeals court, however, found that the Toner Loading Program, which the lockout code was meant to protect, was likely not a copyrightable work.⁸⁸ Because the program was a method of operation, it likely could not get copyright protection, even if there were other methods of operation which could accomplish the same thing.⁸⁹ Additionally, the court rejected the DMCA claim regarding access to the Printer Engine Program for two reasons. First, the program could alternatively be accessed by reading the code in question from the printer memory after purchasing a printer.⁹⁰ Second, the technological measure largely prevented the

81. *Id.* at 14. For a full discussion of the constitutionality issue, see *supra* note 33.

82. *Id.* at 15.

83. See *supra* Section III.A.

84. 17 U.S.C. § 1201(a)(1)(A) (“No person shall circumvent a technological measure that effectively controls access to a work protected under this title.”) (emphasis added).

85. *Lexmark Int’l, Inc. v. Static Control Components, Inc.*, 387 F.3d 522, 530 (6th Cir. 2004).

86. *Id.*

87. *Id.* at 531.

88. *Id.* at 541.

89. *Id.* at 540.

90. *Id.* at 546–48.

programming from running, which was a functionality issue. Because the TPM did not protect the code itself, no DMCA claim could be brought.⁹¹

In his concurring opinion, Judge Merritt called for a broader rule. He proposed establishing in law that “companies . . . cannot use the DMCA in conjunction with copyright law to create monopolies of manufactured goods for themselves,” which might have happened in *Lexmark* if the TPM had effectively blocked access to the Printer Engine Program or if the Toner Loading Program had more creative rather than functional elements.⁹² The concurrence highlights that the burden is currently on the defendant to show that its circumvention falls under one of the exemptions, rather than on the plaintiff to show that the defendant was accessing a copyrighted work, and this could lead to more established and sophisticated companies bringing lawsuits to quash rival companies due to the initial costs of litigation.⁹³ As Judge Merritt notes, such outcomes “clearly stifle rather than promote progress.”⁹⁴ If the court had not seen that the uses by the defendant were purely functional, the court could have found a DMCA violation. This shows the likely potential for a type of aftermarket monopoly misuse that many circuit court judges fear.

Focusing on the nature of the accessed work, as well as how it is used by the circumventer, would help courts analyze whether there is a section 1201 violation. If the work is generally not copyrightable, or if the defendant was making use of functionality, section 1201 cannot apply because the TPM in either scenario does not protect an applicable work under the statute. Using such a test, as outlined in *Baker* or *Lotus*, would allow courts to reach an equitable outcome while still following section 1201 without controversy.

Unfortunately, it is unclear how this interpretation issue will be resolved by remaining circuits or how and if the Supreme Court will issue an interpretation in the near future. The Copyright Office, thus far, has interpreted the statute to align with the Ninth Circuit’s interpretation.⁹⁵ Further, the Copyright Office and most courts have failed to constrain copyright protection of largely functional works appropriately, and rulings have largely ignored the issue of whether the protected work merited copyright

91. *Id.*

92. *Id.* at 551–52 (Merritt, J., concurring).

93. *Id.* at 552.

94. *Id.* at 553.

95. See SECTION 1201, *supra* note 4, at 42–43 (“The Office does not . . . believe enacting an infringement nexus requirement to be advisable, as it would severely weaken the right of copyright owners to exercise meaningful control over the terms of access to their works online—a right . . . essential to the development of the digital marketplace for creative content.”).

protection before commencing the circumvention and access analysis.⁹⁶ Before looking at circumvention, courts need to evaluate whether what was accessed is copyrightable material. Similarly, the Copyright Office's report on section 1201 indicates that the Copyright Office believes accessing software in a device for repair would violate section 1201 if no exemption were granted, even though accessing a device's software for repair involves accessing *functionality*, rather than copyrightable material.⁹⁷ Until copyrightability is regularly addressed in section 1201 cases, another solution will likely be needed.

C. STATUTORY REVISION

Instead of relying on court decisions, Congress could revise the DMCA to clarify limitations as to when anticircumvention provisions apply. Smaller revisions, such as a permanent exemption for repair of devices, have already been suggested by the Copyright Office.⁹⁸ A permanent exemption for all non-infringing uses could clear up the circuit split confusion because, then, infringement and copyrightability analyses would always be required in the court's analysis of section 1201(a) violations.

While there has been disagreement about the necessity of infringement, a permanent statutory exemption requiring infringement would be the correct outcome to promote the policy goals of copyright law and lead to consistent case outcomes. The DMCA has, in fact, encouraged the type of aftermarket protection through copyrightable schemes that the Federal Circuit feared. For example, John Deere has software controls in its newer farm equipment, which require farmers to bring the tractors to a John Deere dealer rather than fix the tractor themselves, or otherwise risk litigation under the DMCA.⁹⁹ This action by John Deere led to the 2015 right-to-repair exemption for farm equipment.¹⁰⁰ It has since been upheld in the Seventh Triennial Rulemaking Process in 2018 and remains as a current temporary exemption.¹⁰¹ As another lesser-known example, Texas Instruments (TI) threatened calculator hobbyists with legal action after they bypassed TPMs in TI's graphing calculators to install operating systems the hobbyists had made.¹⁰² So far, courts have largely failed to evaluate whether a defendant's use was only of functionality rather than

96. *See id.* at 38–40 (discussing different interpretations of how section 1201 applies to the Internet of Things).

97. *See infra* discussion in Section IV.A and notes 140–41.

98. SECTION 1201, *supra* note 4, at 94–95.

99. PERZANOWSKI & SCHULTZ, *supra* note 1, at 144.

100. *Id.* at 144–45.

101. 37 C.F.R. 201.40(b)(9).

102. *Unintended Consequences*, *supra* note 1.

copyrighted work. Passing this type of permanent exemption would prevent violations for functional uses because such uses are non-infringing.

Misuse of anticircumvention law can (sometimes) be resolved through the triennial rulemaking process. However, not all proposed non-infringing uses will receive temporary exemptions from the Copyright Office. For example, the farm equipment exemption somewhat successfully allowed farmers to go back to repairing their own equipment, if they so choose. But, on the other hand, other petitions to create an exemption for modification or tinkering have generally not cleared the Copyright Office's requirement that *all* such uses would likely be non-infringing in order to get a temporary exemption.¹⁰³

Even for the repair petitions which have succeeded, the process still includes plenty of costs that could be avoided if an infringement requirement were implemented instead. The rulemaking process requires considerable amounts of work by attorneys and other advocates, and this burden is incredibly high for public interest organizations and other groups with limited resources.¹⁰⁴ Due to these limitations, the Copyright Office has also recommended, during its examination of section 1201, that Congress enact a permanent exemption, but only for repair.¹⁰⁵ Because modification could potentially infringe on copyright owners' rights, the Office did not recommend it be included in a permanent exemption.¹⁰⁶ The Copyright Office also declined to recommend an exemption which would permit non-infringing or fair uses.¹⁰⁷ However, an exemption permitting non-infringing uses would definitively allow those accessing software for repair or modification to be free from legal liability, because access for functionality is a non-use under copyright law.¹⁰⁸ Implementing a permanent exemption for all non-infringing or fair uses and/or interpreting section 1201(c)(1) to require that circumventers are only liable when infringing will alleviate the burden on both the Copyright Office and public interest organizations in the triennial rulemaking process.

103. See ACTING REGISTER'S RECOMMENDATION, *supra* note 16, at 207–08.

104. See SECTION 1201, *supra* note 4, at 128 (providing an example from the Cyber Law Clinic, which estimated 575 hours dedicated to the most recent rulemaking); *Libraries Again Fight for Exemptions from "Digital Locks" Copyright Law*, DISTRICT DISPATCH (Nov. 11, 2014), <https://www.districtdispatch.org/2014/11/libraries-fight-exemptions-digital-locks-copyright-law/> (describing the process as “expensive and time consuming” for library organizations).

105. SECTION 1201, *supra* note 4, at 94–95.

106. *Id.* at 97.

107. *Id.* at 102.

108. DRASSINOWER, *supra* note 45, at 98 (“Selden’s copyright cannot reach Selden’s system because authors cannot claim as inventors. *Baker* is about the distinction between copyright and patent, works and tools . . .”).

Additionally, revision or re-interpretation of the anticircumvention statute could address the potential First Amendment concerns brought in the ongoing EFF litigation that asserts that the DMCA is unconstitutional.¹⁰⁹ Requiring the plaintiff to show a relationship between the defendant's circumvention and an attempt to infringe the plaintiff's copyright would keep the DMCA in line with existing copyright law, which has not been found to restrict or substantially burden speech. Many parties which oppose the infringement requirement would likely not want the DMCA to be found unconstitutional. Depending on how this litigation continues, the Copyright Office and other parties may push for the infringement requirement interpretation, even if the statute is not revised in the near future.

IV. OTHER JUSTIFICATIONS FOR ALLOWING MODIFICATION AND REPAIR UNDER THE DMCA

As discussed in Part III, courts have found that the consumer's interest in repairing, and to some extent tinkering, outweighs the intellectual property owner's interest in a complete monopoly over the making of the patented product for the granted amount of time.¹¹⁰ Recognizing why courts have taken this position is important in understanding why right to repair advocates push so strongly for this right, with respect to section 1201. Several justifications exist, such as sustainability, incremental innovation, and increasing social good and quality of life.

A. PATENT LAW AND RIGHT TO REPAIR OR MODIFY

To better understand how section 1201 is an overreach of copyright law, it is important to know how patent law, which does protect functional innovation, regards repair or tinkering. Patent law generally allows owners of an item, which contains patented functionality, to repair or modify it if such modification is for noncommercial, personal use.¹¹¹ These allowances are given to the owner of the tangible product or item, even though they infringe on the

109. *See supra* note 33.

110. *See, e.g.,* Aro Mfg. Co. v. Convertible Top Replacement Co., 365 U.S. 336, 346 (1961); Roche Prods. v. Bolar Pharm. Co., 733 F.2d 858, 862 (Fed. Cir. 1984) (quoting *Poppenhusen v. Falke*, 19 F. Cas. 1048, 1049 (C.C.S.D.N.Y. 1861)).

111. 5 CHISUM ON PATENTS § 16.03(1) ("A line of authority indicates that a defendant who makes and uses a patented product or process does not infringe if the use is for purposes of research or experimentation and not for profit."); *id.* § 16.03(3) ("[The] right to use includes a right to make repairs on the product necessary for continued use. Repair includes the replacement of parts under certain circumstances.").

patent owner's intangible right to the innovation, through the patent.¹¹² Owners of items which contain patented functionality can repair those items, so long as those repairs do not effectively replace the product.¹¹³ This means that as long as the repairs do not require one to completely re-construct the product, there is no patent infringement.¹¹⁴ Similarly, under patent law, one can make modifications and experiment with a piece of patented technology under certain narrow restrictions. The owner of an item composed of patented innovations has some leeway, so long as the tinkering is not done for economic gain but rather for “gratifying a philosophical taste, or curiosity, or for mere amusement.”¹¹⁵

The way in which repair and modification are allowed in patent law is particularly interesting, given the Copyright Office's interpretation of the DMCA. Without an exemption, the Copyright Office would hold that circumvention, in order to repair, would violate the DMCA.¹¹⁶ Under the Copyright Office's interpretation of the anti-circumvention provisions, the functional aspects of smart appliances and smart phones get copyright protection, which lasts far longer than the protection they would have received under patent law.¹¹⁷

Offering copyright protection to software functions in these devices could greatly stifle innovation. Patents expire after twenty years, yet offer strong protection of functionality.¹¹⁸ Copyright, on the other hand, lasts considerably longer (ninety-five years) but is meant to have no protection for functional elements and only protects the original work from copying, which is a weaker protection of the creative expressions of a work than what patent law protects in functionality.¹¹⁹ Under the DMCA anticircumvention provisions, some courts and the Copyright Office have allowed functionality to get patent-level protection with the span of time granted by copyright.¹²⁰ This level of

112. *Id.* § 16.03(1) (“Mere use of a patented product or process, even for purposes of personal convenience, ordinarily constitutes infringement.”); *cf.* citations, *supra* note 111.

113. *See Aro Mfg.*, 365 U.S. at 346 (“Mere replacement of individual unpatented parts, one at a time, whether of the same part repeatedly or different parts successively, is no more than the lawful right of the owner to repair his property.”).

114. *See id.* (holding that replacement “is limited to such a true reconstruction of the entity as to ‘in fact make a new article’”).

115. *Roche Prods.*, 733 F.2d at 862 (Fed. Cir. 1984) (quoting *Poppenhusen*, 19 F. Cas. at 1049).

116. SECTION 1201, *supra* note 4, at 92–93.

117. *Id.*

118. PETER S. MENELL, MARK A. LEMLEY & ROBERT P. MERGES, *INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE* 162–63 (2018).

119. 3 NIMMER ON COPYRIGHT § 9.11; 2 NIMMER ON COPYRIGHT § 8.01.

120. *See* SECTION 1201, *supra* note 4, at 30 (“Although the United States has consistently interpreted section 1201 as creating a cause of action separate and independent from copyright

protection for ninety-five years would be disastrous for innovation.¹²¹ Protecting functional innovation to a greater extent under copyright law than what patent law allows is incongruous with the policy goals of intellectual property law.

B. SUSTAINABILITY

One reason why allowing consumers to repair already-purchased devices is that it prevents unnecessary waste. When the owner of a device must take it to a certified repair shop anytime an issue persists, the device-seller has an effective monopoly on the repair market because the owner themselves cannot repair the device, nor can they choose an independent repair shop.¹²² At a certain price point, especially once a device or appliance is no longer under warranty, it becomes more attractive to consumers to simply buy a new version of the product and throw the damaged one into the garbage.¹²³ However, this line of action has had devastating results for the environment and as such, is a reason to favor allowing consumers to repair their own devices.¹²⁴

For example, the mobile phone manufacturing cycle has a massive annual carbon footprint (equivalent to the Philippines's entire yearly emissions), yet many consumers purchase a new phone every one to three years.¹²⁵ Using phones longer could reduce these emissions, but in order to reasonably accomplish this, consumers will need to be able to get repairs for smartphones at a fair price. Overall, a complete picture of what repair rights' impact would be on the environment has been difficult to assess due to a lack of comprehensive empirical studies.¹²⁶ However, it is generally accepted that re-

infringement, courts construing the statute to date have divided over its relationship to the traditional rights of copyright owners.”).

121. Chipman, *supra* note 20, at 1084–85 (discussing the DMCA's chilling effect on research and innovation).

122. Nicholas Deleon, *Right-to-Repair Laws Could Make It Easier to Get a Phone or Laptop Fixed*, CONSUMER REPORTS (Mar. 29, 2018), <https://www.consumerreports.org/consumer-protection/right-to-repair-laws-could-make-it-easier-to-get-a-phone-or-laptop-fixed/>.

123. Roger Harrabin, *EU Brings in 'Right to Repair' Rules for Appliances*, BBC (Oct. 1, 2019), <https://www.bbc.com/news/business-49884827>.

124. *Id.*; Catie Keck, *Right to Repair is Less Complicated and More Important Than You Might Think*, GIZMODO (May 10, 2019), <https://gizmodo.com/right-to-repair-is-less-complicated-and-more-important-1834672055> (“On top of this rigged cycle of predatory price gouging and consumer manipulation, there's another serious problem: electric waste. E-waste describes the electronic stuff that's no longer in use or no longer working The more stuff you buy, the more waste you're likely to generate.”).

125. *The Global Carbon Footprint of Mobiles*, RESTART PROJECT, <https://therestartproject.org/the-global-footprint-of-mobiles/> (last visited Feb. 17, 2020).

126. Cindy Kohtala & Sampsa Hyysalo, *Anticipated Environmental Sustainability of Personal Fabrication*, 99 J. CLEANER PRODUCTION 333, 333–34 (2015).

use and repair of digital products would improve sustainability to some extent over current consumer trends.¹²⁷ Allowing and encouraging repair is the more sustainable option, compared to granting monopolies in the repair industry through the misuse of copyright law.

C. INCREMENTAL INNOVATION

Sustainability, while a strong argument for repair, does not generally provide a justification for modification. One argument that favors modification and tinkering is that such activities will, over time, lead to further innovation and progress. Historically, many manufacturing improvements came about after the company purchasing and using the equipment found ways to make it work better for them.¹²⁸ Some have put forth arguments that today's technology is too complicated for such advancements to still occur.¹²⁹ But, it is worth examining whether that argument is actually true. It turns out that even for more complicated devices, there are still consumers that wish to modify or customize in a way that leads to innovation.

A more customized product for a small set of users is often unprofitable for most manufacturers.¹³⁰ A variety of reasons make such customization untenable. For example, the manufacturer may only specialize in certain component materials, such as carbon fiber, and is unable to accommodate changes that would use different resources.¹³¹ As another example, manufacturers of highly customized products would still need to ensure that each iteration of the product is free from defects or risk liability.¹³² In these contexts, user innovation is a natural result for those that want customization. Additionally, recent studies on user innovation in a variety of fields have shown that users on the leading edge of the technology will often make modifications and innovate.¹³³ This is because lead users will often notice what a product is lacking ahead of the rest of the users or consumers and can then make innovations that other users will also benefit from, even if such changes are relatively small.¹³⁴ In fact, the lightbulb, symbolic of sudden breakthroughs

127. *Id.*

128. ERIC VON HIPPEL, DEMOCRATIZING INNOVATION 21–22 (2005), available at <https://web.mit.edu/evhippel/www/books/DI/DemocInn.pdf>.

129. See generally Marissa MacAneny, Note, *If it is Broken, You Should Not Fix it: the Threat Fair Repair Legislation Poses to the Manufacturer and the Consumer*, 92 ST. JOHN'S L. REV. 331 (2018).

130. VON HIPPEL, *supra* note 128, at 51.

131. *Id.* at 48.

132. *Id.* at 50–51.

133. See *id.* at 23–30 (providing statistical examples of such innovation in PC-CAD users, libraries, extreme sports, and surgeons).

134. *Id.* at 22–23.

in innovation, was developed through small, incremental improvements over time.¹³⁵

In addition to the incremental innovation that can build on existing work, diagnostic and repair access to software systems could further enhance the safety of the existing software. Multiple car manufacturers have released vehicles with software flaws that lead to security risks from hackers or glitches that simply prevent the car from functioning properly.¹³⁶ Allowing consumers to bypass TPMs opens up the possibility of catching these errors with more people looking at the code.¹³⁷ TPMs have also prevented regulatory agency oversight from verifying that automobile and other types of manufacturers are complying with health, safety, and environmental laws.¹³⁸ This same concern can apply to many medical devices as well. Patients have found it immensely valuable to their quality of life to be able to modify devices, such as insulin trackers and pacemakers, to ensure that they can use the data these devices collect, and to ensure that such devices are not susceptible to security risks.¹³⁹

The types of innovation by users are likely to be small, incremental changes much of the time.¹⁴⁰ However, incremental innovation accounts for large amounts of technology progress, making it at least equally as valuable as more pioneering innovations.¹⁴¹ The potential for innovation combined with an increase in oversight for safety provides a compelling rationale for allowing users to modify or repair the products they own. However, section 1201 currently encourages the opposite.

135. See *The History of the Lightbulb*, U.S. DEP'T OF ENERGY (Nov. 22, 2013), <https://www.energy.gov/articles/history-light-bulb> (detailing the history of how the modern-day lightbulb was invented).

136. PERZANOWSKI & SCHULTZ, *supra* note 1, at 147.

137. *Id.*

138. See *id.* at 148 (providing an example of Volkswagen cheating its emissions restriction requirements through proprietary software).

139. See *id.* at 152–53; but see MacAneny, Note, *supra* note 129, at 341–42 (stating that allowing untrained parties to access and modify, or repair, medical devices will lead to a second-hand medical device market free from accountability). A second-hand market for medical devices would violate laws beyond just the DMCA, and it seems more relevant to address such a concern through healthcare law or the FDA.

140. See VON HIPPEL, *supra* note 128, at 21.

141. See William J. Abernathy & James M. Utterback, *Patterns of Industrial Innovation*, 80 TECH. REV. 40, 44, available at <https://teaching.up.edu/bus580/bps/Abernathy%20and%20Utterback,%201978.pdf> (“Though many observers emphasize new-product innovation, process and incremental innovations may have equal or even greater commercial importance.”).

D. COMMUNITY PARTICIPATION AND DEMOCRATIC THEORIES OF COPYRIGHT

Finally, one argument in favor of both repair and tinkering is that both these activities contribute to meaningful lives and enhance community building. Allowing modification and repair of functional articles is important, simply because “users like to innovate.”¹⁴² In a research study on user innovation, many user-innovators indicated that the innovation process “produce[d] learning and enjoyment that [was] of high value to them.”¹⁴³ Under Self-Determination Theory, most, if not all, people need their life to be engaging and active in order to have a satisfying life.¹⁴⁴ One such way this engagement can occur is through working and gaining expertise in certain technologies, whether as a job or as a hobby.¹⁴⁵ Particularly when modification occurs as a hobby, communities around the hobby result, allowing people to engage meaningfully with others who share a similar interest.¹⁴⁶ While this is not the only pathway to community building and decreasing isolation, it makes sense for law to help rather than hinder these types of outcomes.¹⁴⁷

This community building can have impacts beyond improving one individual’s quality of life. While one user may come up with a few incremental changes, the aggregate of their entire interest communities’ changes can both greatly increase the magnitude of innovation and the dissemination of those improvements.¹⁴⁸ Additionally, allowing people to modify mass-produced goods can help promote distributive justice by decentralizing innovation.¹⁴⁹

Furthermore, allowing consumers to tinker will not necessarily cause the economic harm that manufacturers fear. Many new products fail, and one rather large contributing factor to that failure is manufacturers’

142. William W. Fisher III, *The Implications for Law of User Innovation*, 94 MINN. L. REV. 1417, 1457 (2010).

143. VON HIPPEL, *supra* note 128, at 60.

144. Fisher, *supra* note 142, at 1469.

145. *Id.* (“When done at work, innovation helps make that work “meaningful” (in the Marxist sense); the innovator takes control of the tools of her trade and adapts them in hopes of doing the job better. When done outside of work, it makes for more active play.”).

146. *Id.* at 1470.

147. *Id.* at 1471–72 (“The activity of modifying mass-produced products is plainly not the only way in which one can live a good life”) (“[But if] the principal responsibility of the state is to create conditions that provide people access to rewarding lives, then . . . we ought to adjust the legal system to increase the ability of people to engage in [these] activities”).

148. VON HIPPEL, *supra* note 128, at 96. For examples of this type of community innovation, see *id.* at 97–104.

149. See *id.*; Fisher, *supra* note 142, at 1458–59; see also *id.* at 1462 (describing the failure of a green energy project from large manufacturers that then resulted in the successful development of “biomass charcoal” by the local community).

misunderstanding of what users want and look for in a product.¹⁵⁰ Allowing users to incorporate their own innovations into the products they purchased will both accommodate this failure in understanding what consumers are looking for and will attract new consumers who will “pay more for products with which they are permitted and able to tinker.”¹⁵¹

All of these policy goals (sustainability, incremental innovation, and community participation) drive the decisions judges made in patent law to permit repair and personal-use tinkering.¹⁵² However, the DMCA has not promoted these policy norms.

V. CONCLUSION

In all copyright cases, including alleged DMCA section 1201 violations, it is important to consider whether or not copyright infringement occurred; using functionality of a product you own should not implicate copyright liability. The anticircumvention provisions of the DMCA have made copyright more secure in the digital age, but they have also had the unintended effect of extending the scope of copyright beyond incentivizing the creation of original works. Instead, the prohibition on circumvention has hindered further innovation and creativity in industries which make use of software within functional items. These anticircumvention provisions in practice have allowed manufacturers of functional works to protect functionality and get patent-like monopolies for more than triple the amount of time that the patent system would grant.

Section 1201’s protections should not extend to the functional elements of these works because the functional elements are not in fact copyrightable material. Better recognition of the purpose of circumvention could prevent the excessive amounts of protection for functionality described above. Requiring copyright infringement in these types of cases is one way to reach outcomes which protect tinkerers and repairers. Another option is to evaluate whether the circumventer used only functional elements of the work instead of any of the copyrightable expressions, because access to works without copyright protection is allowed under section 1201.

Overall, a permanent exemption for all non-infringing uses, including fair use, would best align with the scope of copyright protection that both protects authors and also encourages others to build on that creativity or innovation. It would help courts reach consistent outcomes in section 1201 cases, and it

150. VON HIPPEL, *supra* note 128, at 107–08.

151. Fisher, *supra* note 142, at 1457.

152. *See, e.g.,* Aro Mfg. Co. v. Convertible Top Replacement Co., 365 U.S. 336 (1961).

would resolve the current circuit split. Allowing non-infringing uses to encourage repair and modification will both help people live a more fulfilling life and likely improve everyone's lives by furthering innovation and device safety.

GOING BEYOND THE MUSIC MODERNIZATION ACT: CREATION IN THE DIGITAL ERA

Shreya M. Santhanam[†]

I. INTRODUCTION

The internet democratized content generation. Gone are the days when only the music elite—those with connections to the industry and access to studios—could share their work with the masses. Now, anyone can create or recreate a song and share it with others, all they need is a microphone, recording device, and the ability to upload their creation onto the internet. Along with this, technology spurred the creation of mashup culture. Audio editing software has made it easy to rearrange and overlay multiple songs. Virtually anyone can create or share such works. However, regardless of the time and effort put into creating a work to share with society, if it contains copyrighted songs it can be immediately taken down by the copyright holders.

Copyright is the primary source of protection for music. Music copyright aims to balance incentivizing creation with compensating creators. Prior to 2018, artists relied on antiquated copyright laws that had not been greatly modified since 1976, to protect their interests. However, these laws were passed in an era where “the internet” and “streaming” were just science fiction. internet piracy disrupted this balance and the music industry struggled to adapt.¹ In response to the growing concern that music copyright needed to be brought in line with current technology, Congress passed the Orrin G. Hatch—Bob Goodlatte Music Modernization Act, or the Music Modernization Act (MMA).² It passed with bipartisan support in the House and Senate and was widely welcomed by various entities and individuals in the music industry.

DOI: <https://doi.org/10.15779/Z38C24QP1T>

© 2020 Shreya M. Santhanam.

† J.D. Candidate, 2021, University of California, Berkeley, School of Law. I would like to thank Erin Delaney, the participants of the Law & Technology Writing Workshop at Berkeley Law, and the Berkeley Technology Law Journal editors for their assistance on this Note.

1. See Paul Resnikoff, The Music Industry has 99 Problems. And They Are . . . , DIGITAL MUSIC NEWS (July 22, 2016) <https://www.digitalmusicnews.com/2016/07/22/music-industry-99-problems-2/>.

2. Orrin G. Hatch-Bob Goodlatte Music Modernization Act, Pub. L. No. 115-264, 132 Stat. 3676 (2018).

The MMA recalibrated the balance between compensation and creation with regards to copyright holders and music streaming platforms. It makes significant progress aligning copyright with modern music distribution, which better guarantees that artists receive accurate royalty payments when their music is accessed on platforms.

Despite the MMA's many gains, there are areas where it falls short. Notably, the MMA did not address how to deal with user generated content, such as covers and mashups, that may contain copyrighted music. Since these creators are not the original copyright holders, they need to license the copyrighted work, or it can be taken down for infringement. But, many people do not know they need these licenses, let alone how to secure them. These works may nonetheless be protected by the fair use defense to copyright infringement. However, as this Note discusses, fair use is a subjective standard, typically adjudicated, not in courts, but by technology companies that host content created by people on their platforms. Because covers and mashups fall within a copyright grey area, platforms are incentivized to deny creators this fair use defense. Consequently, these works are often removed, even if they should be considered "fair use."

This Note explains how the MMA fails to address how music copyright should apply to amateur creation and mashups in the digital age. Specifically, it argues that the law should adapt to allow non-commercial covers and mashups, that contain copyrighted works, to remain online. To accomplish this, there should be a blanket fair use exception for all non-commercial covers and mashups. Part II provides background on the development of music copyright law and the MMA. It also provides definitions for covers and mashups and the legal background surrounding them. Part III addresses fair use and how copyright is enforced in the digital realm. Part IV argues fair use should apply to non-commercial musical works that incorporate copyrighted music. Part V then explores the benefits of a blanket fair use policy, defines a framework, and addresses enforcement concerns.

II. HISTORY OF MUSIC COPYRIGHT LAW

A. PRE-MMA MUSIC COPYRIGHT

Changes in technology have necessitated amendments to music copyright, from its inception to the MMA. Songs receive two types of copyright protections: (1) for the musical composition, and (2) for the sound recordings based on the musical compositions.³ In 1831, music gained formal protection

3. PETER S. MENELL, MARK A. LEMLEY & ROBERT P. MERGES, *INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGY AGE* 671 (2019).

for the first time when musical compositions became protectable under copyright.⁴

Over the next century, music copyright laws evolved in reaction to the introduction of new technologies. For instance, concerns that the player piano could create a potential monopoly instigated the inclusion of mechanical compulsory licenses in the 1909 Copyright Act.⁵ Prior to mechanical licenses, larger companies could create exclusive licensing deals with music composers, effectively freezing out everyone else from using or performing those pieces.⁶ Mechanical compulsory rights allow anyone to produce sound recordings based on a musical composition, essentially by paying the copyright holder a royalty rate determined by law. In the 1920s, various agencies emerged to help administer these licenses and ensure rights holders were paid.⁷

Despite some dissatisfaction with the model, the compulsory license has been retained through subsequent copyright amendments, albeit with some modifications.⁸ For instance, the 1976 Act maintained it but increased the compulsory license rate to keep pace with inflation. Congress revisited the mechanical compulsory license again in the Digital Performance Rights in Sound Recording Act of 1995 (DPRA), as people started listening to music via digital audio transmission, instead of physical albums.⁹ The DPRA extended

4. *Id.* Prior to 1831, sheet music, and thus, musical compositions, were already protected against unauthorized copies under copyright law. William Ellsworth drafted a proposal to amend copyright. In the report he proposed that “the law of copy-right ought to extend to musical compositions, as does the English law.” The exact reasoning for why he decided to formally recognize music as a category to be protected is unclear. Oren Bracha, *Commentary on the U.S. Copyright Act 1831*, in PRIMARY SOURCES ON COPYRIGHT (1450-1900), (L. Bently & M. Kretschmer eds. 2008), available at www.copyrighthistory.org (last visited Feb. 9, 2020) (internal quotations omitted).

5. Howard B. Abrams, *Copyright's First Compulsory License*, 26 SANTA CLARA HIGH TECH. L.J. 215, 222 (2009).

6. *See id.* at 223–24.

7. The Harry Fox Agency, which still exists today, was one of the first agencies that helped administer licenses. *Id.*

8. The Register of Copyrights suggested the eradication of the statutory mechanical license for three reasons: (1) the royalty rate did not change with inflation, (2) the threat of a monopoly was the sole reason for the creation of the license, and (3) it was the fundamental right of a copyright owner “to have the exclusive right to control the commercial exploitation of his work” and this right should also apply to the recording of music. However, proponents of the license feared abolishing it would lead to exclusive licensing deals and lessen the number of covers. As a result, the public would have access to less variety of the same works, smaller companies would be unable to compete with larger ones by offering recordings of the same music, and creators would suffer because their work would not garner as much exposure and royalty revenues. *Id.* at 222 (internal quotations omitted).

9. 2 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 8.23.

the compulsory licenses to include Digital Phonorecord Deliveries (DPDs).¹⁰ A DPD is basically “each individual delivery of a phonorecord by digital transmission of a sound recording which results in a specifically identifiable reproduction by or for any transmission recipient of a phonorecord of that sound recording.”¹¹

As these changes demonstrate, the law mostly kept pace with the new listening formats, such as radio, cassettes, CDs, etc. At least until the internet hit. As the internet took hold globally, piracy became easier and more popular than ever. Websites like Napster enabled people to circumvent legal methods of obtaining music.¹² Napster allowed users to upload files, including copyrighted music. Anyone could download those files, effectively allowing people to illegally obtain music for free. Record sales drastically declined, and many artists struggled to profit off their work.¹³ The music industry filed hundreds of lawsuits against individuals who were downloading and distributing copyrighted music in hopes of preventing further piracy and declines in sales.¹⁴

Eventually, legal online streaming platforms, such as Spotify and Apple, emerged. Streaming services gained popularity for their ease of use, low cost, and huge music repositories.¹⁵ However, streaming services faced industry opposition for failing to pay artists accurate royalties.¹⁶ Various record

10. *Id.*

11. 37 C.F.R. 255.4 (2012). This definition came from the DPRA and remained unmodified until early 2017 when the Copyright Office removed the definition. 82 Fed. Reg. 9354, 9366 (Feb. 6, 2017) (removing and reserving 37 C.F.R. Part 255).

12. The music industry entered somewhat of a dark era. Peer-to-peer file sharing services such as LimeWire and Napster allowed for users to upload copyrighted music to the platform that could then be downloaded by other users. Eventually, multiple members of the Recording Industry Association in America sued Napster for copyright infringement. The Ninth Circuit held that the platform could be held liable, and it subsequently went out of business. *A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004, 1021 (9th Cir. 2001). In 2010, a district court issued a permanent injunction against LimeWire, to prevent further copyright infringement. *Arista Records LLC v. Lime Group LLC*, 715 F. Supp. 2d 481, 511 (S.D.N.Y. 2010).

13. See Eamonn Forde, *Oversharing: how Napster nearly killed the music industry*, THE GUARDIAN (May 31, 2019, 5:00 AM) <https://www.theguardian.com/music/2019/may/31/napster-twenty-years-music-revolution>.

14. Paul R. La Monica, *Music industry sues swappers*, CNN MONEY (Sept. 8, 2003), https://money.cnn.com/2003/09/08/technology/riaa_suits/.

15. Neil Howe, *How Music Streaming Won over Millennials*, FORBES (Jan. 16, 2019), <https://www.forbes.com/sites/neilhowe/2019/01/16/how-music-streaming-won-over-millennials/#353e86325c7c>.

16. In 2014, Taylor Swift pulled her albums from Spotify because she felt that the streaming service was not paying artists fairly. Fast-forward to 2017, and Taylor Swift’s catalog suddenly reappeared on Spotify without an explanation. Swift couldn’t have predicted how

companies sued Spotify, alleging that the company was not properly filing for compulsory licenses and, therefore, was committing mass copyright infringement.¹⁷ Despite these challenges with the compensation model, music streaming services have undoubtedly displaced traditional music technologies.¹⁸ Instead of records, individuals predominately use online services to access music.¹⁹ Streaming services helped change the way people consume music. Now, more people use these services, instead of piracy and illegal file sharing sites. This shift sent ripples throughout the music community by fundamentally altering the economics of the industry and the way artists engage with fans. Streaming services have become a necessity for artists to get paid, reach fans, and qualify for certain awards. Now, instead of record sales, touring and live performances account for a large percent of revenue generated by artist.²⁰

As streaming unseated incumbent technologies as the preferred music consumption method, paying artists accurate mechanical royalties became critical because this became one of the main ways artists were compensated.²¹ It was clear to the music industry, and to Congress, that the law was unable to address this technological change.

successful streaming services would become, and it just was not economically realistic to avoid these platforms. See Kaitlyn Tiffany, *A history of Taylor Swift's odd, conflicting stances on streaming services*, THE VERGE, (June 9, 2017, 11:50 AM) <https://www.theverge.com/2017/6/9/15767986/taylor-swift-apple-music-spotify-statements-timeline>.

17. Prior to the enactment of the MMA, Spotify settled several copyright infringement lawsuits. The suits alleged that Spotify failed to provide proper notice of intent and payment under § 115 of the Copyright Act. Eriq Gardner, *Spotify Settles Copyright Lawsuits Brought by Songwriters*, HOLLYWOOD REPORTER (June 27, 2019, 9:28 AM), <https://www.hollywoodreporter.com/thr-esq/spotify-settles-copyright-lawsuits-brought-by-songwriters-1221403>; see *Bluewater Music Serv. Corp. v. Spotify U.S., Inc.*, 2018 U.S. Dist. LEXIS 173064 (W.D. Tenn. Sept. 29, 2018) (denying Spotify's motion to dismiss for lack of standing and failure to state a claim).

18. See Juli Clover, *Streaming Music Contributed 75% of Total U.S. Music Industry Revenues for 2018*, MACRUMORS (Feb. 28, 2019), <https://www.macrumors.com/2019/02/28/streaming-services-music-industry-revenues-2018/> (discussing the Recording Industry Association of America's 2018 year-end music industry report).

19. See generally Benjamin M. Fly, *How Does Music Consumption Impact the Music Industry and Benefit Artists?* (May 2016) (unpublished B.S. Thesis, University of Arkansas) (on file with ScholarWorks@UARK, University of Arkansas) (an empirical study showing how music consumption has changed over time and how it has impacted revenue generation in the music industry).

20. *Id.*

21. Clover, *supra* note 18.

B. MUSIC MODERNIZATION ACT

The MMA is the most substantial copyright reform passed since the 1970s and is comprised of three previously proposed bills: (1) Musical Works Modernization, (2) Classics Protection and Access, and (3) Allocation for Music Producers.²² Each of the bills addresses a unique long-standing problem in the music industry. The first focused on challenges with the licensing system. The second aimed to provide federal copyright protection for pre-1972 sound recordings. The final bill provided music producers a direct method for receiving compensation for their contributions to musical recordings.²³ In 2018, Congress combined these three bills and unanimously passed the Act.

For the purposes of this Note, only the first title, the Musical Works Modernization Act (MWMA), is relevant. It authorizes the creation of a universal system for allocating copyright ownership for mechanical licenses of musical works. The new framework will ensure songs are properly attributed to copyright owners, enabling accurate compensation.²⁴ The regime aims to fix many of the problems involving copyright of sound recordings and DPD licenses. The MWMA creates a compulsory blanket license for whenever a digital music provider makes a permanent download, limited download, or authorizes interactive streaming.²⁵ In the old regime, streaming services had to license individual songs by serving a notice of intention on the individual copyright holders.²⁶ After serving the notice, they had to pay the applicable royalty fees to the copyright holders. Under the new regime, the blanket license will cover all works that are available for compulsory licensing, thereby eliminating the need to obtain a license on a song-by-song basis and decreasing transaction costs involved in deal making.²⁷ In order to provide flexibility to those in the music industry, parties will still be able to create voluntary licensing deals for songs in lieu of the blanket licenses.²⁸

22. Orrin G. Hatch-Bob Goodlatte Music Modernization Act, Pub. L. No. 115-264, 132 Stat. 3676 (2018).

23. *Id.*

24. Orrin G. Hatch-Bob Goodlatte Music Modernization Act § 101. The blanket licenses will go into effect January 1, 2021.

25. 17 U.S.C. § 115 (a)(4)(A)(ii)(I)(cc) (2018).

26. *Summary of H.R. 1551, the Music Modernization Act (MMA)*, COPYRIGHT ALLIANCE, https://copyrightalliance.org/wp-content/uploads/2018/10/CA-MMA-2018-senate-summary_CLEAN.pdf.

27. 17 U.S.C. § 115 (c)(2); Music Modernization Act Implementing Regulations for the Blanket License for Digital Uses and Mechanical Licensing Collective, 84 Fed. Reg. 49,966 (Sept. 24, 2019) (revising 37 C.F.R. § 201).

28. 17 U.S.C. § 115 (c)(2). Many streaming services are legally hosting copyrighted music on their platform through individual deals with record companies and copyright holders. The

Further, it authorizes the creation of a Mechanical Licensing Collective (MLC) to oversee the new licensing system. The MLC will collect and distribute the royalties and create and maintain a public database that identifies sound recordings and attribute ownership of the works.²⁹ The MLC will be funded in part by contributions made by digital music providers that are engaged in activities related to a blanket license and “significant nonblanket licensees.”³⁰

While the MWMA did not change the current compulsory licensing rates, it details a uniform rate-setting standard. This standard allows Copyright Royalty Judges to determine what reasonable rates would be in a free-market system.³¹ Current mechanical licensing rates are “9.1 cents or 1.75 cents per minute of playing time or fraction thereof, whichever is larger, for physical phonorecord deliveries and permanent digital” downloads.³²

While this Act lays the foundation for the MLC and blanket licenses for streaming platforms, there are several unanswered questions surrounding use of copyrighted music by individuals. Notably, the MMA did not address how the law should handle works created by amateurs that contain copyrighted music.

C. COVERS AND MASHUPS: WHERE THE MMA FELL SHORT

Covers are audio remakes of musical compositions, while mashups tend to be more complex. There are many different forms of mashups. A mashup may combine the lyrics of one song with the music of another, superimpose several songs over each other, or take segments of multiple songs and combine them to create a cohesive composition.³³ While artists can “legally” make covers by obtaining a mechanical or negotiated license, there is no similar licensing

MMA creates a compulsory license as a baseline. Previously, third parties could send copyright holders a Notice of Intent (NOI), which informed copyright holders of the third parties’ intention to record or distribute the copyrighted composition. Now, if someone uses copyrighted music without sending the proper NOI, instead of being liable for copyright infringement damages, they can pay the compulsory fee through the MLC. This decreases the potential for litigation and transaction costs. However, the blanket license does not prevent the creation of deals like those created in the pre-MMA era if the parties involved choose to do so. Music Modernization Act Implementing Regulations for the Blanket License for Digital Uses and Mechanical Licensing Collective, 84 Fed. Reg. at 49,966.

29. 17 U.S.C. § 114(f)(1)(A).

30. *Id.* § 115 (d)(7)(A)(i)(I)–(II). The MLC will be funded by several sources, including compulsory fees from digital music providers and voluntary contributions. *Id.* § 115 (d)(7).

31. *Id.* § 115 (d)(3)(D)(v).

32. Mechanical License royalty rates available at COPYRIGHT.GOV, <https://www.copyright.gov/licensing/m200a.pdf>.

33. Peter S. Menell, *Adapting Copyright for the Mashup Generation*, 164 U. PA. L. REV. 441, 457–58 (2016).

procedure in place for creating mashups. The creator must get permission from the rights holders of each of the songs they plan to use in their mashup. After the enactment of the MMA, there is still no formal regime streamlining the legal process for creating mashups that contain copyrighted work.

Several solutions have been proposed, including that Congress or the music industry implement an extended version of cover licensing where the fee is split up amongst all the songs sampled.³⁴ Alternatively, if there are so many songs sampled that it would be impossible to split up the fee equitably, the artist could be required to pay a fee based off the baseline royalty rate, to the MLC set up under the MMA. The MLC could internalize this cost and use it for regular operations.

The internet made music accessible to the masses. On top of the robust music industry, there are music hobbyists and bloggers who want to incorporate music into their works. Many people who are not in the music industry do not realize the nuances of music copyright, or recognize that they need to pay to use others' works. This is in part because the internet has greatly changed the way people view copyright.³⁵ People feel more entitled to use copyrighted works in ways that are not legally permissible.³⁶ As a result, people often create or upload works that incorporate copyrighted music without getting proper licenses. Due to the realities of copyright enforcement, these works are often treated as infringement and are taken down.

III. DIGITAL COPYRIGHT ENFORCEMENT AND THE ROLE OF PLATFORMS

While the MMA fixed some of the existing problems for legal streaming platforms, it does not address how to deal with other types of popular digital content that may infringe upon music copyright. In this system, platforms that

34. Menell proposes creating a remix compulsory license that is an extension of the cover license. *Id.* at 496.

35. See generally Milijana Mićunović, *Author's rights in the digital age: how Internet and peer-to-peer file sharing technology shape the perception of copyrights and copy wrongs*, 8 LIBELLARIUM 27–28 (2015). Mićunović conducted a study on the effect of technology on copyright violations. The study looked at 535 surveys and found that most respondents had used peer-to-peer file sharing, in violation of copyright. While perceptions of authorship and authors rights have stayed the same, due to the availability of digital production and consumption, users' attitudes and habits have changed. *Id.*

36. In the post Napster era, music copyright has maintained a bad reputation on the internet. Takedowns are frustrating for both creators and fans. Ryan Higa, otherwise known as “nigahiga,” was the most subscribed YouTuber from 2009–2011. A lot of the early videos and parodies he posted were taken down, due to copyright violations. At one point, his entire channel was taken down leading to frustration for him and his fans. Nigahiga, *Copyrighted*, YOUTUBE, <https://www.youtube.com/watch?v=FCTAqdDfL58> (last visited Feb. 9, 2020).

host user generated content, like YouTube or SoundCloud, largely act like the arbiters of online content because they are the de facto enforcers of the Digital Millennium Copyright Act (DMCA). The DMCA is one of the biggest barriers facing creators of mashups and covers.

Congress passed the DMCA in 1998 to combat copyright challenges in the digital era. The DMCA allows copyright owners to send online platforms takedown notices for content they believe infringes their registered copyrighted materials.³⁷ The DMCA contains safe harbor provisions that help absolve platforms of liability if infringing content is found on their websites, if they follow certain procedures. This Section will discuss how, as a result of this system, platforms must determine whether a work contains infringing materials and why they are incentivized to take these works down even if they fall under “fair use.” First, this Section discusses the DMCA safe harbor provision and how it incentivizes platforms to takedown works. Then it discusses fair use, with a particular focus on how it applies to music copyright, and operates on the internet.

A. DMCA § 512

The DMCA provides a series of safe harbor provisions, such as Section 512, to protect platforms from liability for every copyright violation that may occur on their site. These safe harbors protect service providers, such as YouTube, by providing immunity for transmitting, storing, or linking to unauthorized copyright content as long as they do not have “actual knowledge” or awareness of “specific and identifiable instances of infringement.”³⁸ For example, YouTube avoids liability under this standard in two ways: by (1) allowing copyright holders to flag videos for takedown and (2) using its own internal content identification service to flag potential violations.³⁹ YouTube then notifies the video’s uploader of the takedown, and the uploader can choose to appeal.⁴⁰

To ensure it receives protection under the Section 512 safe harbor, a platform will generally immediately remove content for which it received a

37. 17 U.S.C. § 512.

38. *See* *Viacom Int’l, Inc. v. YouTube, Inc.*, 676 F.3d 19, 32 (2d Cir. 2012); 17 U.S.C. § 512.

39. *For Studio Professionals: The Allocation for Music Producers Act (AMP Act)*, ADVOCACY, <https://www.grammy.com/advocacy/issues-policy/studio-professionals> [<https://perma.cc/J9HT-TR9S>].

40. *Id.*

takedown notice from a copyright holder.⁴¹ As a result, the takedown procedure is skewed in favor of the copyright holder, allowing the practice to be abused.

Some safeguards are in place to protect the second-generation creators. For example, copyright holders are supposed to issue takedown notices in good faith. But as the Ninth Circuit noted in *Rossi v. the Motion Picture Association of America*, good faith beliefs under the DMCA encompass a subjective, not objective standard.⁴² As part of this good faith assessment, copyright holders must consider fair use before sending a takedown notice. However, for the reasons described below, the application of fair use to music cases is not straightforward.

B. ROLE OF FAIR USE IN MUSIC COPYRIGHT

Musicians have borrowed elements from previous works for centuries.⁴³ Courts and Congress have recognized that in some circumstances this type of borrowing, or use of another's copyrighted work, is permissible through the doctrine of fair use. Fair use is a defense to copyright infringement that permits second-generation creators, under certain circumstances, to build upon copyrighted works without a license and free of liability.⁴⁴ For example, the Supreme Court has recognized that musical parodies "almost invariably copy

41. Copyright on the internet is enforced in tandem between rights holders and platforms. See Jennifer M. Urban, Joe Karaganis & Brianna L. Schofield, Notice and Takedown in Everyday Practice 17 (UC Berkeley Public Law Research Paper No. 2755628, 2017), <https://ssrn.com/abstract=2755628> [<https://perma.cc/3DYD-4MT8>].

Congress divided the burdens of compliance between OSPs and copyright owners. Congress placed on Internet intermediaries the burden of responding to valid takedown notices by "expeditiously" removing or disabling access to the identified allegedly infringing content. Congress placed the burden on copyright holders to identify infringing material because it considered that they know what material they own, and "are thus better able to efficiently identify infringing copies than service providers [. . .] who cannot readily ascertain what material is copyrighted and what is not." Courts have since affirmed that the DMCA notice and takedown provisions follow longstanding copyright law by "plac[ing] the burden of policing ongoing copyright infringement—identifying potential infringing material and adequately documenting infringement—squarely on the owners of the copyright." *Id.*

42. *Rossi v. Motion Picture Ass'n of Am., Inc.*, 391 F.3d 1000, 1004 (9th Cir. 2004) (addressing 17 U.S.C. § 512(c)(3)(A)(v)).

43. See Edward Lee, *Fair Use Avoidance in Music Cases*, 59 B.C. L. REV. 1874, 1890–92 (2018).

44. PETER S. MENELL, MARK A. LEMLEY & ROBERT P. MERGES, INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGY AGE 752 (2018) (referencing Neil Weinstock Nataneal, *Locating Copyright Within the First Amendment*, 54 STAN. L. REV. 1 (2001)).

publicly known, expressive works” but nonetheless are permissible as fair use.⁴⁵

To evaluate whether the goals of fair use are served by the work, courts turn to the four fair use factors. First, courts look at the “purpose and character of the use” which includes determining “whether such use is of a commercial nature or is for nonprofit educational purposes.”⁴⁶ When evaluating the purpose and character of a work, one must ask, “whether and to what extent the new work is ‘transformative.’”⁴⁷

For instance, in *Cariou v. Prince*, a well-known appropriation artist altered and used several of the plaintiff’s copyrighted photographs.⁴⁸ The portions of the photographs used, “and the amount of each artwork that they constitute[d] var[ie]d] significantly from piece to piece.”⁴⁹ The Second Circuit held that the works were fair use because, despite being used for commercial purposes, they were transformative in nature.⁵⁰ The court found the works to be transformative because they had a different character from the original work, gave the photographs a new expression, and employed new aesthetics.⁵¹ The works in *Cariou* did not present the same material in a different manner, rather they “added something new and presented images with a fundamentally different aesthetic.”⁵² Further, “[t]he more transformative the secondary use, the less likelihood that the secondary use substitutes for the original,” even if “the fair use, being transformative, might well harm, or even destroy, the market for the original.”⁵³

Next, when evaluating fair use, courts look at the “nature of the copyrighted work.”⁵⁴ The second factor is rarely determinative.⁵⁵ The third

45. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 586 (1994).

46. 17 U.S.C. § 107(1).

47. *Campbell*, 510 U.S. at 579 (quoting Pierre N. Leval, Commentary, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1111 (1990)).

48. 17 U.S.C. § 107(3).

49. *Cariou v. Prince*, 714 F.3d 694, 700 (2d Cir. 2013).

50. *Id.* at 706.

51. *Id.* at 708.

52. *Id.* (internal quotations omitted). Derivative works that “merely presents the same material but in a new form, such as a book of synopses of television shows, [are] not transformative. *Id.*”

53. *Castle Rock Entm’t, Inc. v. Carol Publ’g, Inc.*, 150 F.3d 132, 145 (2d Cir. 1998) (citing *Campbell*, 510 U.S. at 591–92).

54. 17 U.S.C. § 107(2). Courts are more likely to find fair use in cases where the underlying copyrighted work is factual.

55. *Davis v. Gap, Inc.*, 246 F.3d 152, 175 (2d Cir. 2001).

factor looks at how much of the copyrighted material the new work used, compared to the “copyrighted work as a whole.”⁵⁶

Finally, the fourth factor looks at “whether the copy brings to the marketplace a competing substitute for the original, or its derivative, so as to deprive the rights holder of significant revenues because of the likelihood that potential purchasers may opt to acquire the copy in preference to the original.”⁵⁷

However, as a multifactor test, fair use includes room for subjectivity which makes it difficult to universally apply. Drawing a line between acceptable uses and theft has never been clear. Adding to this problem, music copyright defendants and courts are often hesitant to turn to the question of fair use.⁵⁸ This may be in part because fair use is an affirmative defense. To invoke it, a defendant must admit they infringed upon a copyrighted work or plead it in the alternative.

Edward Lee theorizes several other reasons for why musicians and courts elect to avoid the fair use defenses outside of music parody cases.⁵⁹ Music parodies are generally considered “transformative,” and so, more cleanly satisfy the fair use test.⁶⁰ In other contexts, Lee found that most music copyright cases were resolved on non-infringement grounds, so the fair use question did not need to be considered.⁶¹ Instead of going to court, musicians within the industry who are accused of infringement often choose to settle by sharing song writing credits and royalties.⁶² Fair use is also a double edged sword for people within the industry; being able to freely borrow others’ works means also allowing others to borrow their own work.⁶³ Further, if the work is considered fair use it does not need to be licensed. Thus, the primary artist may lose out on revenue.

Lee also proposes a historical reason for the avoidance of fair use. The fair use doctrine in copyright emerged separately from, but around the same time as, music copyright protections.⁶⁴ Since the fair use doctrine was so new and

56. 17 U.S.C. § 107(3).

57. *Id.* § 107(4); *Authors Guild v. Google, Inc.*, 804 F.3d 202, 223 (2d Cir. 2015).

58. *See* Lee, *supra* note 43, at 1876–78 (noting the prominence of fair use defenses in other areas of copyright law and its limited use in music copyright).

59. *Id.* at 1874.

60. *Id.* at 1921; *see also* *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 577, 586 (1994) (holding that commercial parodies can qualify as fair use).

61. Only nine percent of all reported music copyright cases from 1987 to January 2018 mentioned fair use. Lee, *supra* note 43, at 1900–1903.

62. *Id.* at 1910.

63. *Id.* at 1915.

64. *Id.* at 1917.

not as robust as it is today, earlier music cases focused on permissible use instead of fair use.⁶⁵ The lack of precedent can explain some courts' reluctance to recognize fair use in non-parody music cases. Courts have "avoided th[e] difficult question of how transformative purpose and character can be applied to music borrowing outside of parodies."⁶⁶

The principles of fair use and their application to music sheds light on why it is failing online creators. In practice, both "good faith" and "fair use" require subjective determinations about permissible uses of copyrights. It is not in the interest of a copyright holder to objectively consider whether a work using their material is fair use. Likewise, because the availability of the Section 512 DMCA safe harbor hinges on service providers promptly responding to takedown notices, online service providers are not incentivized to standup for potential fair uses of copyrighted works. Consequently, creators are at a disadvantage. They are not part of the music industry, so they do not have access to the same strategies used by those in the industry to avoid infringement suits. Further, the counter notice process, which allows creators to respond to DMCA notices, is largely considered ineffective, again, because of service providers' reliance on Section 512 safe harbors.⁶⁷

Studies confirm that takedown notices are stacked against alleged infringers. A study that looked at over one hundred million automated DMCA takedown notices, found that over thirty percent of the notices were flawed or questionable.⁶⁸ In a separate study, the same researchers found that one in fifteen notices sent by rights holders had characteristics that weighed towards fair use. A large portion of these requests targeted "potential fair uses as mashups [or] remixes."⁶⁹ These findings reveal how covers and mashups that should be considered fair use, are not being treated as such. The copyright system needs to be modified to allow for permissive uses of copyrighted works to remain on the internet.

65. *Id.*

66. *Id.* at 1921 (explaining courts' reluctance to use a fair use analysis in non-parody music cases).

67. Urban et al., *supra* note 41.

68. Urban and her colleagues conducted three studies to determine how the DMCA and online copyright enforcement has evolved since its conception. The data presented came from Study 2, which examined random takedown notices (out of a sample of 108 million requests) that were generated by automated procedures. Further, one in twenty-five cases, or for 4.5 million requests, the "targeted content did not match the identified infringed work." *Id.* at 2, 11.

69. *Id.* at 95. The first study focused on qualitative data based on surveys completed by OSPs and rights holders. *Id.* at 1, 95. The study was under-inclusive in finding works that had characteristics of fair use. It did not consider musical covers when discussing works that had fair use elements. *Id.* at 162.

C. PRACTICAL APPLICATION OF DIGITAL COPYRIGHT MANAGEMENT TO USER GENERATED MASHUPS AND COVERS

The realities of digital enforcement, coupled with the subjectivity of fair use, have made it virtually impossible for platforms to apply fair use to online works. While an individual may upload a work that should fall under fair use, it may nevertheless be taken down. In practice, platforms lack the flexibility to consider these factors when they receive takedown notices for works that contain copyrighted material, because they can be held liable for “knowingly” hosting infringing content.

For example, the MMA has not changed the ways third party creators can use copyrighted music in their videos or mashups on hosting sites such as YouTube. While YouTube is one of the largest music streaming services in the world, it is also a host site where individuals upload their own content to channels. To ensure user compliance with copyright, YouTube provides users with a bank of songs that are “free” to use.⁷⁰ These songs and effects can be used without violating copyright. If a creator wants to use music outside of the bank, YouTube maintains a separate library that lists the current copyright policies for use of various songs.⁷¹ By using the directory and guidance, the creators can determine what steps they need to take to use the song without violating copyright; most often the creator must get an individual license to use the song.⁷²

If the creators do not follow the proper policies and upload something that incorporates copyrighted music, then the copyright holder may issue a takedown notice, or YouTube’s Content ID filter will flag the song for takedown. Once the user has received a takedown notice they can (1) do nothing and YouTube will remove the video, (2) remove and/or replace the music, (3) get a license, (4) in some cases allow the copyright holder to monetize the video and get ad revenue from the song, or (5) challenge the notice.⁷³ Currently, YouTube’s Content ID filter can block works containing copyrighted material a few moments after the video is uploaded onto a

70. See YouTube Studio, *Audio Library*, YOUTUBE, <https://www.youtube.com/audiolibRARY/music?nv=1> (last visited Feb. 9, 2020).

71. See YouTube Studio, *Music Policies*, YOUTUBE, https://www.youtube.com/music_policies?ar=1576251027059&nv=1 (last visited Feb. 9, 2020).

72. YouTube Creators, *YouTube Copyright Basics*, YOUTUBE, (July 1, 2013), <https://www.youtube.com/watch?v=Cp1Jn4Q0j6E> (last visited Feb. 9, 2020).

73. See generally YouTube About, *Copyright on YouTube*, YOUTUBE, <https://www.youtube.com/about/copyright/#learn-about-copyright> (last visited Feb. 9, 2020).

channel, which can further discourage creators from sharing their work.⁷⁴ In comparison, on SoundCloud, an online platform where creators can share original content and mashups of songs, compositions are often taken down if any portion of the song contains unlicensed copyrighted work.⁷⁵

Creators must navigate this complex system in order to share their work with the public. Even in seemingly innocent uses, where the song is not the focal point of the video, the work can be taken down for violating copyright. For instance, in the early days of YouTube, Universal Music issued a takedown notice of a video of a toddler dancing to Prince's song "Let's Go Crazy."⁷⁶ The case ultimately settled, but a Ninth Circuit panel held that under the DMCA, specifically 17 U.S.C. § 512(c)(3)(A)(v), copyright holders must consider fair use before sending a notification.⁷⁷ Here, the video contained twenty-nine unmodified seconds of Prince's song, but takedown notices can also be issued for modified works, like covers or mashups, that contain the copyrighted work without a license.⁷⁸

Applied correctly, fair use would cover these works. However, the system is not properly protecting them because platforms are incentivized to take them down, to avoid liability for hosting works that infringe copyright. The next Sections address what can be done to protect these works.

IV. MASHUPS AND COVERS CREATED FOR NON-COMMERCIAL USE SHOULD FALL UNDER FAIR USE

Parts IV and V focus on covers and mashups. This Part argues why non-commercial uses of mashups and covers should be considered fair use by default and, therefore, should not be taken down for copyright infringement nor subject to the mechanical licensing regime. Part V then discusses how such a policy can be implemented given concerns with digital enforcement, copyright holders, and second-generation creators.

74. Emily Hong, *What Beyoncé and Justin Bieber Taught Me About Fair Use*, SLATE (Jan. 25, 2016), <https://slate.com/technology/2016/01/what-beyonce-and-justin-bieber-taught-me-about-fair-use.html>.

75. *See Learn about Copyright*, SOUNDCLOUD, <https://soundcloud.com/pages/copyright#what-is-copyright-infringement-and-how-can-i-avoid-it> (last visited Feb. 9, 2020).

76. *Lenz v. Universal Music Corp.*, 815 F.3d 1145, 1149 (9th Cir. 2016).

77. *Id.* at 1151.

78. *See* *Viacom Int'l, Inc. v. YouTube, Inc.*, 676 F.3d 19, 26 (2d Cir. 2012). To avoid liability, YouTube allows copyright holders to flag videos for takedown and YouTube has its own internal content identification service to flag potential violations. YouTube then notifies the video's uploader of the takedown, and the uploader can choose to appeal or engage in the steps described earlier. *For Studio Professionals: The Allocation for Music Producers Act*, *supra* note 39.

A. SHOULD MASHUPS AND COVERS BE CONSIDERED FAIR USE OR SOMETHING ELSE?

This Section applies the four fair use factors to non-commercial uses of mashups and covers to determine if they should fall under the doctrine. The analysis is constrained by the fact that fair use is “an open-ended and context-sensitive inquiry.”⁷⁹ Not all mashups and covers are the same—they incorporate different amounts of the copyrighted work. However, at its core, the fair use analysis asks whether non-commercial mashups and covers are “the type of use that furthers the essential goal of copyright law and should be excused from liability for infringement”⁸⁰

Section IV.A.1, below, addresses why non-commercial uses of copyrighted music should be treated differently from commercial uses. Section IV.A.2 discusses in depth why mashups and covers are inherently transformative. Section IV.A.3 addresses the fourth factor, arguing that covers and mashups do not serve as a replacement for the original and are unlikely to usurp the market.

1. *Non-Commercial Uses of Copyrighted Works Should Be Given More Leeway Than Commercial Uses*

Use of copyrighted music online for non-commercial purposes should be treated differently than those used for commercial gain. The intent and actions taken by the creator must be considered when assessing if their use of copyrighted music is “commercial” in nature. People post videos on social media for a variety of reasons. For the purposes of this Note, these users fall into three primary archetypes: (1) those who use it as an outlet for their creativity, (2) those who hope to be discovered and become an “influencer,” and (3) those who are already “influencers,” that have sizable fan followings and make a profit off of their videos and online content. These three classes should be treated differently under fair use, as the non-commercial factor can be determinative. For the sake of this analysis, the creations being discussed are exclusively covers or mashups.

Those who fall in the first category, people who just want to share their work with others, should be considered “amateurs” who are not using music for commercial purposes. Since these creators are not profiting from the copyrighted work, they should be given the benefit of fair use. Those in the second group, who want to eventually use their work for commercial purposes, should be treated like amateur creators since they are not fiscally profiting from

79. *Blanch v. Koons*, 467 F.3d 244, 251 (2d Cir. 2006).

80. *Estate of Smith v. Cash Money Records, Inc.*, 253 F. Supp. 3d 737, 749 (S.D.N.Y. 2017).

the work.⁸¹ Finally, the works of influencers should not be considered fair use, as the creators are commercializing and profiting off the work of copyright holders..

Fair use seeks to balance the tension between incentivizing new creation and compensating copyright holders. Ultimately, fair use intended to preclude a finding of infringement where the goals of copyright are better served by use, rather than disuse, of the work.⁸² This is one such case. By taking the time to create a cover or create a mashup, amateur artists are injecting their own creativity into the piece, making it their own.⁸³ Here, the focus is on balancing second-generation creators' incentives to produce new works against ensuring primary creators receive appropriate revenue for their work.

2. *There Should Be a Presumption of Non-Infringement for Covers and Mashups, Under a Fair Use Analysis, Because They Are Transformative by Nature*

The first element of the fair use test favors findings of fair use for covers and mashups. All mashups and covers are transformative, to a certain degree. Remixes, and even just overlaying two songs over each other, requires a modicum of creativity beyond the original work. Even if the work incorporates a large sample of the song, by adding other elements and labor to it, the second-generation creator is contributing to the work beyond plain consumption and is transforming it. At times mashups and covers can be completely different takes and do not feel like the original song.

While there have been virtually no non-parody music cases that dealt with fair use,⁸⁴ other cases can give insight into why covers and mashups should receive its protections. For instance, in *Cariou*, the court found a series of collages to be transformative because they added something new and

81. Some use YouTube in hopes of becoming “YouTube famous.” Once they gain a certain number of followers, they can then monetize their videos. However, it is very challenging to become YouTube famous. Max Benator, *Can You Still Become YouTube Famous?*, VOX (May 5, 2015), <https://www.vox.com/2015/5/5/11562306/can-you-still-become-youtube-famous>; Creator Academy, *Make money on YouTube*, YOUTUBE, <https://creatoracademy.youtube.com/page/lesson/revenue-basics#strategies-zippy-link-3> (last visited Feb. 19, 2020).

82. *See* Castle Rock Entm’t, Inc. v. Carol Publ’g Grp, Inc., 150 F.3d. 132, 141 (2d Cir. 1998).

83. Fair use can be unpredictable where it potentially intersects with derivative work. The distinction before transformative and derivative works is part of the tension in using fair use in music copyright cases. “A fair use decision in one of the music cases would likely require greater theorization of the whole concept of transformative works and its relationship with and distinction from an infringing derivative work.” Lee, *supra* note 43, at 1923.

84. *See id.* at 1921.

presented the images with a new aesthetic.⁸⁵ Likewise, the popularity and special magic of mashups is how different they are from the songs they borrow from.

Much like covers and mashups that incorporate varied amounts of the original song, the collages in *Cariou* were transformative, despite the fact that “the amount of each artwork” used “var[ie]d significantly from piece to piece.”⁸⁶ Copyright law makes allowances for small degrees of copying, like the kind woven into many mashups. Under the *de minimis* doctrine, using small amounts of copying is okay.⁸⁷ While some mashups may incorporate very little of copyrighted songs, many covers and other mashups incorporate “substantially” more from the original works.⁸⁸ Works that use substantially more, versus a small amount, of a copyrighted work are less likely to be fair use. But the amount of use is not dispositive.

Often, the dispositive consideration is whether a challenged work is used commercially or is otherwise transformative in nature. Mashups and covers are typically transformative because they are far more than passive uses of copyrighted work. A passive use of copyright would be just uploading or downloading the original versions of a copyrighted work, as is, which courts have treated as infringement.⁸⁹ Just copying the work does not require any creativity or labor on the part of the creator. Unlike passive uses, covers and mashups are transformative, in some cases because they only incorporate small amounts of the original works, and in others, because the creator uses their

85. See *Cariou v. Prince*, 714 F.3d 694, 708 (2d Cir. 2013).

86. See *id.* at 700.

87. Menell, *supra* note 33, at 465. In *Bridgeport Music, Inc. v. Dimension Films*, the Sixth Circuit held that “no substantial similarity or *de minimis* inquiry should be undertaken at all when the defendant has not disputed that it digitally sampled a copyrighted sound recording.” 410 F.3d 792, 798 (6th Cir. 2005). This ruling effectively eliminated the plaintiff’s burden in music sampling cases to show substantial similarity and the defendant’s ability raise a *de minimis* inquiry. However, the Sixth Circuit indicated that a defendant may still prevail if they can establish fair use. *Id.* at 805.

88. Some cover artists may sound very similar to the original artists; in fact, there are tribute and cover bands that tour and perform exclusively other people’s works for fiscal gain. Since they are profiting off others’ works, it makes sense that they have to pay in order to perform these songs. However, covers posted online that sound like the original, but are not used commercially, should not be held to the same scrutiny, since they are just contributing more works to society. *But see* *Midler v. Ford Motor Co.*, 849 F.2d 460. In *Midler*, Ford tried to recruit famous musicians to sing popular songs of the time in an advertisement. When one of the singers refused to participate, they had a “sound alike” imitate the plaintiff’s voice as much as possible. The appeals court found that, although a voice is not copyrightable, the plaintiff still had a claim for the tort of appropriation under California law. Notably though, the “sound alike” was used for commercial purposes and with the sole purpose of impersonating or imitating the original artist.

89. See *BMG Music v. Gonzalez*, 430 F.3d 888, 890–91 (7th Cir. 2005).

creativity to combine and overlay the copyrighted elements in unique ways. For these reasons, covers and mashups routinely satisfy the first element of the fair use defense.

3. *Covers and Mashups Do Not Negatively Impact the Market for Original Songs*

The next fair use factor also cuts towards finding fair use for non-commercial covers and mashups. The minimal potential impact of mashups and covers on the market cuts in favor of considering covers and mashups fair use. “[T]he more the copying is done to achieve a purpose that differs from the purpose of the original, the less likely it is that the copy will serve as a satisfactory substitute for the original.”⁹⁰ Although covers and mashups take elements from other songs, they are still different from the original work. Courts have held that the more transformative the secondary work is, “the less likelihood” that it will serve as a replacement for the original.⁹¹

Ultimately, covers and mashups do not serve as replacements for the original song, but rather can help promote exposure of it. In the 1970s, record companies allowed radio stations to play songs for free because it was good for publicity and encouraged record sales. Similarly, now it is commonplace for new artists to share their music for free on TikTok in order to get exposure and develop a fan base.⁹² In the era of social media, where exposure is everything, use and proper attribution of copyrighted music can help boost the music industry. These derivative works can even lead to more streams for the original work, which would provide compensation to the original artists through the MMA’s blanket licensing scheme.⁹³ Many platforms already require mashups or covers to identify what songs they are sampling. Making this a hard legal requirement for creators would further the promotion of new music.

A natural concern with making covers and mashups fall under the fair use doctrine is that it may cut into the original artist and copyright holder’s profit. If non-commercial mashups or covers are considered fair use, there may be situations where the cover or mashup is more successful or popular than the

90. *Authors Guild v. Google, Inc.*, 804 F.3d 202, 223 (2d Cir. 2015) (summarizing a standard set out in *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 591 (1994)).

91. *Castle Rock Entm’t, Inc. v. Carol Publ’g Grp, Inc.*, 150 F.3d. 132, 145 (2d Cir. 1998) (citing to *Campbell*, 510 U.S. at 591–92).

92. Duncan Cooper, *How TikTok Gets Rich While Paying Artists Pennies*, PITCHFORK (Feb. 12, 2019), <https://pitchfork.com/features/article/the-great-music-meme-scam-how-tiktok-gets-rich-while-paying-artists-pennies/>.

93. Menell, *supra* note 33, at 462 (referencing Mark A. Lemley & Philip J. Weiser, *Should Property or Liability Rules Govern Information?*, 85 TEX. L. REV. 783, 827, 827 n.208 (2007)).

original work. In these cases, neither the primary and second-generation creators would be profiting from the work. But the platform hosting the content may be generating ad revenue from people visiting their site to listen to the mashup or cover. As a result, potential revenue may be diverted from the original artists to the platform.

There are already technological solutions to this concern. Currently, YouTube allows the rights holder to claim ad revenue for the work that contains their copyrighted work as an alternative to taking the work down.⁹⁴ If non-commercial uses of covers are fair use, copyright holders would no longer be able to monetize and keep the ad revenue from videos uploaded to YouTube.⁹⁵ The main value the copyright holder would get from the work would be from publicity. As described above, even if the engagement for the original work is lower than it is for the mashup or cover, this can lead to legitimate streams which will generate some revenue for the artists that they would not have gotten without publicity from the mashup or cover.

Further, copyright aims to provide incentives for artists to create, not necessarily to maximize their profits. Not fiscally compensating artists for non-commercial mashups and covers made by amateur artist will not take away all incentives for primary creators. Fair use tells us that the law should be less concerned with non-commercial uses, if it is not taking away the market for the original work.⁹⁶ Courts have also acknowledged that in cases where the work falls under fair use and is transformative, it is okay if it “harm[s], or even destroy[s], the market for the original.”⁹⁷ Even though there will be some loss in licensing fees, these works do not harm the original market at the level where it would disincentivize primary creators to continue creating.

Finding fair use for these works aligns with the policy basis of the doctrine—increasing the public’s access to works by promoting new, socially beneficial creations that use copyrighted materials. Therefore, there should be a presumption in favor of non-infringement for covers and mashups.

94. Currently, copyright holders can revenue share on videos that contain their copyrighted work. See YouTube Help, *How Content ID works*, YOUTUBE, <https://support.google.com/youtube/answer/2797370?hl=en> (last visited Feb. 9, 2020).

95. *Id.*

96. Courts are more likely to find fair use in cases where the work does not deter from the market for the work. See *Authors Guild v. Google, Inc.*, 804 F.3d 202, 223 (2d Cir. 2015).

97. *Castle Rock Entm’t, Inc. v. Carol Publ’g Grp, Inc.*, 150 F.3d 132, 145 (2d Cir. 1998) (citing to *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 591–92 (1994)).

B. IT IS SOCIALLY BENEFICIAL FOR NON-COMMERCIAL COVERS AND MASHUPS TO FALL UNDER FAIR USE

The four factors cut towards treating covers and mashups as fair use. But they rarely receive such treatment because platforms are incentivized to take down works challenged by a copyright holder. Even if users think their work falls under fair use, in practice, users rarely use the counter notice process. So, once their work is taken down, it stays down.

This Section discusses why this overinclusive takedown practice harms creators and stifles creativity and content generation for amateurs and upcoming artists. Therefore, treating all non-commercial covers and mashups as fair use would be socially beneficial.

1. *Takedown Notices Disproportionately Harm Amateur Artists and Discourage Them from Creating New Works*

Taking down non-commercial mashups or covers for copyright violations may disincentivize second-generation creators from producing or sharing their work with the public. Copyright law aims to incentivize creation, so allowing new artists to access the market should be prioritized. Aside from takedowns there are already safeguards in place on some platforms. YouTube, for example, allows copyright holders to monetize ads on covers instead of taking the video down. This enables copyright holders to be compensated for use of the work through the platform.⁹⁸ However, on other platforms, like SoundCloud, if creators use copyrighted music without permission, the entire track gets taken down.⁹⁹ If the file is a direct re-share of the original work, which happened often in the peer-to-peer file sharing era, the platform should take the work down to protect the copyright holder's interests. In contrast, mashups and covers should be allowed to stay up. Doing so encourages further creativity and increases public access to these works. Since these individuals

98. YouTube policy allows creators to monetize their channel through the YouTube partner program. Creators can earn money through various means, such as advertising revenue, by meeting certain requirements. See YouTube Help, *How to earn money on YouTube*, YOUTUBE, <https://support.google.com/youtube/answer/72857?hl=en> (last visited Feb. 9, 2020). On the other hand, SoundCloud allows users with premier accounts to monetize individual tracks they create. To qualify for monetization members must own all the publishing, masters, and distribution rights to the song. So, mixes, including those that are properly licensed, do not qualify. See SoundCloud 101, *What is the SoundCloud Creator Guide*, SOUNDCLLOUD <https://creators.soundcloud.com/guide> (last visited Feb. 9, 2020).

99. SoundCloud Help Center, *Your mashup was taken down for copyright infringement*, SOUNDCLLOUD, <https://help.soundcloud.com/hc/en-us/articles/115003563368-Your-mashup-was-taken-down-for-copyright-infringement> (last visited Feb. 9, 2020).

are not profiting off their creations through ad revenue, there should be a presumption in favor of fair use for these works.

Penalizing upcoming artists by taking down their work or flagging their channel for copyright infringement has real consequences. Many platforms have enforcement mechanisms that disproportionately impact online creators. On SoundCloud, just one copyright violation prevents the user from being able to monetize their channel in the future, creating an additional barrier for new artists to break into the industry.¹⁰⁰ YouTube has a similar policy.¹⁰¹ Even if these creators want to start commercializing and profiting from their original works, they will be unable to reap the full benefits of online platforms. Knowing they will not be able to profit off their content may dissuade them from generating their own original content. The consequence of losing all ability to monetize is grossly disproportionate to the harm of one allegation of copyright infringement. Even if creators make works that involve copyrighted material, fair use aims to incentivize forms of creation that can benefit society. When the penalty for infringement is so burdensome that it prevents new works, the goals of copyright are compromised.

Treating covers and mashups as fair use may cause copyright holders to lose out on some revenue, as the works would not be subjected to the mechanical licensing fee, but it will help encourage participation in the music industry. The access-incentive balance is arguably skewed in favor of the rights holder. Compensation under copyright is intended to reward and encourage artists to keep producing new works. Copyright seeks to balance creation with author's rights—economic benefits for authors are only half the equation.

Looking at copyright only through an economic lens may lead to the underemphasis of access and promoting creativity.¹⁰² Discussions around democratic theories of copyright focus on prioritizing access, and participation in culture.¹⁰³ Covers and mashups are cultural phenomena that allow people to

100. SoundCloud Help Center, *How do copyright strikes work?*, SOUNDCLLOUD <https://help.soundcloud.com/hc/en-us/articles/360005032554-How-do-copyright-strikes-work> (last visited Feb. 9, 2020).

101. YouTube has a copyright strike policy. The first one serves as a warning, and then users must go to copyright school. If you get three strikes, your account can be terminated, all your videos will be deleted, and you will be banned from creating new channels. YouTube Help, *Copyright strike basics*, YOUTUBE, <https://support.google.com/youtube/answer/2814000?hl=en> (last visited Feb. 9, 2020).

102. Oren Bracha & Talha Syed, *Beyond Efficiency: Consequence-Sensitive Theories of Copyright*, 29 BERKELEY TECH. L.J. 229, 233–34 (2014).

103. *Id.*

engage with music.¹⁰⁴ Encouraging a participatory society should not—and does not need to—be in direct conflict with economic incentives. Too much focus on the initial rights holders’ economic interests can smother emerging artists. Conversely, too little attention makes it difficult to transform musical talents into a living. Music helps encourage creativity and self-expression. Sharing new music with others should be prioritized. The law seeks to balance these competing aims, but has been stymied by platforms’ draconian takedown policies.

By retaining the compulsory license in the Copyright Act of 1976, Congress recognized the importance of covers in promoting new music at the expense of delinking control from compensation. Opponents of the compulsory license wanted to get rid of it to realign copyright law with other types of property laws by giving artists back the right to exclude others from using their work.¹⁰⁵ Lawmakers actively chose to keep the compulsory license and thereby showed that music copyright should not be treated as a pure property right. Congress recognized that compromising traditional property features was necessary to guard its twin interests in incentivizing new works and promoting access to the arts.¹⁰⁶ Like covers, mashups give consumers access to more variations of similar works. Technology now allows for broad access and use of songs. Covers and mashups are socially beneficial and finding fair use for these works would promote the goals of copyright.

2. *Allowing Mashups and Covers to Remain Online Can Help New Artists Break into the Industry*

Aside from increasing access to new works and encouraging creation, covers and mashups can help launch careers. Therefore, those who create covers and mashups with the intent of eventually monetizing their work should be treated the same as amateurs, until they start commercializing their works. YouTube and SoundCloud have helped launch several artists’ careers, such as Adele, Carly Rae Jepsen, and The Weeknd, to name a few.¹⁰⁷ Often, new artists will cover famous songs to get noticed. Once they develop their own following

104. Mashups are a cultural phenomenon. See generally John Shiga, *Copy-and-Persist: The Logic of Mash-Up Culture*, 24 CRITICAL STUD. MEDIA COMM. 93, 93–94 (2007) (discussing how the internet allowed for the emergence of mashup-culture); Menell, *supra* note 33, at 444.

105. Abrams, *supra* note 5, at 222.

106. *Id.* at 222–23.

107. Michelle Regalado, *9 Singers Who Came to Fame Online*, SHOBIZ CHEATSHEET (Dec. 12, 2020), <https://www.cheatsheet.com/entertainment/9-singers-who-came-to-fame-online.html/>.

or get signed to record companies, they begin to create their own work.¹⁰⁸ Or, if these artists already have original work, covering popular songs may allow the artist to gain notoriety and garner interest in their original songs. Other artists, like Chance the Rapper, got their start by releasing original work on SoundCloud.¹⁰⁹ Treating these artists leniently, when they are coming up, can help them gain a following, thereby encouraging the promotion and creation of new works by these artists.

3. *Works Used for Commercial Purposes Should Not Be Treated as Fair Use*

Artists who monetize their channel and profit off the copyrighted materials should not be treated as leniently as amateurs. These creators should have to go through the proper licensing process, whether it be using the compulsory license, or through a separate deal. As mentioned earlier, there are no official licensing regimes in place for mashups.¹¹⁰ If these creators are making a profit off the work, they should have to give a portion of the proceeds proportional to the amount sampled to the original holders.

On SoundCloud, this is easier to manage as individual tracks are monetized. YouTube also has a policy in place that allows monetized channels to turn off revenue on certain videos.¹¹¹ These creators would have the option to keep videos containing copyrighted work up by: (1) licensing and monetizing the video, or (2) choosing not to commercially benefit from the work.¹¹²

The four factors point to finding fair use for non-commercial covers and mashups. Treating these works as such is socially beneficial and is in line with the purposes of copy right law.

108. Justin Bieber, Shawn Mendes, and 5 Seconds of Summer are examples of musicians who got noticed for performing covers on social media. *Id.*

109. Chance the Rapper was propelled to stardom through online streaming platforms. Now several Grammy wins later he still uploads all his albums to SoundCloud and refuses to charge fans for his albums. In 2017, Chance the Rapper “saved” SoundCloud. It was rumored that the platform was going to shut down. Chance the Rapper tweeted that he was “working on the SoundCloud” thing and that it is “here to stay” after discussions with the CEO. *See* Ashley King, *After Helping to Save SoundCloud In 2017 Chance the Rapper Hits 1 Billion Streams on the Platform* (Aug. 17, 2019), <https://www.digitalmusicnews.com/2019/08/17/chance-the-rapper-soundcloud-billion/>; John Lynch, *Chance the Rapper says SoundCloud is ‘here to stay,’ after a report says its cash will run out in 50 days*, BUSINESS INSIDER (July 14, 2017), <https://www.businessinsider.com/chance-the-rapper-soundcloud-here-to-stay-amid-financial-struggles-2017-7>.

110. *See supra* Section II.C.

111. Creator Academy, *supra* note 81.

112. *Id.*

Properly understood, fair use should insulate creative works like mashups and covers from abusive copyright takedowns. However, for the reasons discussed above, fair use has not adapted to the new challenges created by the internet. The current system needs to be changed to allow for these works to remain accessible to the public.

V. DEFINING A FRAMEWORK

Creating a statutory exception for non-commercial mashups and covers is one way to ensure that fair use is applied *fairly* to non-commercial covers and mashups. After all, fair use is judged by platforms. And the safe harbor provisions encourage platforms to side with the rights holders to avoid liability.¹¹³ As a result, second-generation creators' rights are stifled. Something needs to change to incentivize platforms to modify their enforcement mechanisms. Creating a legislative carve-out will help recalibrate fair use to accommodate the realities of the internet and encourage platforms to consider second-generation creator's rights without fear of legal repercussions.

A. PROPOSAL

This proposal aims to recalibrate the balance between increasing access to new works and protecting rights holders' interest. The statute needs to balance the interests of the copyright holder, the second-generation creator, and platforms. It must create a statutory, rebuttable presumption that mashups and covers are fair use. The presumption for fair use exists to prevent the automatic takedown of mashups and covers and ensure fair use is being fairly applied. It has to be rebuttable to protect against works that are incorrectly, or falsely, labeled as mashups and covers, but are unmodified uploads of the copyright holder's work. Creators need to signal to platforms and copyright holders that their work is a cover or mashup, and it is not being monetized. For this to be successful in practice, the liability placed on platforms needs to change. Platforms should not be held liable for having a good faith belief that the work is a cover or mashup, if it turns out to not be one. If they can be held liable, they will still be legally incentivized to continue an over inclusive practice of taking down works that should fall under fair use. By not holding them liable for a good faith belief, they will not be penalized for allowing covers and mashups to remain on their websites, shifting the balance towards promoting access.

113. *See supra* Part III.

1. *Presumption for Fair Use for Non-Commercial Mashups and Covers*

Under this proposal, there would be a presumption of fair use for non-commercial mashups and covers. To qualify, first, the work must be non-commercial, meaning the second-generation creator cannot be deriving monetary value from the work. Second, it must be a mashup or a cover. It cannot be direct uploads of a copyrighted work. To qualify as a cover, and not a direct copy, the creator must either generate the vocal component, instrumental component, or both, in their song. As iterated earlier, covers and mashups are inherently transformative because users actively contribute and modify the work; it is not passive consumption. Further, they are not meant to replace the original song, but can boost interest in it.

Under the proposed policy mashup artists will not have to get licenses for every song they sample for non-commercial works. To ensure copyright holders receive some benefit from the secondary work, the second-generation creators must indicate and attribute all the copyrighted works included, as to not confuse the sources. Crediting the original artist may encourage people to listen to the original source, thereby providing some compensation for the copyright holder.

Crucial to this is that the cover or mashup does not greatly impact the market for the original songs, or that the second-generation creator fiscally profit off the labor of the original creators. Therefore, the mashup and cover must be created and shared for non-commercial purposes. The combination of these factors supports a presumption of fair use and should be automatically given to non-commercial mashups and covers.

2. *Rebuttable Presumption*

Just claiming the work is a cover or mashup is not enough to allow a work to remain online. Rights holders cannot be deprived of their ability to enforce their copyright on works that do not fall under the fair use presumption. Copyright holders should be able to flag unauthorized uses of their work and specify why the work is not a cover or mashup. If they merely flag the work as infringement, it would be no different from the current system, and a specification requirement will prevent frivolous notices.

To satisfy the specification requirement, rights holders would need to indicate what aspect of the work is infringing their copyright in an impermissible way. For instance, they would be able to indicate if the work is a direct upload of their work, was being used for commercial purposes, or did not properly credit the underlying work to the rights holder.

3. *Good Faith Exception*

Currently, platforms act as strict gatekeepers and takedown essentially anything that may implicate copyrighted music to prevent liability.¹¹⁴ This proposal will prevent platforms from taking down videos that use copyrighted works that should be covered by fair use. SoundCloud’s policy directly states that any use of copyrighted work, even for non-commercial use is not fair use, and it will take down any infringing works.¹¹⁵ Creating a good faith exception for platforms would prevent them from over-policing content through such liability shifting techniques.¹¹⁶

By requiring creators to identify their works as mashups or covers upfront, platforms are put on notice that the works contain copyrighted materials that the second-generation creator does not own. Currently, platforms can be held liable under the DMCA if they have “actual knowledge” or awareness of “specific and identifiable instances of infringement.”¹¹⁷ Under existing law, it is possible for platforms to be held liable if it hosts something that was labeled as a non-commercial mashup or cover that turns out to not be one, because it can be argued that the platforms had “actual knowledge” that a specific work contained or could contain infringing content. Therefore, platforms would still be incentivized to takedown works. To prevent this behavior, platforms should not be held liable for having a good faith belief that the work was a mashup or cover. The good faith belief would come from users self-reporting the status of their work. If later a copyright holder flags a video for not falling within the permissible uses of copyrighted music, the good faith requirement would alleviate liability concerns for platforms.

However, platforms would still have to maintain filters. Internally, filters would have to be modified to distinguish between monetized and non-monetized channels when running their algorithm. The concept of fair use is hard for an algorithm or filter to detect, since they contain copyrighted work,

114. Julia Alexander, *YouTubers and Record Labels are Fighting, and Record Labels Keep Winning*, THE VERGE (May 24, 2019), <https://www.theverge.com/2019/5/24/18635904/copyright-youtube-creators-dmca-takedown-fair-use-music-cover>.

115. SoundCloud Help Center, *Best Practices for uploading a track under fair use*, SOUNDCLOUD, <https://help.soundcloud.com/hc/en-us/articles/115003445707-Best-practices-for-uploading-a-track-under-fair-use> (last visited Feb. 9, 2020).

116. YouTube’s Content ID filter can be employed to immediately flag and block videos that may fall into a “gray” area of fair use, such as fan videos and mashups. For instance, a creator tried to upload a video combining the audio from Justin Bieber’s song *Sorry* and dancing from Beyoncé’s music video for “Grown Women.” The video was almost immediately flagged by Content ID. YouTube blocked the video, and the creator specific channel privileges. Hong, *supra* note 74.

117. *See Viacom Int’l, Inc. v. YouTube, Inc.*, 676 F.3d 19, 26, 30 (2d Cir. 2012); 17 U.S.C. § 512.

so having a presumption that these works are fair use will help eliminate some of the ambiguity. But, if the platform has something like YouTube's Content ID filter, it will not be a good faith belief, if the filter identifies that the entire uploaded work is a direct copy.

Nevertheless, the good faith belief would prevent the overinclusive takedown of works by absolving platforms of potential liability.

B. CREATING A BLANKET FAIR USE EXCEPTION FOR COVERS AND MASHUPS WILL BENEFIT CREATORS AND NOT MAKE ENFORCEMENT MORE CHALLENGING

From a policy perspective, implementing this proposal will help promote the arts and right the balance between incentivizing second-generation content creation, platforms, and rights holders. In this era of free-flowing ideas, works created for non-commercial purposes should continue to thrive to encourage content generation and access to creative works. Fair use is failing online creators, and not just those who make covers and mashups. For instance, people who record guitar tutorials and other educational videos often get their work taken down for DMCA copyright violations.¹¹⁸ It is difficult for creators to counter these allegations.¹¹⁹ While copyright holders are supposed to consider fair use before issuing takedown notices, they do not always do so in practice. Further, automatic takedown processes, like YouTube's Content ID filter, further complicate the ability to consider fair use before sending takedown notices. By giving these works a presumption of fair use, the works will be able to stay up, increasing access and content of works.

While copyright holders may disagree with how their work is being used in a cover or a mashup, artists have not been able to completely control who uses their music since 1909.¹²⁰ In order to use someone's song, you did not need to get their permission, you just had to file a notice of intent with the copyright office.¹²¹ Unlike other areas of law where people have the right to exclude others from their property, copyright has long recognized music as incompatible with traditional property rights. The compulsory licensing system severed the copyright holder's right to exclude others from using their musical composition by delinking control from compensation. In the pre-internet era, even though people could not control who used their music, it was much easier to prevent violations of copyright, because use and access of copyrighted

118. Use of copyrighted materials for educational purposes is typically considered fair use. Alexander, *supra* note 114.

119. *Id.*

120. MENELL, *supra* note 3, at 672.

121. *Id.*

works were more restricted.¹²² Now, the internet makes it much easier for people to access and use copyrighted music, so these same principles need to be extended in the digital age.

Applying and advertising a uniform policy can decrease confusion for creators. Creating a compulsory licensing scheme for mashups is challenging because, unlike covers, mashups sample multiple songs and may overlay tracks. The proposed policy, for finding fair use for non-commercial uses, allows mashup and cover artists to create without having to get permission and individual licenses for each song used.

Creating an exception for non-commercial covers and mashups will help advance creation and copyright enforcement. Mashups allow people to express creativity and discover their self-identity. People are going to continue to create work using copyrighted materials, regardless if they share their work online or not.¹²³ The abundance of takedown notices for works that should fall within the fair use exception has caused some people to become skeptical of the entire copyright system.¹²⁴ A presumption for fair use will allow these people to share their work, thus, promoting creativity and not alienating people from the copyright system.¹²⁵

Further, implementing this proposal will not make enforcement across platforms much more burdensome for copyright holders. Copyright holders have always held the burden of identifying infringing works that they own as they “are thus better able to efficiently identify infringing copies than service providers [. . .] who cannot readily ascertain what material is copyrighted and what is not.”¹²⁶ Additionally, record companies and copyright holders are already required to consider fair use before issuing a takedown notice.¹²⁷ Under a presumption for fair use, the rights holders just need to consider two things.

122. Menell, *supra* note 33, at 482.

123. YouTube is constantly taking down copyrighted work through takedown notices and issuing Content ID notices. For videos tagged by Content ID, YouTube lets the copyright holder decide how they want to proceed. They can either block the whole video, monetize the video by running ads on it, or track viewership. YouTube Help, *supra* note 94. On SoundCloud, creators are not allowed to upload any work that may contain copyrighted work. This restriction extends to “private works” that are not share publicly as the creator can still share the work with a limited audience. SoundCloud Help Center, *How to avoid copyright infringement*, SOUNDLOUD, <https://help.soundcloud.com/hc/en-us/articles/115003566668-How-to-avoid-copyright-infringement> (last visited Feb. 9, 2020).

124. See Peter K. Yu, *The Copyright Divide*, 24 CARDOZO L. REV. 331, 381 (2003) (discussing how the internet has led to a rise in “anti-copyright,” “anti-establishment culture” and people who subscribe to the mindset that all information should be free).

125. See *supra* Part IV.

126. Urban et al., *supra* note 41, at 17.

127. Alexander, *supra* note 114.

First, whether the work is commercial or non-commercial and, second, if it is a mashup or cover and not just an unmodified upload of the original work. The platform will need to indicate the commercial nature of the work to viewers and record companies. Currently, there is no formal process to evaluate if record companies are taking fair use into consideration when issuing takedown notices, so creating a presumption of fair use will shift the balance in favor of second-generation creators.¹²⁸ Even if a notice is issued, creators will have a legal basis to argue why their work should remain online. The balance shifts from “infringing” to “non-infringing,” unless proven otherwise.

Just as the MMA fixed the balance between rights holders and streaming services, implementing such a system will help correct the balance between second generation creators, copyrights holders, and platforms.

VI. CONCLUSION

The MMA helped address some of the music copyright issues created by the internet and new forms of listening. Specifically, the MWMA helped create a system to ensure accurate compensation, so copyright holders can benefit from their work. However, the MMA dealt largely with challenges faced by those in the music industry and did not consider music copyright issues that second-generation creators have when using copyrighted work.

The DMCA liability structure encourages platforms to takedown videos if there is any chance of infringement. Fair use, an affirmative defense to infringement, is a multi-step test and can be difficult to apply. As such, works that should be considered fair use are also taken down. Non-commercial covers and mashups should be considered fair use, until proven otherwise, in order to incentivize second-generation creators and increase society’s access to creative musical works.

Creating a statutory, rebuttable presumption for fair use will benefit second-generation creators, while protecting copyright holders’ ability to remove unauthorized uploads of their work.

128. *Id.* Any takedown requests sent to YouTube counts as a strike against the user, even if they are invalid. Urban et al., *supra* note 41, at 47.

RIMINI STREET V. ORACLE AND THE PROBLEM OF HIGH TRANSACTIONAL COSTS IN COPYRIGHT LITIGATION

Alistair McIntyre[†]

I. INTRODUCTION

By default, parties to litigation in the United States must pay their own litigation costs.¹ The Copyright Act is an exception to this rule, allowing a court to award “full costs” of litigation to the prevailing party.² But what are “full costs”? While “taxable costs” are clearly defined by statute,³ for years circuit courts disagreed as to whether “full costs” were broader than taxable costs.⁴ The Supreme Court recently resolved this issue in *Rimini St., Inc. v. Oracle USA* by concluding that “full costs” have the same scope as “taxable costs.”⁵ This decision represents a significant change to copyright law and is likely to alter the cost-benefit analysis of many potential copyright litigants.

This Note proceeds in four parts. Part II discusses the legal background of *Rimini*. It touches on the American rule of fee shifting, *Crawford Fitting Co. v. J.T. Gibbons, Inc.*’s effect on judicial interpretation of cost-shifting statutes, and fee shifting under the Copyright Act.⁶ Additionally, it explains the circuit split leading to *Rimini*. Part III overviews *Rimini* itself, covering the facts of the dispute, the procedural history, a summary of the arguments presented to the Supreme Court, and the opinion. It ends with an argument that *Rimini* represents the better of two choices. Part IV presents an empirical overview of cost-shifting in recent copyright litigation. It additionally discusses both the

DOI: <https://doi.org/10.15779/Z38D50FZ3Z>

© 2020 Alistair McIntyre.

[†] J.D., 2021, University of California, Berkeley, School of Law.

1. Jane P. Mallor, *Punitive Attorneys’ Fees for Abuses of the Judicial System*, 61 N.C. L. REV. 613, 613 (1983).

2. 17 U.S.C. § 505 (2018).

3. 28 U.S.C. § 1821 (1996); 28 U.S.C. § 1920 (2008). The statute refers to fees that are “taxable as costs,” which is synonymous with “taxable costs” for purposes of this Note.

4. *See Rimini St., Inc. v. Oracle USA, Inc.*, 139 S. Ct. 873, 876–77 (2019).

5. *Id.* at 881.

6. *See Crawford Fitting Co. v. J.T. Gibbons, Inc.*, 482 U.S. 437 (1987); *see also* *Pinkham v. Camex, Inc.*, 84 F.3d 292, 295 (8th Cir. 1996) (limiting awarded costs to those stated in § 1920); *InvesSys, Inc. v. McGraw-Hill Cos., Ltd.*, 369 F.3d 16, 22 (1st Cir. 2004) (allowing a broader definition of “full costs” than that defined by §§ 1920 and 1821). In some instances, *Crawford Fitting* has been superseded by statutes. *See, e.g.*, *Hamilton v. Sheridan Healthcare, Inc.*, 2015 U.S. Dist. LEXIS 191801, at *18 n.14 (S.D. Fla. Jan. 21, 2015).

impact of *Rimini* on the state of copyright law and details the litigation landscape on which *Rimini* is acting. Part V argues that, while *Rimini* was correctly decided, heavy transactional costs should be addressed legislatively so copyright law can better achieve its goals.

II. BACKGROUND

A. THE AMERICAN RULE OF COST-SHIFTING

1. *The History and Current Status*

The default rule in the United States is that cost-shifting in litigation is not permissible, though there are some statutory and contractual exceptions.⁷ This American rule contrasts with the English rule, where litigation expenses are awarded to the prevailing party as a matter of course.⁸ The English rule arose out of English common law, but was abandoned in the early days of the U.S. courts for unclear reasons.⁹

While U.S. common law typically follows the American rule, there are exceptions to the rule. The common benefit doctrine, for example, allows a lawyer who recovers a common fund for the benefit of persons other than his client to be awarded reasonable attorney's fees from that fund.¹⁰ This prevents third parties from free-riding and gaining benefits from a plaintiff attorney's efforts. Another exception, the bad faith doctrine, allows courts to award attorney's fees to a prevailing party when the opponent has acted in bad faith during litigation.¹¹ Additionally, there are statutory exceptions to the American rule. One such exception is the Copyright Act,¹² which will be discussed below.

2. *Advantages and Disadvantages of the American Rule*

The United States' continued use of the American rule reflects policy choices built around the central goal of encouraging "expedient and inexpensive determination of every action."¹³ Courts deliberately favor the rule in the absence of overriding statutes because of several policy advantages.¹⁴

7. Mallor, *supra* note 1, at 613.

8. Mary Jo Hudson, Comment, *Expert Witness Fees as Taxable Costs in Federal Courts – The Exceptions and the Rule*, 55 U. CIN. L. REV. 1207, 1212 (1987).

9. *Id.*; Mallor, *supra* note 1, at 615.

10. *Boeing Co. v. Van Gemert*, 444 U.S. 472, 478 (1980). Note that common-benefit doctrine and common-fund doctrine are synonymous for the purposes of this Note.

11. *Hall v. Cole*, 412 U.S. 1, 5 (1973).

12. 17 U.S.C. § 505 (2018).

13. Hudson, *supra* note 8, at 1211.

14. *Id.*; see Nicholas Vennekotter, *Full Cost in Translation: Awarding Expert Witness Fees in Copyright Litigation*, 87 FORDHAM L. REV. 1721, 1725 (2019).

Firstly, the rule protects against unwarranted litigation expenses.¹⁵ Because expenses lie with the party that generates them, there is a strong incentive to minimize the expenses incurred while pursuing a positive outcome in court. Otherwise, under the English rule, parties might be incentivized into an “arms race” of legal expenditure.¹⁶ Secondly, the rule preserves broader access to courts by removing the possibility of being burdened by the opponents’ costs.¹⁷ Parties with limited resources might otherwise hesitate to bring a claim to court, with the possibility of facing crushing expenses at the end of litigation.¹⁸ Lastly, the rule lets courts avoid difficult issues of reviewing the litigation expenses sought as part of an award.¹⁹ Deciding which expenses relate to the litigation and which do not could be a difficult problem in sufficiently complicated cases.

But the American rule is imperfect. When costs are substantial relative to the award, a prevailing plaintiff falls short of being made whole.²⁰ If costs are so high that a prospective plaintiff would not come close to being made whole, a copyright holder is actually *discouraged* from enforcing her rights in court.²¹ The policy of encouraging broad access to courts is turned on its head. Additionally, the American rule does not reduce parties’ incentive to increase their opponents’ costs, unlike the English rule.²² For instance, one party might make discovery requests that are costly for the other party to fulfill. Under the American rule, costs can still be high.

15. Hudson, *supra* note 8, at 1211.

16. Theodore Eisenberg & Geoffrey P. Miller, *The English Versus the American Rule on Attorney Fees: An Empirical Study of Public Company Contracts*, 98 CORNELL L. REV. 327, 336 (2013); *see also* Avery Wiener Katz, *Indemnity of Legal Fees*, in ENCYCLOPEDIA OF LAW & ECONOMICS 7300 Section 3 (Boudewijn Bouckaert & Gerrit De Geest eds., 1999) (ebook), <https://reference.findlaw.com/lawandeconomics/literature-reviews/7300-indemnity-of-legal-fees-art.html#> (jurisdictions that follow the English rule often enact measures to prevent “arms race” behavior, such as limiting shifted expenditures to those that are “reasonable” or specifying shiftable costs by statutory schedule).

17. Hudson, *supra* note 8, at 1211.

18. *Id.*

19. *Id.*

20. *See* Brief for Ralph Oman as *Amicus Curiae* Supporting Respondents at 10, Rimini St., Inc. v. Oracle USA, Inc., 139 S. Ct. 873 (No. 17-1625) (2019).

21. *See id.*

22. Eisenberg & Miller, *supra* note 16, at 336.

3. Crawford Fitting *Introduces Ambiguity Regarding Taxable Costs*

In *Crawford Fitting Co. v. J. T. Gibbons, Inc.*, the Supreme Court failed to deal with an ambiguity that later generated a circuit split.²³ Although this holding did not arise in the context of intellectual property protections, it was viewed as broad enough to reach other areas, including copyright claims.²⁴ The Court considered whether prevailing parties could generally be awarded expenses, such as expert witness fees.²⁵ The Court identified § 1920 as defining the types of costs that could be awarded against a losing party under federal law.²⁶ Nonetheless, the Court left some wiggle room for future decisions: “Any argument that a federal court is empowered to exceed the limitations explicitly set out in §§ 1920 and 1821 without plain evidence of congressional intent to supersede those sections ignores our longstanding practice of construing statutes *in pari materia*.”²⁷ In effect, the Court asked for lower courts to interpret legislative intent, in determining whether § 1920 applied in other situations involving expert witness costs.²⁸

B. THE COPYRIGHT ACT

1. *Cost-Shifting Under the Copyright Act*

The Copyright Act of 1976 exemplifies a statutory exception to the American rule, which was subject to debate under the *Crawford Fitting* framework.²⁹ 17 U.S.C. § 505 permits courts to award “full costs” to the prevailing party. Unlike other fields of intellectual property law that only allow award of costs to prevailing plaintiffs,³⁰ the Copyright Act allows award of costs to prevailing defendants too.³¹ This encourages defendants to pursue

23. See *Crawford Fitting Co. v. J.T. Gibbons, Inc.*, 482 U.S. 437 (1987); see also *Pinkham v. Camex, Inc.*, 84 F.3d 292, 295 (8th Cir. 1996) (finding that the Eighth Circuit limited awarded costs to those stated in § 1920) and *InvesSys, Inc. v. McGraw-Hill Cos., Ltd.*, 369 F.3d 16, 22 (1st Cir. 2004) (finding that the First Circuit allowed a broader definition of “full costs” than that defined by §§ 1920 and 1821). In some instances, *Crawford Fitting* has been superseded by statutes. See, e.g., *Hamilton v. Sheridan Healthcare, Inc.*, 2015 U.S. Dist. LEXIS 191801, at *18 n.14 (S.D. Fla. Jan. 21, 2015).

24. *Crawford Fitting*, 482 U.S. at 438–39; see *Twentieth Century Fox Film Corp. v. Entm’t Distrib.*, 429 F.3d 869, 885 (9th Cir. 2005).

25. *Crawford Fitting*, 482 U.S. at 438.

26. *Id.* at 440.

27. *Id.* at 445 (citing *United States v. United Continental Tuna Corp.*, 425 U.S. 164, 168–69 (1976) and *Train v. Colo. Pub. Int. Res. Group*, 426 U.S. 1, 24 (1976)).

28. See *id.*

29. See *Rimini St., Inc. v. Oracle USA, Inc.*, 139 S. Ct. 873, 877 (2019).

30. See, e.g., 15 U.S.C. § 1117(a) (2008); 35 U.S.C. § 284 (2011).

31. See *Fogerty v. Fantasy, Inc.*, 510 U.S. 517, 534 (1994).

“meritorious copyright defenses . . . to the same extent” a plaintiff would.³² This supports the underlying policy goal of demarcating the boundaries of copyright law clearly, so creative works are not tied up in nuisance litigation and instead available to the public.³³

While § 505 might seem clear on its face, circuits developed differing definitions of the scope of “full costs,” which came to a head in *Rimini*.³⁴ “Costs” is a term of art; under Rule 54(d)(1) of the Federal Rules of Civil Procedure, “costs” refer to expenses that can be taxed to the losing party under 28 U.S.C. § 1920.³⁵ That section sets forth six categories of these taxable costs:

- (1) Fees of the clerk and marshal;
- (2) Fees for printed or electronically recorded transcripts necessarily obtained for use in the case;
- (3) Fees and disbursements for printing and witnesses;
- (4) Fees for exemplification and the costs of making copies of any materials where the copies are necessarily obtained for use in the case;
- (5) Docket fees under [28 U.S.C. §] 1923 [];
- (6) Compensation of court appointed experts, compensation of interpreters, and salaries, fees, expenses, and costs of special interpretation services under [28 U.S.C. §] 1828 [].³⁶

28 U.S.C. § 1821 further defines compensation for witnesses. The attendance fee for a witness is limited to a mere forty dollars per day.³⁷ Travel and subsistence expenses are also reimbursable.³⁸

Cost-shifting in copyright law is different than other areas of intellectual property law. For this reason, patent and trademark laws are of little use in guiding interpretation of § 505. The Lanham Act specifies trademark law’s treatment of cost-shifting. Within the Lanham Act, 15 U.S.C. § 1117(a) addresses cost shifting, providing that a prevailing “plaintiff shall be entitled . . . to recover . . . the costs of the action.” These costs are limited to

32. *Id.* at 527.

33. *Id.*

34. *Rimini*, 139 S. Ct. at 876.

35. 10 MOORE’S FEDERAL PRACTICE § 54.103 (Matthew Bender 3d Ed. 2019).

36. 28 U.S.C. § 1920 (2008).

37. 28 U.S.C. § 1821 (1996).

38. *Id.*

those defined in 28 U.S.C. § 1920³⁹ and are mandatory.⁴⁰ § 1117(a) does not address expert fees, but the Second Circuit has indicated that expert witness fees in excess of forty dollars per day are outside the scope of costs awardable under the Lanham Act.⁴¹

Likewise, the Patent Act allows limited cost shifting.⁴² 35 U.S.C. § 284 provides that “the court shall award the claimant damages . . . together with interest and costs.” Again, the scope of these costs is limited by 28 U.S.C. § 1920.⁴³ Like in trademark law, costs are mandatory and only awarded to a prevailing plaintiff.⁴⁴ Because § 284 does not directly address expert fees, courts have held that such fees in excess of forty dollars per day may not be shifted.⁴⁵

Prior to *Rimini*, circuit courts agreed that, minimally, the Copyright Act’s “full costs” included all the costs listed in § 1920.⁴⁶ The question posed by *Crawford Fitting* was whether the scope of “full costs” extended beyond § 1920.

C. THE CIRCUIT SPLIT OVER “FULL COSTS”

After *Crawford Fitting*, circuit courts were left to decide whether 17 U.S.C. § 505’s “full costs” represented “congressional intent” to provide for awards outside of the scope of §§ 1920 and 1821. Of the seven circuits to consider whether the “full costs” of § 505 were limited to the six categories identified in § 1920, only two circuits answered in the affirmative.⁴⁷

39. *See, e.g.*, *Uphoff v. Elegant Bath, Ltd.*, 176 F.3d 399, 411 (7th Cir. 1999) (“Costs are defined in 28 U.S.C. § 1920 and do not include expert witness fees unless the expert is appointed by the court.”).

40. *See Kingdomware Techs., Inc. v. United States*, 136 S. Ct. 1969, 1977 (2016) (“the word ‘shall’ [] connotes a requirement”).

41. *Merck Eprova AG v. Gnosis S.p.A.*, 760 F.3d 247, 266 (2d Cir. 2014) (“under 28 U.S.C. § 1821 and the Lanham Act, the district court likely should not have awarded more than \$40 per day in expert witness fees”).

42. *See* 35 U.S.C. § 284 (2018).

43. *See, e.g.*, *Parks v. Booth*, 102 U.S. 96, 106 (1880). (holding that “the complainant was not entitled to an allowance for any expenses beyond the taxable costs”; § 1920 defines the scope of taxable costs).

44. *See Kingdomware*, 136 S. Ct. at 1977; *see also Parks*, 102 U.S. at 106.

45. *See, e.g.*, *Amsted Indus. v. Buckeye Steel Castings Co.*, 23 F.3d 374, 379 (Fed. Cir. 1994).

46. *See, e.g.*, *Pinkham v. Camex, Inc.*, 84 F.3d 292, 295 (8th Cir. 1996); *Twentieth Century Fox Film Corp. v. Entm’t Distrib.*, 429 F.3d 869, 884–85 (9th Cir. 2005).

47. *Compare Pinkham*, 84 F.3d at 295 (finding that the Eighth Circuit limited awarded costs to those in § 1920) *and* *Artisan Contrs. Ass’n of Am. v. Frontier Ins. Co.*, 275 F.3d 1038, 1040 (11th Cir. 2001) (following the Eighth Circuit’s lead on “full costs”), *with InvesSys, Inc. v. McGraw-Hill Cos., Ltd.*, 369 F.3d 16, 22 (1st Cir. 2004) (allowing a broader definition of “full costs” than that defined by §§ 1920 and 1821), *Adsani v. Miller*, 139 F.3d 67, 75 (2d Cir.

The Eighth Circuit was one of the two to limit “full costs” to the six categories in § 1920.⁴⁸ The parties presented no authority regarding the meaning of “full costs.”⁴⁹ With only the statutory language of § 505 before it, the court found there was no congressional intent to treat “full costs” under § 505 differently from costs under other federal statutes.⁵⁰ In *Artisan Contractors Association of America v. Frontier Insurance Co.*, the Eleventh Circuit came to the same conclusion for similar reasons.⁵¹ Neither court addressed surplusage arguments that were determinative in other circuits.⁵²

In contrast, the other circuits determined that “full costs” exceeded § 1920’s categories. For example, the Ninth Circuit, like many other courts that found “full costs” exceeded § 1920’s categories, focused on a close textual analysis of § 505 and a surplusage argument.⁵³ The court referenced *Crawford Fitting* in searching for clear evidence of congressional intent.⁵⁴ In this search, the court opined that limiting § 505’s “full costs” to those found in § 1920 “effectively read[] the word ‘full’ out of the statute.”⁵⁵ Given that “full” became surplusage if costs were limited by § 1920, the court found “clear evidence of congressional intent” that § 505 authorized a broader scope of costs.⁵⁶

III. RIMINI STREET INC. V. ORACLE

This circuit split set the stage for *Rimini*, and for good reason. In a given high-stakes copyright case, the amount of costs that could be awarded to the prevailing party in the Ninth Circuit could be millions of dollars more than the amount awardable in the Eighth Circuit, due solely to the difference in interpreting “full costs.”⁵⁷

A. FACTS OF THE DISPUTE

Oracle is a software development company that licenses its data and operations software to businesses and non-profits.⁵⁸ In addition to licensing,

1998), *Coles v. Wonder*, 283 F.3d 798, 803 (6th Cir. 2002), *Susan Wakeen Doll Co. v. Ashton-Drake Galleries*, 272 F.3d 441, 458 (7th Cir. 2001), and *Twentieth Century Fox*, 429 F.3d at 885.

48. *Pinkham*, 84 F.3d at 295.

49. *Id.*

50. *Id.*

51. *Artisan Contractors*, 275 F.3d at 1039.

52. *See id.* at 1039; *Pinkham*, 84 F.3d at 292.

53. *Twentieth Century Fox Film Corp. v. Entm’t Distrib.*, 429 F.3d 869, 885 (9th Cir. 2005).

54. *Id.*

55. *Id.*

56. *Id.*

57. *See Rimini St., Inc. v. Oracle USA, Inc.*, 139 S. Ct. 873, 876 (2019).

58. *Id.*

Oracle provides software maintenance services.⁵⁹ Rimini also provides software maintenance for Oracle software, competing with Oracle itself.⁶⁰ To better compete, Rimini additionally provided Oracle software updates to its customers.⁶¹ Without a proper license, providing software updates constitutes copyright infringement.⁶² Rimini obtained the software updates from Oracle's website without a license, also violating the website's terms of use.⁶³

B. PROCEDURAL HISTORY

Oracle alleged that Rimini copied Oracle's software without a license as part of its customer service offering, in violation of the Copyright Act.⁶⁴ Oracle brought a copyright suit against Rimini and its CEO in the U.S. District Court of Nevada, additionally alleging violations of state computer access laws.⁶⁵

The jury found that Rimini had infringed Oracle's copyrights and violated state computer access statutes.⁶⁶ The jury awarded Oracle \$35.6 million in damages for copyright infringement and \$14.4 million for violations of state computer access statutes.⁶⁷ The district court additionally ordered Rimini to pay a \$28.5 million reimbursement for attorneys' fees, \$4.95 million for costs under § 1920, and \$12.8 million for nontaxable litigation expenses outside the scope of § 1920, like expert witnesses, e-discovery, and jury consulting.⁶⁸ Rimini appealed, and while the Ninth Circuit reduced the taxable cost award to \$3.4 million, the \$12.8 million for nontaxable litigation expenses was untouched.⁶⁹ The court held that the \$12.8 million award for expenses was proper under Ninth Circuit precedent, which did not limit "full costs" to the categories identified in §§ 1920 and 1821.⁷⁰

Rimini filed a petition for *certiorari* which was granted by the Supreme Court.⁷¹ At issue was whether the "full costs" of § 505 included non-taxable expenses outside the scope of §§ 1920 and 1821.⁷²

59. *Id.*

60. *Id.*

61. *Rimini Street, Inc. v. Oracle USA, Inc.*, OYEZ, <https://www.oyez.org/cases/2018/17-1625> (last visited Oct. 17, 2019).

62. *Id.*

63. *Rimini St.*, 139 S. Ct. at 876.

64. *Id.*

65. *Id.*

66. *Id.*

67. *Id.*

68. *Id.*

69. *Id.*

70. *Id.*

71. *Id.*

72. *Id.*

C. ARGUMENTS PRESENTED TO THE SUPREME COURT

1. *Arguments for Broad Costs*

One argument presented to the Court was familiar from the circuit courts' proceedings: interpreting "full costs" as taxable costs would result in surplusage of the word "full."⁷³ "Maxim[] of quantity," a rule of linguistics, supports this argument.⁷⁴ The maxim states that only necessary information, and no more, should be included in a statement.⁷⁵

Other arguments focused on the fact that the "full cost" language of § 505 existed within the Copyright Act of 1831 before § 1920 was enacted.⁷⁶ The *Cramford Fitting* principle—not construing congressional authorization of broad litigation expenses without clear evidence of intent—should look forward from the enactment of §§ 1920 and 1821.⁷⁷ But §§ 1920 and 1821 were passed *after* "full costs" had been used in the context of federal copyright law.⁷⁸ Congress could well have employed the "full cost" language in § 1920 but instead specified "taxable costs," suggesting that the two are distinct.⁷⁹ Alternatively, because Congress is not always consistent in separating "costs" from "fees" or "expenses," it is conceivable that "full costs" include expert witnesses or discovery, which are more commonly referred to in statutes as "expenses."⁸⁰

Another argument pointed out that discovery and expert costs make copyright litigation disproportionately expensive and may discourage parties from pursuing meritorious litigation or defense.⁸¹ The cost of copyright litigation is estimated to be over three times that of other litigation.⁸² In 2011, the average cost of a copyright infringement case was estimated to be between \$384,000 and \$2 million.⁸³ And even for a relatively low-stakes case, costs can

73. See Brief of *Amici Curiae* Scholars of Linguistics in Support of Respondents at 8, *Rimini St., Inc. v. Oracle USA, Inc.*, 139 S. Ct. 873 (No. 17-1625) (2019).

74. *Id.*

75. *Id.* The maxim of quantity requires a speaker to "not make [her] contribution more informative than is required," and requires listeners to give real meaning to each of the speaker's words. *Id.* (quoting H. Paul Grice, *Logic and Conversation*, in PETER COLE AND JERRY L. MORGAN, SYNTAX AND SEMANTICS 3: SPEECH ACTS 44 (Academic Press 1975)).

76. *Id.* at 13.

77. Brief of BSA | The Software Alliance as *Amicus Curiae* in Support of Respondents at 15–16, *Rimini St., Inc. v. Oracle USA, Inc.*, 139 S. Ct. 873 (No. 17-1625) (2019).

78. *Id.*

79. See *id.* at 13.

80. *Id.* at 13, 15–16.

81. Brief for Ralph Oman as *Amicus Curiae* Supporting Respondents, *supra* note 20, at 2.

82. Shyamkrishna Balganes, *Copyright Infringement Markets*, 113 COLUM. L. REV. 2277, 2285 (2013).

83. *Id.* at 2280.

be high.⁸⁴ In 2017, for copyright cases with less than one million dollars at stake, median costs were \$200,000.⁸⁵ Such high costs may deter copyright holders from enforcing their rights or discourage defendants from contesting frivolous infringement claims if settlement is sufficiently inexpensive in comparison.⁸⁶

Yet another argument looked to evidence of congressional intent during passage of the Copyright Act.⁸⁷ Experts were regularly used in copyright litigation before the passage of the 1976 Copyright Act.⁸⁸ At the time the Copyright Act was passed, Congress seems to have understood that copyrightable works could include complex material, like software, necessitating expert testimony.⁸⁹ Accordingly, Ralph Oman, *amicus curiae*, argued that Congress, knowing expert testimony is necessary in copyright litigation, would have retained the “full cost” language without intent to cover broad costs.⁹⁰

Lastly, proponents of the broad definition of costs argued that determining the scope of non-taxable costs beyond § 1920 would not create administrability problems.⁹¹ Existing caselaw provides standards for awarding discretionary costs under § 505 and could be used to provide guidance for awarding non-taxable costs.⁹² The scope of attorneys’ fees and costs taxable under § 1920 are already disputed in litigation.⁹³

84. Brief for Ralph Oman as *Amicus Curiae* Supporting Respondents, *supra* note 20, at 11–12 (citing AM. INTELLECTUAL PROP. LAW ASS’N, 2017 REPORT OF THE ECONOMIC SURVEY 44 (June 2017)).

85. *Id.*

86. U.S. COPYRIGHT OFFICE, COPYRIGHT SMALL CLAIMS: A REPORT OF THE REGISTER OF COPYRIGHTS 24 (2013), <https://www.copyright.gov/docs/smallclaims/usco-smallcopyrightclaims.pdf>; Brief of BSA | The Software Alliance as *Amicus Curiae* in Support of Respondents, *supra* note 77, at 8.

87. *See* Brief for Ralph Oman as *Amicus Curiae* Supporting Respondents, *supra* note 20, at 9.

88. *Id.*

89. *See, e.g.*, H.R. Rep. No. 94-1476, at 54 (1976) (stating that “computer programs” may be copyrightable).

90. Brief for Ralph Oman as *Amicus Curiae* Supporting Respondents, *supra* note 20, at 9.

91. *See, e.g.*, Brief of BSA | The Software Alliance as *Amicus Curiae* in Support of Respondents, *supra* note 77, at 9.

92. *Id.* at 9–10 (citing *Fogerty v. Fantasy, Inc.*, 510 U.S. 517 (1994) and *Kirtsaeng v. John Wiley & Sons, Inc.*, 136 S. Ct. 1979 (2016)).

93. *Id.*

2. *Arguments for Narrow Costs*

One argument for a narrow interpretation of “full costs” disputed that “full” was surplusage when the language was initially passed.⁹⁴ This argument put “full cost” in its historical context.⁹⁵ At the time the term “full costs” was originally written into statute, there were state cost-shifting statutes that allowed two thirds, double, or treble costs.⁹⁶ Patrick Gillen, *amicus curiae*, argued that “full costs” merely specifies that the court should award a hundred percent of taxable costs, not some multiple of that amount.⁹⁷

Proponents likewise argued that policy considerations support a narrow reading of “full costs.” One such policy consideration is that allowing award of non-taxable costs reduces the incentive to register a copyrightable work.⁹⁸ Although registration of a work is not compulsory under the Copyright Act, registration is desirable because it provides notice to would-be infringers.⁹⁹ In circuits like the Ninth, which broadly defines “full costs,” copyright owners could recover large cost awards without having fulfilled the timely registration requirement.¹⁰⁰ Proponents of narrow costs argue that copyright holders should not be able to recover exorbitant amounts, including fees, without putting others on notice of their copyright.¹⁰¹ Additionally, the broad definition of “full costs” can lead to odd situations where an attorney’s fees are not recoverable, but the cost of expert, who was hired by that attorney, is recoverable.¹⁰²

Another policy consideration is encouraging consistency in litigation, but a lack of clear standards for non-taxable costs increases the variability and uncertainty in copyright litigation.¹⁰³ In pre-Rimini copyright caselaw, there are several different standards for awarding non-taxable costs, varying by jurisdiction.¹⁰⁴ These various standards include whether the costs in question

94. See Brief of *Amicus Curiae* Professor Patrick T. Gillen in Support of Petitioners at 18, *Rimini St., Inc. v. Oracle USA, Inc.*, 139 S. Ct. 873 (No. 17-1625) (2019).

95. *Id.*

96. *Id.*

97. *Id.*

98. Brief for the American Intellectual Property Law Association as *Amicus Curiae* in Support of Neither Party at 2, *Rimini St., Inc. v. Oracle USA, Inc.*, 139 S. Ct. 873 (No. 17-1625) (2019).

99. *Id.* at 5, 9.

100. *Id.* at 7.

101. See, e.g., *id.*

102. *Id.*

103. *Id.* at 3.

104. *Id.* at 10.

were “reasonable or reasonably incurred”¹⁰⁵ or “necessary to the case,”¹⁰⁶ or whether the losing party was morally blameworthy.¹⁰⁷ Court treatment of computer-assisted research costs exemplifies the uncertainty around non-taxable costs.¹⁰⁸ Such research costs are sometimes awarded as non-taxable costs, other times lumped into attorneys’ fees, and still other times denied entirely.¹⁰⁹

D. OPINION OF THE SUPREME COURT

The Court unanimously rejected Oracle’s arguments of close textual analysis, historical meaning, and surplusage as unpersuasive.¹¹⁰

Firstly, the Court rejected Oracle’s argument that “full costs” must refer to all costs of litigation, including those beyond the costs specified in §§ 1821 and 1920.¹¹¹ Oracle argued, if Congress had intended § 505 to refer only to “costs” in §§ 1821 and 1920, § 505 would have specified “costs” but not “full costs.”¹¹² The Court disagreed that “costs” could include expenses beyond federal statutes. It stated, “‘[C]osts’ refers to the costs generally available under the federal costs statute—§§ 1821 and 1920. ‘Full costs’ are all the costs generally available under that statute.”¹¹³

Secondly, the Court rejected the argument that “full costs” is a historical term of art going beyond the “costs” of §§ 1821 and 1920.¹¹⁴ In part, this argument was based on the fact that the “full costs” phrase was borrowed from

105. *Id.* at 10–11 (internal quotations omitted) (citing *Ronaldo Designer Jewelry, Inc. v. Prinzo*, 2017 U.S. Dist. LEXIS 133121 (S.D. Miss. Aug. 21, 2017) and *Compass Homes, Inc. v. Heritage Custom Homes, LLC*, 2015 U.S. Dist. LEXIS 101338, at *34–35 (S.D. Ohio Aug. 3, 2015)).

106. *Id.* at 11 (internal quotations omitted) (citing *Clarity Software, LLC v. Fin. Indep. Grp., LLC*, 2016 U.S. Dist. LEXIS 70602, at *24–25 (W.D. Pa. May 31, 2016)).

107. *Id.* (citing *Under a Foot Plant, Co. v. Exterior Design, Inc.*, 2017 U.S. Dist. LEXIS 141634, at *12 (D. Md. Sept. 1, 2017) (citing 4 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 14.09 (Matthew Bender, rev. ed. 2021) [hereinafter NIMMER ON COPYRIGHT]); *Ducks Unlimited, Inc. v. Boondux*, 2018 U.S. Dist. LEXIS 38676, at *22 (W.D. Tenn. Mar. 9, 2018) (citing NIMMER ON COPYRIGHT, *supra*)).

108. *Id.* at 12–14.

109. *Id.* at 13–14 (citing *Compass Homes*, 2015 U.S. Dist. LEXIS 101338, at *34–35 (awarding reasonably incurred computer research costs as non-taxable costs); *InvesSys, Inc. v. McGraw-Hill Cos., Ltd.*, 369 F.3d 16, 22 (1st Cir. 2004) (awarding computer assisted research costs as attorneys’ fees); *Tempest Publ’g, Inc. v. Hacienda Records & Recording Studio, Inc.*, 141 F. Supp. 3d 712, 725 (S.D. Tex. 2015) (denying research costs because such costs are not recoverable under § 1920)).

110. *Rimini St., Inc. v. Oracle USA, Inc.*, 139 S. Ct. 873, 878–880 (2019).

111. *Id.* at 878.

112. *See id.*

113. *Id.* at 879.

114. *Id.*

English copyright law in 1831, before §§ 1821 and 1920 were enacted.¹¹⁵ But, the Court explained that lower courts “should not undertake extensive historical excavation” in interpreting cost statutes, citing *Cranford Fitting*.¹¹⁶ The Court also pointed out that, though “full costs” was taken from English law, Oracle failed to show the meaning of “full costs” at the time of enactment in 1831.¹¹⁷ “Full costs” may well have been used to contrast with half, double, or treble damages found in state law at the time of enactment.¹¹⁸ Further, the Court characterized §§ 1821 and 1920 as “a comprehensive schedule of costs for proceedings in federal court.”¹¹⁹

Lastly, the Court rejected Oracle’s surplusage arguments.¹²⁰ Oracle asserted that the word “full” would be surplusage if costs were restricted to only those in §§ 1821 and 1920.¹²¹ The Court disagreed because “full” made the costs mandatory from 1831 to 1976 (i.e., a court could not award a mere portion of the costs but was obliged to award “full” costs), at which point the Copyright Act was amended to make award of costs discretionary.¹²² Additionally, the Court pointed out that, if “full costs” were to mean all costs generated during litigation, there would be a different surplusage problem: “award[ing] a reasonable attorney’s fee to the prevailing party as part of the costs” would also be redundant.¹²³ The Court also pointed out that some redundancy within statutory language is not uncommon.¹²⁴

Oddly, the Court sidestepped higher-level policy arguments. For instance, the Court did not explain why a narrow interpretation of costs might be favorable for incentivizing parties to be efficient with their litigation spending. The Court also failed to address the argument that the prospect of high litigation costs might discourage meritorious litigation.

E. RIMINI GOT IT RIGHT

Rimini represents the best of the two options, even though the Court omitted some policy considerations from the opinion. The decision to narrowly interpret “full costs” retains an incentive to keep discovery and expert costs low and avoids administrability issues. Further, the possibility of high

115. *Id.*

116. *Id.*

117. *Id.* at 879–80.

118. *Id.* at 880.

119. *Id.* at 879.

120. *Id.* at 880.

121. *Id.*

122. *Id.*

123. *Id.* at 881.

124. *Id.*

litigation costs in copyright litigation is, by itself, not a compelling reason to allow shifting of broad costs.

With the costs of discovery and experts remaining mostly untaxable under *Rimini*, there is still a strong incentive on both parties to keep costs low. Allowing expert costs or discovery costs to be taxed would create an incentive for both sides to throw increasingly more resources at the case, in an expense “arms race.” If additional expense can marginally increase a party’s chances of winning the case, and the losing party ultimately pays for that additional expense, there is less pressure to keep costs low, especially if winning seems likely. Arguments have been put forth that the high cost of discovery and experts is a reason to shift costs to the losing party,¹²⁵ but one can only imagine how costs would grow if there were no incentives to keep them low.

Secondly, the approach taken by the circuits overruled by *Rimini* did not adequately define the scope of awardable costs. As pointed out by Justice Sotomayor, when costs are not limited by § 1920, “there [is] nothing statutorily or otherwise that would limit a judge’s discretion of awarding costs.”¹²⁶ Drawing a line between costs that “were caused by the suit” and costs that were not would be an issue that each court would have to face.¹²⁷ Certainly, if the Court had decided that “full costs” extended beyond § 1920, there would need to be further guidance to aid a court’s discretionary awarding of costs.

Relatedly, sanctions issued by the court are a better means of punishing litigation misconduct than shifting non-taxable costs. Even without the option to shift non-taxable costs, courts may award sanctions, in part, based on bad faith or fault.¹²⁸ *Rimini* does not change this, so the incentive for both parties to behave during litigation is still present. Further, sanctions are better tied to the level of misconduct, rather than being tied to the common economic costs of copyright litigation as non-taxable costs are.

Lastly, *Rimini* keeps copyright law in step with trademark and patent law, with respect to the scope of awardable costs. One brief in *Rimini* argued that “full costs” should be broad because, otherwise, copyright litigation’s high

125. See, e.g., Brief for the National Music Publishers’ Association and Recording Industry Association of America, as *Amici Curiae* in Support of Respondents at 9, *Rimini St., Inc. v. Oracle USA, Inc.*, 139 S. Ct. 873 (No. 17-1625) (2019).

126. Transcript of Oral Argument at 52, *Rimini St., Inc. v. Oracle USA, Inc.*, 139 S. Ct. 873 (2019) (No. 17-1625).

127. *Id.*

128. Eric C. Surette, Annotation, *Sanctions Available Under Rule 37, Federal Rules of Civil Procedure, Other Than Exclusion of Expert Testimony, for Failure to Obey Discovery Order Not Related to Expert Witness*, 156 A.L.R. Fed. 601 (2019).

costs might discourage certain plaintiffs from bringing suit.¹²⁹ But patent and trademark litigations are also complex and can involve costly discovery and experts.¹³⁰ Further, patent litigation tends to be even costlier than that of copyright, when comparing cases with similar monetary stakes.¹³¹ Other intellectual property statutes align with the policy argument that litigation costs should be low by restricting awardable costs.¹³² This demonstrates that the mere existence of high litigation costs is not a reason to broaden the scope of shiftable costs.

IV. THE MODERN REALITY OF COPYRIGHT LAW

A. RIMINI REPRESENTS A CHANGE IN THE LAW FOR THE MAJORITY OF COPYRIGHT CASES

In the years prior to *Rimini*, most copyright infringement filings were made within jurisdictions that were awarding costs outside the scope of §§ 1920 and 1821.¹³³ It is perhaps unsurprising that the circuits containing New York (with its large publishing industry) and California (with its notable film industry) account for so much copyright litigation activity. Over thirty percent of copyright infringement filings between 2008 and 2017 were made within the Ninth Circuit.¹³⁴ Over fourteen percent of copyright cases were filed in the Second Circuit during that same period of time.¹³⁵ Figure 1 shows each circuit's share of all copyright filings in the United States in the decade leading up to *Rimini*.

Of course, both the Second and Ninth Circuits are among those that were awarding broad, non-taxable costs. As shown in Figure 2, a comfortable majority of copyright filings between 2008 and 2017 were made in jurisdictions

129. Brief for Ralph Oman as *Amicus Curiae* Supporting Respondents, *supra* note 20, at 3–4.

130. See Chris Neumeyer, *Managing Costs of Patent Litigation*, IPWATCHDOG (Feb. 5, 2013), <https://www.ipwatchdog.com/2013/02/05/managing-costs-of-patent-litigation/id=34808/>; Charles P Lickson, *Trademark Protection: Is Litigation Worth the Cost?*, IPWATCHDOG (May 23, 2013), <https://www.ipwatchdog.com/2013/05/23/trademark-protection-is-litigation-worth-the-cost/id=40711/>.

131. See AM. INTELLECTUAL PROP. LAW ASS'N, 2017 REPORT OF THE ECONOMIC SURVEY 41–44 (June 2017).

132. See 15 U.S.C. § 1117(a) (2018); 35 U.S.C. § 284 (2018).

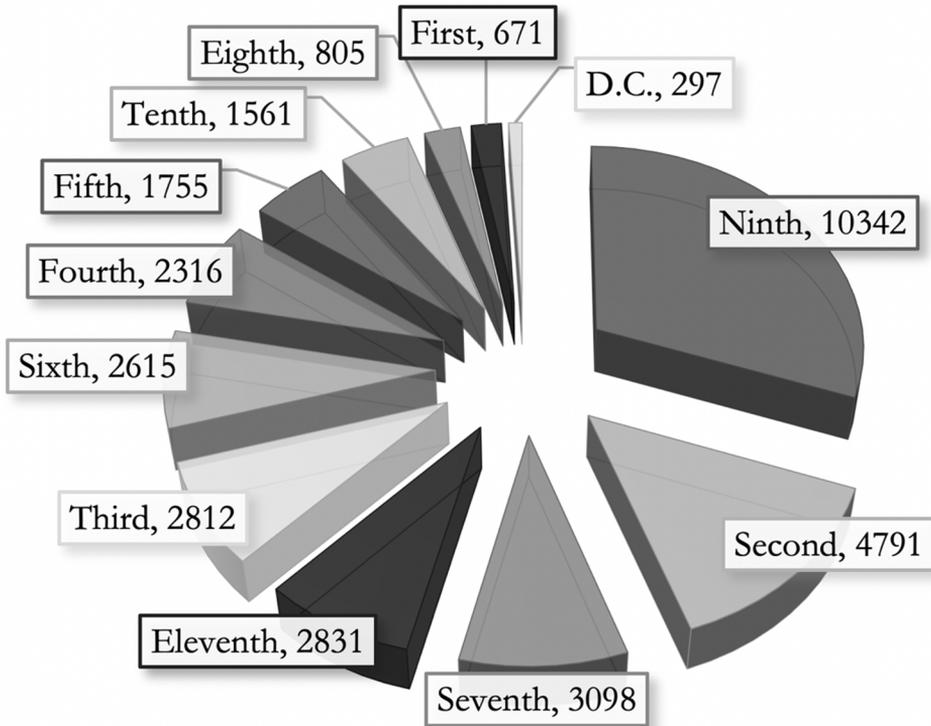
133. See Brief for Ralph Oman as *Amicus Curiae* Supporting Respondents, *supra* note 20, at 4a.

134. *Id.* at 10–11 (citing *Twentieth Century Fox Film Corp. v. Entm't Distrib.*, 429 F.3d 869, 884–85 (9th Cir. 2005), *cert. denied*, 548 U.S. 919 (2006)).

135. *Id.* at 11 (citing *Capitol Records, Inc. v. MP3tunes, LLC*, 2015 WL 7271565, at *6 (S.D.N.Y. Nov. 12, 2015)).

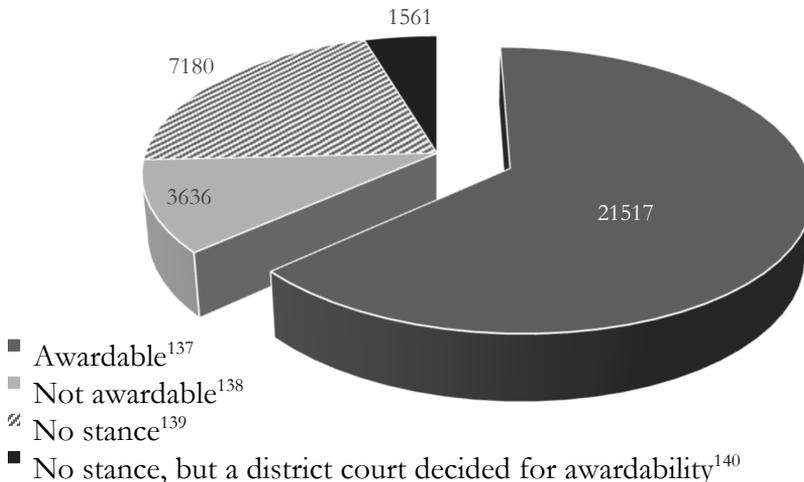
allowing award of broad costs: sixty-three percent. In effect, the *Rimini* ruling has changed the law of cost-shifting for the bulk of future copyright filings.

Figure 1: Copyright Infringement Filings by Circuit, 2008–2017¹³⁶



136. Brief for Ralph Oman as *Amicus Curiae* Supporting Respondents, *supra* note 20, at 4a.

Figure 2: Copyright Infringement Filings by Circuit Court’s Treatment of Non-Taxable Costs, 2008–2017



B. COSTS AND COMPLEXITY OF COPYRIGHT IN THE DIGITAL AGE

The fact that digital media is so easily copied could not have been anticipated during the early days of U.S. copyright law. Indeed, modern copyright infringement filings can regularly involve huge numbers of copyrights, infringers, and infringing acts. In one case, *UMG Recordings, Inc. v. Grande Communications Networks LLC*, the complaint alleged over a million acts of infringement.¹⁴¹ In another case, *BMG Rights Management (U.S.) v. Cox*

137. Brief for Ralph Oman as *Amicus Curiae* Supporting Respondents, *supra* note 20, at 41; *see also* *InvesSys, Inc. v. McGraw-Hill Cos., Ltd.*, 369 F.3d 16, 22 (1st Cir. 2003) (allowing a broader definition of “full costs” than that defined by §§ 1920 and 1821); *Adsani v. Miller*, 139 F.3d 67, 75 (2d Cir. 1998); *Coles v. Wonder*, 283 F.3d 798, 803 (6th Cir. 2002) (allowing a broader definition of “full costs” than that defined by §§ 1920 and 1821); *Susan Wakeen Doll Co. v. Ashton-Drake Galleries*, 272 F.3d 441, 458 (7th Cir. 2001) (allowing a broader definition of “full costs” than that defined by §§ 1920 and 1821); *Twentieth Century Fox*, 429 F.3d at 885 (allowing a broader definition of “full costs” than that defined by §§ 1920 and 1821).

138. Brief for Ralph Oman as *Amicus Curiae* Supporting Respondents, *supra* note 20, at 41; *see also* *Pinkham v. Camex, Inc.*, 84 F.3d 292, 295 (finding that the Eighth Circuit limited awarded costs to those in § 1920); *Artisan Contractors Ass’n of Am. v. Frontier Ins. Co.*, 275 F.3d 1038, 1040 (11th Cir. 2001) (following the Eighth Circuit’s lead on interpretation of “full costs”).

139. Brief for Ralph Oman as *Amicus Curiae* Supporting Respondents, *supra* note 20, at 41.

140. *Id.*; *see also* *Energy Intelligence Grp., Inc. v. CHS McPherson Refinery, Inc.*, 2019 U.S. Dist. LEXIS 14632, at *19 (D. Kan. Jan. 30, 2019) (awarding of non-taxable costs by a district court within the Tenth Circuit).

141. Plaintiffs’ Original Complaint at 2, *UMG Recordings, Inc. v. Grande Communs. Networks LLC*, No. 1:17-cv-365 (W.D. Tex. Apr. 21, 2017), ECF No. 1.

Communications, over a million acts of infringement were committed against over a thousand copyrighted works.¹⁴² Not only was there “extensive document and electronic discovery,” but *BMG* further involved thirty-six depositions of twenty-eight witnesses, eleven of which were experts, for a total of 270 hours of testimony.¹⁴³ This demonstrates that copyright litigation can require “the time-consuming task of piecing together incomplete, scattered records” to prove a defendant has infringed.¹⁴⁴ And all this effort comes at a cost.

1. *The Registration Requirement Is Problematic in the Digital Age*

While the costs and complexity of copyright litigation have changed since the early days of copyright law, authors’ motivation to register works has also changed. One argument against allowing award of non-taxable costs was that doing so would undermine incentives to register.¹⁴⁵ However, it is arguable that registration for creative works is not as feasible as it was in the past. This is a consequence of the ease of publishing content to the web.

Consider, for example, a small-time YouTube content producer who makes hundreds of videos and publishes them to her YouTube channel, generating income through ads shown before the videos. At the publishing stage, she will not necessarily know which of these videos will be widely viewed and valuable enough to warrant registration. Thus, registration for each and every video does not seem realistic.

If one of her videos is infringed, the Copyright Act’s statutory damages might seem particularly enticing because the actual damages resulting from copying of the video might be impossible to calculate. Choosing to pursue statutory damages could therefore promise faster and cheaper litigation. But a failure to timely register the video means that the full range of remedies for infringement, like statutory damages and attorneys’ fees, would be unavailable

142. *BMG Rights Mgmt. (U.S.) LLC v. Cox Communs., Inc.*, 149 F. Supp. 3d 634, 639–40 (E.D. Va. 2015), *aff’d in part, rev’d in part on other grounds*, 881 F.3d 293 (4th Cir. 2018). In *BMG*, the copyright holder brought a claim for infringement of more than 1,400 musical works. There were allegedly 2.5 million instances of internet users making the copyright holder’s works available for download. *Id.*

143. Brief for the National Music Publishers’ Association and Recording Industry Association of America, as *Amici Curiae* in Support of Respondents, *supra* note 125, at 12; Brief in Support of *BMG*’s Motion for Attorneys’ Fees and Costs at 27, *BMG Rights Mgmt. (U.S.) LLC v. Cox Communs., Inc.*, 149 F. Supp. 3d 634 (E.D. Va. 2015).

144. Brief for the National Music Publishers’ Association and Recording Industry Association of America, as *Amici Curiae* in Support of Respondents, *supra* note 125, at 12.

145. Brief for the American Intellectual Property Law Association as *Amicus Curiae* in Support of Neither Party, *supra* note 98, at 2.

in any subsequent infringement case.¹⁴⁶ This substantially diminishes the possible award and decreases the content producer's chance of being made whole after an infringing act. This in turn reduces her likelihood of bringing a suit in the first place. This hypothetical illustrates that, in the age of quick and easy publishing, a regime valuing registration does little to protect some creators of copyrightable work.

Copyright holders of high-value works are adequately enticed to register largely to access the possibility of shifting attorneys' fees. The ability to pursue statutory damages is not likely to provide a similar level of motivation, however. Courts are given discretion to award between \$750 and \$30,000 per work, though these upper and lower bounds can be altered depending on whether the infringement was willful.¹⁴⁷ For cases involving high-value works, statutory damages might be so low as to be inadequate, representing a poor alternative to calculating actual damages. While the prospect of shifting hefty attorneys' fees is a strong incentive, access to statutory damages is not a compelling reason for a copyright holder to register a high-value work. However, the discretionary range for statutory damages could be adjusted upward. Statutory damages might then be a viable incentive to timely register the high-value works.

C. TYPICAL AWARDS AND COSTS OF LITIGATION

1. *An Examination of Copyright Cost Statistics*

The complexity of copyright litigation seems to be evidenced by its comparative costs. Copyright litigation is estimated to be three times higher than the typical cost of litigation.¹⁴⁸ But copyright cases can vary greatly in terms of value, ranging from relatively low-stakes litigation over thousands of dollars to tens of millions of dollars.¹⁴⁹ Are costs in low-stakes cases more burdensome compared to those of high-stakes cases?

146. See 17 U.S.C. § 412 (2008). Even if the work is not registered, statutory damages are available if infringement occurred within three months of publication.

147. 17 U.S.C. § 504 (2010). A court may increase statutory damages to \$150,000 if the defendant is found to have infringed willfully. On the other hand, the court is allowed to lower statutory damages to \$200 if defendant "was not aware and had no reason to believe that his or her acts constituted an infringement of copyright." *Id.* § 504(c)(2).

148. Balganes, *supra* note 82, at 2285.

149. Compare *Tempest Publ'g, Inc. v. Hacienda Records & Recording Studio, Inc.*, 141 F. Supp. 3d 712, 716 (S.D. Tex. 2015) (having damages of merely \$5,000) with *Rimini St., Inc. v. Oracle USA, Inc.*, 139 S. Ct. 873, 876 (2019) (having damages of over \$35 million).

Typically, costs do not scale closely with the total amount of money at stake.¹⁵⁰ In other words, for a very high-stakes case, costs represent a smaller proportion of the total money at stake than for low-stakes cases. For copyright infringement suits in 2017 with less than one million dollars at stake, median cost of litigation through discovery, motions, and claim construction was \$100,000, or over ten percent of the value at risk.¹⁵¹ Through post-trial and appeal (when applicable), median costs jump up to \$200,000, over twenty percent of the total money at stake.¹⁵²

For the copyright cases with highest stakes—those with more than \$25 million at stake—the median cost of litigation through discovery, motions, and claim construction was merely \$600,000, less than 2.4% of the total value at stake.¹⁵³ Through pre- and post-trial and appeal, median costs rose to one million dollars, or less than four percent of total value at stake.¹⁵⁴ Compared to the copyright cases with lowest stakes, costs represent a much smaller proportion of the money at risk.¹⁵⁵

150. See AM. INTELLECTUAL PROP. LAW ASS'N, *supra* note 131, at 44 (presenting statistics from 2017).

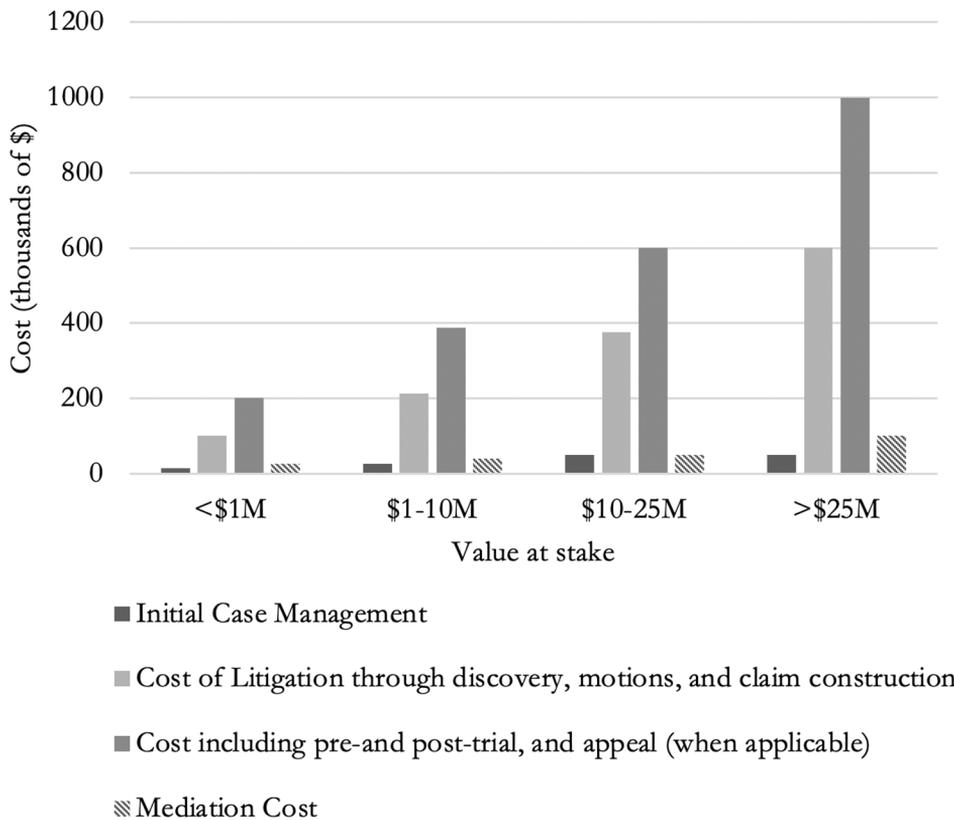
151. See *id.*

152. See *id.*

153. *Id.*

154. *Id.*

155. *Id.*

Figure 3: Median Litigation and Mediation Costs Plotted Against Monetary Stakes¹⁵⁶

One critical omission in these statistics is that some copyright holders were discouraged from bringing claims because of burdensome costs. The U.S. Copyright Office has acknowledged that copyright owners are “often” deterred from enforcing their rights in court due to the burden of high litigation costs.¹⁵⁷ Disproportionately large litigation costs seem to be a problem for lower-value claims, generally.¹⁵⁸ However, this is likely an even bigger problem for lower-stake copyright claims, given the tendency for higher costs in copyright litigation relative to other litigation.¹⁵⁹ As the Copyright Office acknowledged, “Especially in the case of lower-value copyright claims, the potential for monetary recovery can be quickly overcome by the costs of discovery, motion practice, and other litigation expenses.”¹⁶⁰ It seems unfair

156. *Id.* (presenting data from 2017).

157. U.S. COPYRIGHT OFFICE, *supra* note 86, at 24.

158. *See id.* at 25.

159. *See* Balganes, *supra* note 82, at 2285.

160. U.S. COPYRIGHT OFFICE, *supra* note 86, at 24.

that copyright holders are so disincentivized from bringing a claim for a low-value work when there is no other option to combat infringement.

2. *Example Cases*

While the median statistics show costs are typically sizeable but smaller than the award at stake, it is important to consider fringe situations as well. Copyright law should be able to sufficiently address both average and atypical cases. There are edge case examples where the harmed party remains less than whole after litigation. Some unusual copyright cases are included as illustration of phenomenon below.

Mattel, Inc. v. MGA Entertainment, Inc. is an example of an exceptional case because it is one of the costliest copyright cases to date.¹⁶¹ The court characterized Mattel's complaint as "stunning in scope and unreasonable in the relief it requested," adding that the claim "imperiled free expression [and] competition."¹⁶² MGA, the defendant, was awarded \$105 million in attorneys' fees and over \$31 million for costs that included non-taxable costs.¹⁶³ The cost award included reimbursement for expert opinion.¹⁶⁴ This shows the "high end" of what costs in copyright litigation might reach.

Costs may be incurred even by meritorious defendants in the face of objectively unreasonable complaints.¹⁶⁵ In *Kourtis v. Cameron*, the plaintiff pursued a claim for years that was substantially identical to a previously denied claim, with no new evidence.¹⁶⁶ During this time, the defense incurred expenses, which the district court was able to compensate.¹⁶⁷ The prevailing defendant was awarded nearly \$38,000 for the cost of an expert witness and \$166,000 in attorneys' fees.¹⁶⁸ Under *Rimini*, the defendant would have been \$38,000 poorer, even after winning this objectively unreasonable suit.

161. *Mattel, Inc. v. MGA Entm't, Inc.*, 2011 U.S. Dist. LEXIS 85998 (C.D. Cal. Aug. 4, 2011), *aff'd*, 705 F.3d 1108 (9th Cir. 2013). Mattel sued MGA over the entirety of their "Bratz" line of dolls, requesting more than one billion dollars in copyright damages. *Id.* at *15–16. While Mattel had copyrighted concept sketches and sculptures that were infringed by a few of the dolls, Mattel's request for relief was based on an "expectation that [Mattel] owned the 'ideas' in [these] copyrighted works." *Id.* at *16. In the words of the district court, "Mattel can't claim a monopoly over fashion dolls with a bratty look or attitude, or dolls sporting trendy clothing—these are all unprotectable ideas." *Id.* (quoting *Mattel, Inc. v. MGA Entm't, Inc.*, 616 F.3d 904, 917 (9th Cir. 2010)).

162. *Id.* at 42–43.

163. *Id.* at 43.

164. *Id.*

165. *See Kourtis v. Cameron*, 358 F. App'x 863, 867–68 (9th Cir. 2009).

166. *Id.*

167. *Id.* at 868.

168. *Id.*

In particularly large cases, nontaxable costs—which are now not awardable under *Rimini*—can range into the hundreds of thousands of dollars, even without running the full course of litigation.¹⁶⁹ For example, one court awarded defendants over five million dollars in attorneys’ fees and \$424,000 in non-taxable costs.¹⁷⁰ These expenses were incurred prior to summary judgment.¹⁷¹

Another costly case, *Pringle v. Adams*, illustrates that prevailing defendants may incur significant expenses defending a copyright infringement suit, even when the plaintiff has engaged in bad-faith behavior.¹⁷² In this case, the plaintiff disposed of a hard drive that he should have known would be relevant to litigation.¹⁷³ As a result, defendants’ motion for summary judgment was granted.¹⁷⁴ Even though this action was dismissed prior to trial, the defendants collectively incurred about \$2.17 million in attorneys’ fees and \$351,000 in non-taxable costs.¹⁷⁵

Similarly, a successful plaintiff may also incur substantial costs in litigating against a bad-faith defendant.¹⁷⁶ The court in *Capitol Records, Inc. v. MP3tunes, LLC*, found the defendant, not only willfully infringed, but took objectively unreasonable positions during litigation and induced fraudulent testimony from some of the witnesses.¹⁷⁷ The plaintiffs sought over \$705,000 in non-taxable costs, but the court granted less than half of those costs.¹⁷⁸ As a point of comparison, the plaintiffs were granted \$2.7 million in taxable costs.¹⁷⁹

In other extreme cases, a prevailing plaintiff may *lose* money on the case when costs beyond § 1920 are not awarded.¹⁸⁰ In *Tempest Publishing, Inc. v. Hacienda Records & Recording Studio, Inc.*, the plaintiff incurred \$15,400 in taxable costs.¹⁸¹ But, the court awarded only half these costs, no non-taxable costs, and

169. *See* Perfect 10, Inc. v. Giganews, Inc., 847 F.3d 657, 674 (9th Cir. 2017).

170. *Id.*

171. *Id.* at 665.

172. *See* Pringle v. Adams, 2014 U.S. Dist. LEXIS 101525, at *23–25 (C.D. Cal. July 23, 2014) (discussing plaintiff’s bad faith action in 2012 U.S. Dist. LEXIS 46332 (C.D. Cal. Mar. 30, 2012)).

173. *Pringle*, 2012 U.S. Dist. LEXIS 46332, at *21.

174. *Id.* at *29.

175. *Pringle*, 2014 U.S. Dist. LEXIS 101525, at *25–26.

176. *See* Capitol Records, Inc. v. MP3tunes, LLC, 2015 WL 7271565, at *6 (S.D.N.Y. Nov. 12, 2015).

177. *Id.* at *1.

178. *Id.* at *6. Label plaintiffs sought \$537,817.16, while publisher plaintiffs sought \$167,488.29, which together total to \$705,305.45. *Id.* The court granted merely \$105,288.33 to the label plaintiffs and \$142,763.29 to the publisher plaintiffs, totaling \$248,051.62. *Id.*

179. *Id.* .

180. *See* Tempest Publ’g, Inc. v. Hacienda Records & Recording Studio, Inc., 141 F. Supp. 3d 712 (S.D. Tex. 2015).

181. *Id.* at 726.

\$5,000 for copyright infringement.¹⁸² Even when a party successfully enforces its copyright, it might ultimately be in a worse position.

V. DISCUSSION: SOLVING THE PROBLEM OF HIGH COPYRIGHT LITIGATION COSTS

While the Supreme Court came to the correct conclusion regarding copyright law cost-shifting in *Rimini*, the fact that these costs can overshadow awards is a problem that should be further addressed. A potential solution should recognize the differences between low-stakes and high-stakes cases, because the economics of litigation are different for each. This Part's first Section discusses the wide variety of copyright infringers. After, the second Section argues that high litigation costs stymie the goals of copyright. The final Section ends with potential legislative solutions for both high-stakes and low-stakes copyright litigation.

A. COPYRIGHT INFRINGERS CAN RANGE FROM INDIVIDUALS TO CORPORATIONS

One of the quirks of copyright law is that the context in which the infringing act is committed can be so different. For example, a teenager may infringe by copying a song he enjoys off the internet. This is in stark contrast to a multi-billion-dollar corporation copying another multi-billion-dollar corporation's utilitarian software to maximize profits by avoiding licensing fees. But copyright litigation surprisingly treats each of these acts as being similar.

Other areas of law present cheaper, quicker alternatives to full-blown litigation. For general, civil litigation, there are small claims courts. Criminal law has traffic courts. Even patent law, with proceedings at the PTO, offers an option that can be less costly than litigation in a district court. Given the diversity of infringers, and damage caused by infringement, it seems odd that copyright law does not already have some alternative to litigation for low-stakes cases.

B. COSTS IN COPYRIGHT LITIGATION ARE SUFFICIENTLY HIGH THAT THE PURPOSE OF COPYRIGHT LAW IS NOT ALWAYS FULFILLED

When transactional costs, such as expert witness fees and discovery expenses, are too high, copyright litigation no longer accomplishes the policy goals of copyright law or litigation in general. These goals include making the

182. *Id.* at 717.

prevailing party whole, deterring infringement, and balancing access to works, with the author's right to exclude.

Litigation generally is intended to make a prevailing party whole, or at least as close to whole as is reasonable. This is consistent with the intent of § 505.¹⁸³ In a copyright suit, if non-taxable costs cut substantially into the award, the prevailing party will not be close to being made whole. Logically, if a prospective plaintiff knows that copyright litigation is an inadequate avenue for seeking a remedy, they are less likely to bring a meritorious action in the first place. Of course, a victim of infringement who does not bring a claim will remain far from "whole." Relatedly, absent the financial incentives to bring meritorious claims, infringing acts are less likely to be punished through litigation.

The prospect of losing litigation is intended to act as deterrence for infringement by punishing the losing party, the infringer. This punishment is tied to incentivization of good (i.e., non-infringing) behavior by parties that are in a position to infringe. It is important that the punishment is proportional to the harm caused, as to incentivize an appropriate level of infringement avoidance on the part of a potential infringer. Draconian punishments lead to over-protection of copyrightable work, whereas weak or rare punishment leads to under-protection.

More specifically, copyright law is meant to balance access to works of authorship and the authors' rights to exclude.¹⁸⁴ Copyright protection is meant to address the problem of "free-riding" on the backs of authors, which is addressed by creating legal excludability.¹⁸⁵ "Free-riding" includes enjoying a work without paying for it as well as profiting off an author's work by reselling it. The mechanisms of enforcing this excludability carry their own costs, such as the cost of litigation or the cost of copyright management.¹⁸⁶ On the other hand, access to a work, or lack thereof, is associated with "deadweight loss," the cost to society of certain people being unable to access material due to unwillingness or inability to pay the copyright price.¹⁸⁷ Heavy transactional costs, in enforcing excludability, disturb this balance. Such costs may make copyright holders less likely to enforce their rights. This, in turn, likely encourages more free-riding if potential infringers are aware that infringement

183. *Fogerty v. Fantasy, Inc.*, 510 U.S. 517, 525 (1993) (citing *McCulloch v. Albert E. Price, Inc.*, 823 F.2d 316, 323 (9th Cir. 1987)) ("because section 505 is intended in part to . . . make the plaintiff whole, fees are generally awarded to a prevailing plaintiff").

184. See Peter Eckersley, *Virtual Markets for Virtual Goods: The Mirror Image of Digital Copyright?*, 18 HARV. J. LAW & TECH. 85, 118 (2004).

185. *Id.* at 117–118.

186. *Id.*

187. *Id.*

will not be punished. If copyright holders increase the price of accessing their materials in order to compensate for increased enforcement costs, fewer consumers can pay the price and deadweight loss increases. Striking the right balance of access and exclusion through legislation allows for minimization of deadweight loss.

C. LEGISLATIVE ACTION COULD EASE THE SITUATION

1. *Low-Stakes Infringement: Change to Copyright Law Required*

With small harms caused (and often shallow pockets on the part of the infringer), it makes little sense to rely on the “high fine/low enforcement” cost-deterrence model of litigation to enforce copyright against low-stakes infringement.¹⁸⁸ As noted above, costs of litigation are likely to be disproportionately large in low-stakes copyright cases. Streamlining the process for non-commercial infringers through a compulsory licensing model or copyright small-claims process might help avoid high transactional costs, however.¹⁸⁹ At the time of writing this Note, Congress is considering a bill that would introduce a new small-claims proceeding at the U.S. Copyright Office, capping damages at \$30,000.¹⁹⁰ Alternatively, a system that acts like an issuance of a “ticket” for small-time infringers could help redirect non-commercial users from infringement to more favorable copyright usage, such as subscription plans.¹⁹¹ Importantly, such a system would allow the copyright holder to recover some money from a small-time infringer, as opposed to the current system where the economics (and sometimes unfavorable public relations) make it difficult for a copyright holder to retrieve any sort of repayment from the “little guy.” Further, because the small-time user might be more accountable for infringement, they would feel more pressure to adhere to proper copyright usage.

Either system should also have some diminished statutory damages accessible to works that were not timely registered. This would address the problem faced by the hypothetical YouTube content producer. Under such a system, she would be able to get something from an infringer, even if actual damages were impossible to calculate. The diminished statutory damages would, by necessity, be lower in amount than the normal statutory damage in order to continue incentivizing timely registration.

188. Peter S. Menell, *This American Copyright Life: Reflections on Re-Equilibrating Copyright for the Internet Age*, 61 J. COPYRIGHT SOC'Y OF THE U.S.A. 201, 269 (2014).

189. *Id.* at 218, 277.

190. *See* CASE Act of 2019, S. 1273, 116th Cong. (2019).

191. Menell, *supra* note 188, at 277.

One major consideration for a low-stakes claim system is that expanded subpoena power would be necessary to detect file sharers.¹⁹² This development would be required because, while identifying a person storing copyrighted work on a server can be done by an internet service provider without court intervention, identifying sharing conducted on a peer-to-peer network is not so straightforward.¹⁹³ This expanded power should include safeguards against abuse.¹⁹⁴ As learned from the success of the *inter partes* review option in patent law,¹⁹⁵ an alternative to litigation for copyright that is efficient and cheap is likely to succeed.

2. *High-Stakes Infringement: A Small Adjustment to the Law*

A streamlined system makes less sense for high-stakes suits, which are typically between big commercial parties. As mentioned above, high-stakes cases tend to have proportionally lower costs. The nature of infringement in high-stakes cases is more likely to be very complex, as it was in Rimini, warranting full-fledged discovery. Additionally, the decision to infringe was more likely tied to economic harm that lends itself to the calculation of actual damages. For example, it has been suggested that Rimini Street made the decision to infringe because they could save on the cost of licensing.¹⁹⁶ Litigation seems to be workable for disputes involving large commercial defendants, but additional options for plaintiffs to avoid burdensome costs would certainly be welcomed.

Increasing the range of statutory damages would offset high costs cutting into the prevailing party's award, but also keep incentives for parties to minimize costs. Statutory damages remove the need to gather extensive evidence of actual economic harm to the plaintiff, and therefore allow the plaintiff to focus discovery (and experts) on the infringing act itself. The current range for statutory damages, capped at \$30,000 per work, absent a showing of willful infringement,¹⁹⁷ might be too low for particularly valuable works. If the upper bound of statutory damages were better calibrated to approximate actual damages, plaintiffs could have a more enticing option

192. *Id.* at 273–74.

193. *Id.*

194. *Id.* at 274.

195. Josh Landau, *Inter Partes Review: Five Years, Over \$2 Billion Saved*, PATENT PROGRESS (Sep. 14, 2017), <https://www.patentprogress.org/2017/09/14/inter-partes-review-saves-over-2-billion/> (last visited Nov. 2, 2019).

196. *See* Brief in Opposition at 7, Rimini St., Inc. v. Oracle USA, Inc., 139 S. Ct. 873 (No. 17-1625) (2019).

197. 17 U.S.C. § 504 (2018).

during litigation that would allow adequate recovery while reducing discovery cost.

VI. CONCLUSION

The American rule, under which parties pay their own costs, and the English rule, under which the losing party pays all litigation costs, represent extremes of cost-shifting in litigation. Neither rule is perfect. The Copyright Act's treatment of cost-shifting falls somewhere in the middle of the two extremes, with its allowance of limited cost taxation under *Rimini*. This is almost certainly better than allowing broad cost-shifting in copyright cases, as there needs to be some pressure to keep litigation costs low. But when transactional costs of litigation are too high, litigation ceases to be an appropriate avenue for seeking relief. This seems to be especially problematic for low-stakes copyright litigation, where litigation costs tend to make up a larger proportion of money at stake. By contrast, the costs of high-stakes litigation represent a smaller proportion of the overall stakes. Congress should create an alternative to litigation to address this problem. Legislation directed at low-stakes infringement should create a copyright small-claims court or a compulsory licensing regime. The internet promotes fast, easy publishing but the copyright system values registration, which is not feasible for this mass of low-value work. Reduced statutory damages should be available for nonregistered works to give the creators of these works some protection. For high-stakes copyright litigation, reform is less necessary. A simple recalibration of statutory damages might allow for faster and cheaper litigation, even for high-stakes cases.

DO ANDROIDS DREAM OF COPYRIGHT?: EXAMINING AI COPYRIGHT OWNERSHIP

Gia Jung[†]

I. INTRODUCTION

If Shakespeare were an android, would we care? Would it make his work less worthy of praise, or less important to society? If Shakespeare, reincarnated as an android by a clever programmer, published a new play today, would we deny that play copyright? As outlandish as this hypothetical may seem, these are immediate questions about the state of copyright for works generated by artistic artificial intelligence (AI). The U.S. Patent and Trademark Office (USPTO) is aware, but uncertain, of the role and copyrightability of AI works. Already, programmers and companies have been registering for copyright on works produced by generative AI¹ despite the recent imposition of a “human authorship requirement” in the Compendium of Copyright Practices.² Because the Copyright Office, which administers the registration and recording of copyright, only needs someone to claim that a work is theirs, companies and programmers have been taking advantage of the lax investigation into claims of ownership.³ But *Naruto v. Slater*⁴ and the newest issue of the Compendium show that if those copyrights were ever challenged, they would likely be invalidated. Scholarly arguments as to the best formal regime⁵ and a request by the USPTO for comments⁶ characterize, but not clarify, this uncertainty. As it stands now, there is a disparity between what companies are doing in practice and what the Copyright Office allows on paper.

DOI: <https://doi.org/10.15779/Z383J3922D>

© 2020 Gia Jung.

[†] J.D., University of California, Berkeley, School of Law, Class of 2021.

1. Dani Deahl, *The USPTO Wants to Know if Artificial Intelligence Can Own the Content it Creates*, THE VERGE (Nov. 13, 2019, 1:45 PM), <https://www.theverge.com/2019/11/13/20961788/us-government-ai-copyright-patent-trademark-office-notice-artificial-intelligence>.

2. U.S. COPYRIGHT OFFICE, COMPENDIUM OF U.S. COPYRIGHT OFFICE PRACTICES § 313.2 (3d ed. 2014).

3. See, e.g., *infra* note 35.

4. *Naruto v. Slater*, No. 15-cv-04324-WHO, 2016 U.S. Dist. LEXIS 11041, 2016 WL 362231 (N.D. Cal. Jan. 28, 2016), *aff'd*, 888 F.3d 418 (9th Cir. 2018).

5. See, e.g., Kalin Hristov, *Artificial Intelligence and the Copyright Dilemma*, 57 IDEA 431 (2017).

6. See Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation, 84 Fed. Reg. 58,141 (Oct. 30, 2019).

This Note looks particularly at how artistic works produced by AI should be protected and incentivized. As AI-as-an-industry grows and becomes more advanced, so do its issues surrounding copyright. Rather than banning AI-produced works to keep the existing regime unchanged, it is time to reexamine why copyright hinges on human authorship, rather than creativity. Copyright presents three distinct hurdles for a work to be copyrightable.⁷ The first is originality, which mandates that a work must be independently created or not copied.⁸ The next is creativity, which is used not to judge the aesthetic merit of the work, but to filter against function.⁹ This standard ensures that copyright protects the substance of a work, not its general idea or theme. Finally, and most important to this Note, is authorship, which is used to establish the creator and rights-holder of a copyrighted work.¹⁰ This hurdle is highest for AI generated works because traditionally, originality and authorship were so tied together that authorship implied creativity, and vice versa.¹¹ AI-generated art, as authorless creative works, breaks the standard “if it looks creative, it must have a human author” way of understanding authorship and confronts how non-human authorship has been and can be accommodated.¹² This Note clarifies how and why AI-generated creative works meet these hurdles, and how incentive-based rationales for copyright protections apply.

This Note seeks to elucidate the policy arguments underlying the copyright system that support or oppose the adoption of AI authorship of artistic works. These will be used to explain institutional changes that can best serve the producers, users, and the future of generative AI. To that end, this Note begins with a brief description of generative artificial intelligence, arguing that the way AIs are designed makes output inherently creative, rather than operating as a tool of the programmer. Next, Part II explores some of the case precedent and history on machine and non-human copyright. From there, Parts III and IV proceed through the elements of originality and authorship, distinguishing copyrightability from ownership and exploring the implications of applying copyright to artificial intelligence.

7. U.S. COPYRIGHT OFFICE, *supra* note 2, § 302.

8. *Id.* § 308.1.

9. *Id.* § 308.2; *see also* Dennis S. Karjalam, *Copyright and Creativity*, 15 UCLA ENT. L. REV. 169, 201 (2008) (arguing that creativity is a necessary, not sufficient condition for copyright such that functionality funnels works either into patent or copyright).

10. U.S. COPYRIGHT OFFICE, *supra* note 2, §§ 302, 404, 405.

11. *See infra* Part II (showing how legal jurisprudence that affects this area became entangled).

12. *See Artificial Intelligence: The Ins and Outs of Copyright and AI*, U.S. PATENT & TRADEMARK OFFICE (Jan. 31, 2019), <https://rev-vbrick.uspto.gov/#/videos/d6e591c3-64cf-4d74-ab35-9f387a2da4b2>.

Part V proposes a model of copyright ownership for artistic works by artificial intelligence. A brief overview of current scholarship shows that popular proposals are incomplete in their considerations of liability, incentive, and equitable return on investment. This Note advocates for a model that draws on the works made for hire doctrine and current U.K. law addressing computer-generated copyright. AI works would have specific provisions for term limits and authorship to achieve an equitable balance between effort and reward. AI would be treated as the author-in-fact and the programmer as the author-in-law to clearly indicate attribution of risks and benefits. This Note concludes with a call for a clear model that weighs current issues and upcoming technological advances, so that the future of copyright may ensure appropriate accountability and administrability for creative works by artistic AI.

A. WHAT IS ARTIFICIAL INTELLIGENCE?

Confusion about what AI is extends beyond the general public into the field itself. Scholars' and practitioners' definitions vary from broad to narrow and analogize from aspects of human functions that are also hard to define—such as the ability to learn, consciousness, and self-awareness.¹³ What is most generally accepted and has made its way into the Oxford English Dictionary is that AIs are “computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.”¹⁴ In traditional forms of simple AI, programmers write many specific instructions for the program to follow and hope that their instructions are the best way to achieve their end goal.¹⁵ Current modes of generative AI are more complex and independent.¹⁶ This Note focuses on deep learning and generative adversarial neural networks as the two dominant modes of generative AI.

Deep learning and neural networks seek to create programs that behave as though they are interconnected brain cells. Programmers create multiple layers of processing “neurons” to allow the AI to learn how to recognize a solution on its own by looking for signature patterns of appearance, behaviors, or features at varying scales.¹⁷ Each level of neurons provide insight and passes

13. Matthew U. Scherer, *Regulating Artificial Intelligent Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. & TECH. 353, 360 (2016).

14. *Artificial Intelligence*, OXFORD ENGLISH DICTIONARY, https://en.oxforddictionaries.com/definition/artificial_intelligence (last visited Feb. 9, 2020).

15. Chris Rodley, *Deep Dinosaur*, CHRIS RODLEY (June 19, 2017), <https://chrisrodley.com/2017/06/19/dinosaur-flowers/>.

16. *Id.*

17. *Id.*

that information to the next, more senior level.¹⁸ After learning to recognize something based off of labeled data or supervised self-learning, deep learning programs can then apply that knowledge to recognize raw information and produce outcomes based on those understandings.¹⁹

Generative adversarial neural networks (GANs) set two neural networks against each other to improve the quality of the results. One generates imitations of the target material, and one evaluates whether or not the productions are forgeries. Learning in tandem, as the latter gets better at detecting forgeries, the former must become better at creating them.²⁰ In contrast to deep learning, GANs do not rely on a large amount of training data, as the interaction between the two networks itself creates a large amount of data from a limited source that continues to be cross-checked against each other.²¹

1. *Artistic and Creative Generative AI*

By their very design, deep learning and GANs (hereinafter generative AI) seek to imitate the process of human learning and creativity. Consequently, they produce creative works that are outside the control of the original programmer. Some scholars, including Professor Jane Ginsburg, dismiss the concept of AI authorship outright, claiming that AI is a mere tool, and its results are absolutely dependent on the inputs of the programmer.²² If this were the case, the extensive scholarship, debate, and calls for input by the USPTO would be moot; copyright already protects machine created works where there is creative input or intervention from a human author.²³ The issue at hand arises because the machines described here are generating works with an increasing level of independence from human intervention. And, as others have noted, the incorporation of randomness into a generative AI means it is only constrained to the same parameters as a human author—grammar,

18. See Simon Löfwander, *About Artificial Intelligence, Neural Networks, & Deep Learning*, AYIMA (Jan. 24, 2017), <https://www.ayima.com/blog/artificial-intelligence-neural-networks-deep-learning.html>.

19. *Id.*

20. Bernard Marr, *Artificial Intelligence Explained: What are Generative Adversarial Networks (GANs)?*, FORBES (June 12, 2019, 12:23 AM), <https://www.forbes.com/sites/bernardmarr/2019/06/12/artificial-intelligence-explained-what-are-generative-adversarial-networks-gans/#32aad65a7e00>.

21. *Id.*

22. Jane C. Ginsburg & Luke Ali Budiardjo, *Authors and Machines*, 34 BERKELEY TECH. L.J. 343, 396 (2019).

23. U.S. COPYRIGHT OFFICE, *supra* note 2, § 313.2.

cohesion, and genre-specific devices.²⁴ To that end, it could be said that in this context, the programmer has as much claim to an AI's work as Shakespeare does to any iteration of the star-crossed lovers narrative.²⁵

This Note specifically addresses artistic generative AI. These types of AI will consistently be referred to as “creative” or “artistic” as a way of reflecting the independent process of recognition, analysis, and generation that the AI undergoes to produce a unique work. The three below examples illustrate the generative AI referenced throughout this Note. The first, Chris Rodley’s “Deep Dinosaur” (Figure 1) produces novel and striking images by combining two sets of inputs—here, vintage flowers and dinosaurs.²⁶ Rodley and his Deep Dinosaurs are not the product of a large project like Google’s Deepmind, but an independent creative project undertaken by a PhD candidate at the University of Sydney.²⁷ Using a technique known as style transfer or deep style, the AI learns to recognize characteristics of images, then reproduce them, synthesizing the two characteristics to produce a recognizable amalgam.²⁸

24. See Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 2012 STAN. TECH. L. REV. 5, 12 (2012).

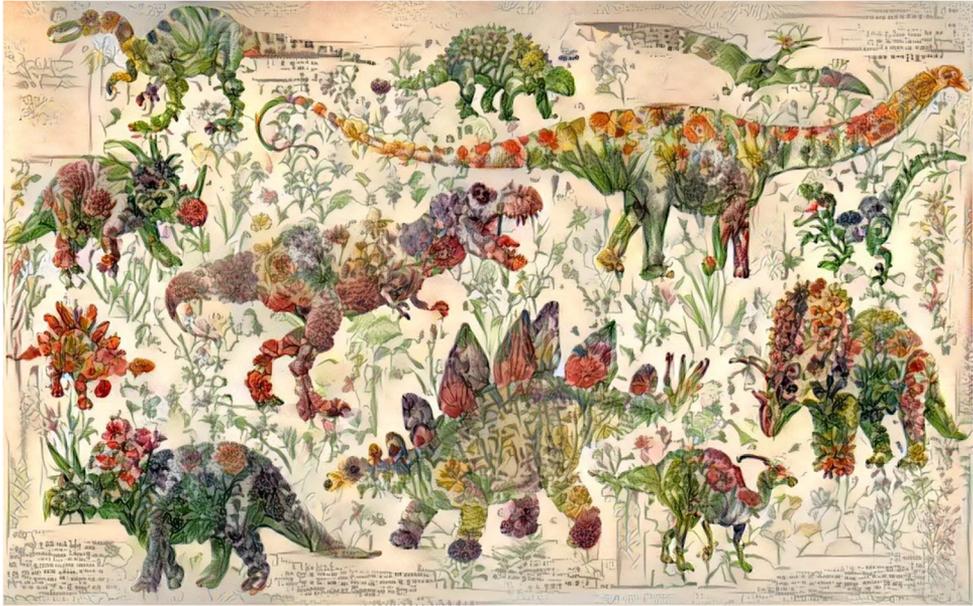
25. By this, I mean works like *High School Musical* (instead of feuding families, high-school cliques) and *West Side Story* (instead of feuding families, rival New York gangs) wherein the conceit is the same, but a court would likely not consider them derivative works of *Romeo and Juliet*. See also *Nichols v. Universal Pictures Corporation*, 45 F.2d 119 (2d Cir. 1930).

26. Rodley, *supra* note 15.

27. Chris Rodley, *Me*, CHRIS RODLEY, <https://chrisrodley.com/me/> (last visited Feb. 9, 2020).

28. Rodley, *supra* note 15.

Figure 1: Deep Dinosaur



The second is “Deep-speare,” a deep learning AI that writes sonnets (Figure 2).²⁹ Created by researchers for the 56th Annual Meeting of the Association for Computational Linguistics, Deep-speare was trained on the rules, structure, and elements of around 2,600 real sonnets.³⁰ From this data, Deep-speare was able to generate sonnet quatrains “nearly indistinguishable from published poems by humans” when read by the average person.³¹

29. Luke Dormehl, *Move Over, Shakespeare: This Sonnet-Writing A.I. is the Poet We Need*, DIGITAL TRENDS (July 30, 2018, 12:32 PM), <https://www.digitaltrends.com/cool-tech/ai-generates-shakespearean-sonnets/>.

30. Jey Han Lau, Trevor Cohn, Timothy Baldwin, Julian Brooke, and Adam Hammond, *Deep-speare: A Joint Neural Model of Poetic Language, Meter and Rhyme*, 56 PROC. ANN. MEETING ASS'N FOR COMPUTATIONAL LINGUISTICS 1948 (2018).

31. *Id.* (noting that a literature expert was not fooled—the stress, rhyme, and meter were all human-passable but the AI underperformed on readability and emotion).

Figure 2: Deep-speare Example Quatrain

“With joyous gambols gay and still array,
no longer when he ‘twas, while in his day
at first to pass in all delightful ways
around him, charming, and of all his days.”

The third is Endel, an app-based company that uses a generative algorithm to create “[p]ersonalized sound environments” for focus, relaxation, and sleep.³² In 2018, Endel partnered with Warner Music Group to produce and distribute these soundscapes on Apple Music and iTunes.³³ To monetize the songs, Warner needed to register them for copyright, and requested the songwriting credits for each. But besides creating the program itself, Endel as a company did nothing to generate the audio—the AI reacts to inputs like “rain” and creates the song entirely on its own.³⁴ Ultimately, the company “decided to list all six employees at Endel as the songwriters for all 600 tracks.”³⁵

The above examples do not begin cover the broad and fascinating world of generative AI. Rather, they serve as points of reference for thinking through specific aspects of AI copyright dilemmas. But before investigating how potential solutions for AI copyright ownership could apply to works like these, it is important to first understand how and why these issues arise out of the existing copyright regime.

II. LEGAL JURISPRUDENCE ON NON-HUMAN COPYRIGHT OWNERSHIP

Case precedent displays a grudging willingness to challenge traditional conceptions of copyrightability in order to accommodate mechanical and non-human copyright. The U.S. Constitution Article I, Section 8, authorizes

32. ENDEL, <https://endel.io/> (last visited Feb. 10, 2020).

33. *Endel to Release 20 Algorithm-Powered Albums to Help You Sleep, Focus & Relax*, ENDEL (Jan. 22, 2018), <https://endel.io/presskit/Endel-PressRelease-20MusicAlbums.pdf> [hereinafter *Endel.io*].

34. *See id.*; *Science*, ENDEL, <https://endel.io/science/> (last visited Feb. 10, 2020).

35. Dani Deahl, *We’ve Been Warned About AI And Music For Over 50 Years, But No One’s Prepared*, THE VERGE (Apr. 17, 2019, 10:30 AM), <https://www.theverge.com/2019/4/17/18299563/ai-algorithm-music-law-copyright-human>.

Congress “[t]o promote the [p]rogress of [s]cience and useful [a]rts, by securing for limited [t]imes to [a]uthors and [i]nventors the exclusive [r]ight to their respective [w]ritings and [d]iscoveries.”³⁶ This order, written intentionally broadly,³⁷ does not explicitly preclude non-human authors. However, the concept of “romantic authorship,” in which creativity can only stem from “an individual creative personality, a solitary originator of stylistically consistent works,” has long been the guiding concept for understanding how authorship and creativity interact.³⁸ As a result, disputes in the copyright field have developed over the scope of what constitutes “authors” and “writings” when it is not clear that the author was entirely human.

A. PHOTOGRAPHY AND *TORAH SOFT*—EXTENDING COPYRIGHT FROM MACHINE TO MAN

One of the first cases considering how machine-produced works could be copyrightable, *Burrow-Giles Lithographic Co. v. Sarony*,³⁹ established that works are copyrightable so long as they are “representatives of original intellectual conceptions of the author.”⁴⁰ In 1865, President Abraham Lincoln authorized a bill amending the Copyright Act to extend protection to photographs and photographic negatives, likely due in part to their “prominent role in bringing the horrors of the Civil War to the public.”⁴¹ Nearly twenty years later, this case tested the constitutionality of that 1865 legislation, a seemingly logical progression from maps, to sketchings and engravings, and now photography. In considering whether photographs, which were and are purely mechanical reproductions of their subjects, met the requisites of originality and creativity, the Court understood the concept of authorship could enable the product of a machine to be copyrightable.⁴² In concluding the case in favor of the photographer, Justice Miller commented that “photography is to be treated for the purposes of the act as an art,” such that the author uses the camera to

36. U.S. CONST. art. I, § 8, cl. 8.

37. *ArtI.S8.C8.1.1 Origins and Scope of the Power*, CONGRESS.GOV, https://constitution.congress.gov/browse/essay/I_8_8_1_1/ (last visited Feb. 11, 2020) (“Congress has broad leeway to determine how best to promote creativity and utility through temporary monopolies.”).

38. Bridy, *supra* note 24, at 4. See also ZACHARY LEADER, REVISION AND ROMANTIC AUTHORSHIP, 11 (1999) (“The Romantic author is often portrayed as spontaneous, extemporizing, otherworldly, and alone . . . Romantic attitudes to authorship profess a preference for what comes naturally, with a concomitant devaluing of secondary processes, including second thoughts.”).

39. 111 U.S. 53 (1884).

40. *Id.* at 58.

41. WILLIAM F. PATRY, COPYRIGHT LAW AND PRACTICE 244 (1996).

42. *Burrow-Giles Lithographic*, 111 U.S. at 58.

implement a creative idea.⁴³ To grant copyright, the Court extended the creative thought of the author through the camera, a logical inference from the control wielded by the photographer.⁴⁴

Notably, the Court in *Burrow-Giles* declined to establish whether unstaged photographs that had no active human author or human intervention in their composition could engender copyright.⁴⁵ However, the dicta implied that an unstaged photograph, since it did not include creative human participation, could not validly hold copyright.⁴⁶ Though not used explicitly in other cases, Professor Annemarie Bridy notes that the reasoning behind this dicta continued to perpetuate an unnecessary “dichotomy between creative and mechanical labor” in which automation is in opposition to creative authorship.⁴⁷ Because the case’s reasoning rested largely on Sarony’s involvement in setting up the photograph, later cases concluded that where mechanical creations arose, a human artist had to be substantially involved.⁴⁸

Following that dichotomy, the National Commission on New Technological Uses of Copyrighted Works (“Commission”)’s 1974 report expressed strong skepticism regarding precedent linking mechanical works to authorship.⁴⁹ Specifically directed to study the emergence of new, computer assisted works, the Commission declined to imagine the future implications of existing generative word processors.⁵⁰ Instead, they decided computers could only be “an inert instrument,” which, like the camera, functions solely as a creative tool that must be activated and directed by a human.⁵¹ The Commission thus advised Congress that no change should be made to accommodate automatic systems.⁵² This recommendation was challenged only ten years later by the Office of Technology Assessment (OTA).⁵³ The authors of the 1986 OTA report recognized that programs autonomously producing

43. *Id.* at 61.

44. *Id.* at 60 (deciding that the photograph is art on the basis that the photograph was given “visible form” by Sarony’s “desired expression”).

45. *Id.* at 59 (on the question of whether the ordinary production of a photograph without the photographer’s intervention, “we decide nothing”).

46. Bridy, *supra* note 24, at 5–6.

47. *Id.* at 6.

48. *See id.*

49. NAT’L COMM’N ON NEW TECH. USES OF COPYRIGHTED WORKS, FINAL REPORT OF THE NATIONAL COMMISSION ON NEW TECHNOLOGICAL USES OF COPYRIGHTED WORKS 4, 108–15 (1978).

50. *Id.*

51. *Id.* at 109.

52. *Id.*

53. U.S. OFFICE OF TECH. ASSESSMENT, INTELLECTUAL PROPERTY RIGHTS IN AN AGE OF ELECTRONICS AND INFORMATION (1986), 65–73.

text or musical works were a far cry from the word processing capabilities implied by the Commission.⁵⁴ In particular, the OTA report referenced the “blurring of the distinction between the copyrighted work and its product.”⁵⁵ The report concluded by suggesting that interactive computer programs, if not considered co-authors of the output produced, at least raised several troubling questions of copyright ownership and creativity requirements.⁵⁶

U.S. case law has not yet dealt with the copyrightability of an artistic work produced by AI.⁵⁷ However, a Southern District of New York case, *Torah Soft Ltd. v. Drosnin*,⁵⁸ comes within the realm of dealing with the output of an algorithm. Torah Soft, makers of a biblical-code-finding algorithm, sued over infringement of printouts of output generated by their computer program.⁵⁹ According to Bible code researchers, the Hebrew Bible is purported to be embedded with a code that is revealed by finding words and phrases which appear in the Bible at equidistant letter skips.⁶⁰ This software analyzed and found these code words and phrases.⁶¹ Interestingly, this court did not concern itself with authorship. The court suggested in passing that copyright protection afforded to the computer program may also extend to the output files. If the program does the “lion’s share of the work” in creating the output files and the end-user’s input is “marginal,” then it follows that the protection extends.⁶² Regardless, the court focused on whether the outputs contained protectable elements.⁶³ Because the software was made to comply with religious rules that were functional in nature, the court determined that the outputs were also functional and therefore not protectable.⁶⁴

54. *Id.*

55. *Id.*

56. *Id.*

57. This statement is accurate as of December 2019. *But see* BETA WRITER, LITHIUM-ION BATTERIES: A MACHINE-GENERATED SUMMARY OF CURRENT RESEARCH (2019) (a machine-generated book that claims copyright but is not registered in the U.S. Copyright Office’s Public Catalog); Leo Kelion, *AI system ‘should be recognised as inventor,’* BBC NEWS (Aug. 1, 2019), <https://www.bbc.com/news/technology-49191645> (“[T]wo professors from the University of Surrey have teamed up with the Missouri-based inventor of Dabus AI to file patents in the system’s name with the relevant authorities in the UK, Europe and US.”).

58. 136 F. Supp. 2d 276 (S.D.N.Y. 2001).

59. *Id.* at 280.

60. *Id.*

61. *Id.*

62. *Id.* at 283.

63. *Id.* at 283–84.

64. *Id.* at 287.

B. PSYCHOGRAPHY—GRANTING GOD COPYRIGHT

Another interesting line of cases that involve non-human works are those dealing with psychography, or automatic writing. Psychography is the production of writing or drawing supposedly by a spiritual agent, in which the human is merely a scribe.⁶⁵ Although parties seeking copyright protection in these cases claim that authorship actually lies with non-human, usually celestial or spiritual, beings, courts have found a sufficient nexus to human creativity to sustain copyright.⁶⁶

In each of these cases, courts apply *Feist Publications, Inc., v. Rural Telephone Service Co.*⁶⁷ to determine whether a compilation possesses the requisite originality for copyright protection. In *Feist*, the Court established that copyright demanded a minimal degree of creativity, which did not include an alphabetically arranged phonebook.⁶⁸ Nonetheless, compilations of facts may be copyrightable if arranged creatively, i.e., beyond a merely functional arrangement.⁶⁹ Referencing *Burrow-Giles*, the Court held that an author who claims infringement must prove the existence of “intellectual production, of thought, and conception.”⁷⁰ This showing can be demonstrated through active and intentional choices, where changes were not accidental or externally motivated.⁷¹ *Feist* thus sets a low bar for determining whether a work is creative or merely functional.

Along these lines, the Ninth Circuit in *Urantia Found. v. Maaberra*⁷² upheld a copyright granted on behalf of a “non-human” author.⁷³ There, the claimed work was dictated by a deity listed as “the Divine Counselor, the Chief of the Corps of Superuniverse Personalities, and the Chief of the Archangels of Nebadon.”⁷⁴ It was transcribed by the human whose name ended up on the copyright.⁷⁵ Pushing past the non-human element, the court found that the originality requirement necessary for a valid copyright was satisfied because the human beings who “compiled, selected, coordinated, and arranged” the

65. *Psychography*, OXFORD ENGLISH DICTIONARY, <https://en.oxforddictionaries.com/definition/psychography> (last visited Feb. 11, 2020).

66. Bridy, *supra* note 24, at 20.

67. 499 U.S. 340 (1991).

68. *Id.* at 362.

69. *Id.*

70. *Id.*

71. Bridy, *supra* note 24, at 8.

72. 114 F.3d 955 (9th Cir. 1997).

73. *Id.* at 964.

74. *Id.* at 957.

75. *Id.*

book did so “in such a way that the resulting work as a whole constitutes an original work of authorship.”⁷⁶

A similar case, *Penguin Books U.S.A., Inc. v. New Christian Church of Full Endeavor, Ltd*⁷⁷ held that, “[a]s a matter of law, dictation from a non-human source should not be a bar to copyright.”⁷⁸ There, a “Voice” the transcriber identified as Jesus dictated a manuscript and gave legal advice.⁷⁹ According to the author, Jesus advised that the copyright page should not have the author’s name in case people confused the copyright author with the actual author (Jesus and the Holy Spirit).⁸⁰ In a compromise, the copyright registration listed the work’s author as “[Anonymous](Helen Schucman).”⁸¹ The court held that, irrespective of independent editorial judgment in the process of recording the work, copyright could subsist in a non-human work so long as there is a sufficient human nexus. Helen’s acknowledgement of her involvement was sufficient.⁸²

C. *NARUTO V. SLATER*—THE MONKEY SELFIE THAT DROVE THE COPYRIGHT OFFICE BANANAS

As made clear in *Naruto v. Slater*, however, this liberal view of non-human authorship does not apply to animals.⁸³ Now widely known as the “Monkey Selfie Case,”⁸⁴ *Naruto v. Slater* dealt with whether Naruto, a crested macaque, had standing to state a claim under the Copyright Act over photographs of his own “independent, autonomous action.”⁸⁵ The dispute began when Slater—the wildlife photographer who set the cameras up and claimed ownership of the selfie—asked that the photos be taken off of Wikipedia. Wikipedia, in return, argued that the photo was uncopyrightable because it was taken by an

76. *Id.* at 958.

77. *Penguin Books U.S.A., Inc. v. New Christian Church of Full Endeavor, Ltd.*, 96 Civ. 4126 (RWS), 2000 U.S. Dist. LEXIS 10394 (S.D.N.Y. July 21, 2000).

78. *Id.* at *36.

79. *Id.* at *7–14 (“Schucman heard from the Voice that copyright registration should be sought for the Course.”).

80. *Id.* at *16.

81. *Id.* at *16–17.

82. *Id.* at *19, *32–33, *63–67.

83. *Naruto v. Slater*, No. 15-cv-04324-WHO, 2016 U.S. Dist. LEXIS 11041, 2016 WL 362231 (N.D. Cal. Jan. 28, 2016), *aff’d*, 888 F.3d 418 (9th Cir. 2018).

84. See, e.g., Joshua Berlinger, *PETA, Photographer Reach Settlement In ‘Monkey Selfie’ Case*, CNN (Sept. 12, 2017, 6:19 AM), <https://www.cnn.com/2017/09/12/asia/monkey-selfie-settlement/index.html>; Andres Guadamuz, *Can The Monkey Selfie Case Teach Us Anything About Copyright Law?*, WIPO MAG. (Feb. 2018), https://www.wipo.int/wipo_magazine/en/2018/01/article_0007.html.

85. *Naruto*, 2016 U.S. Dist. LEXIS 11041, at *3.

animal.⁸⁶ Seeing an opportunity to potentially strike a blow for animal rights, the People for the Ethical Treatment of Animals (PETA) filed a suit on behalf of Naruto as his “next friends.”⁸⁷ PETA alleged that Slater and the publishers of books containing the selfies violated Naruto’s copyright.⁸⁸

As this case was ongoing, the U.S. Copyright Office released an update to the Compendium Of U.S. Copyright Office Practices, which issued “The Human Authorship Requirement.”⁸⁹ Relying on citations from the *Trade-Mark Cases*⁹⁰ and *Burrow-Giles*,⁹¹ the Copyright Office concluded that it will only register an original work of authorship provided that the work was “created by a human being.”⁹² The requirement explicitly notes that a “photograph taken by a monkey” would not be considered an original work of authorship.⁹³ With this on the books, the court acknowledged and the Ninth Circuit affirmed that the Copyright Act does not confer standing upon animals like Naruto and dismissed the case.⁹⁴

Naruto’s holding that non-human entities cannot be considered “authors in law” leaves unanswered questions that have broad implications for ownership of non-human art. Slater’s active arrangement of the camera would seem to meet creativity standards under *Feist* and *Burrow-Giles*. However, unlike the psychography cases, the court did not address whether Slater would be able to enforce his own copyright on the photos.⁹⁵ Instead, Slater and PETA reached a settlement allowing Slater to continue to use the photos so long as he donates twenty-five percent of the revenue earned from them to a charity for crested macaques.⁹⁶ However, the ruling and the newest edition of the Compendium seem to suggest that the photos belong in the public domain. If that is the case, to what extent must a work be “created by a human being”?⁹⁷ How can those with financial interests in the output of non-human entities protect their investment? What sort of regime that protects non-human

86. *Id.* at *1–2.

87. *Id.*

88. *Id.* at *2.

89. U.S. COPYRIGHT OFFICE, *supra* note 2, § 306.

90. 100 U.S. 82 (1879).

91. 111 U.S. 53 (1884).

92. U.S. COPYRIGHT OFFICE, *supra* note 2, § 306.

93. *Id.* § 313.2.

94. *See* *Naruto v. Slater*, No. 15-cv-04324-WHO, 2016 U.S. Dist. LEXIS 11041, at *2, 2016 WL 362231, at *1 (N.D. Cal. Jan. 28, 2016), *aff’d*, 888 F.3d 418 (9th Cir. 2018).

95. *See generally* *Naruto v. Slater*, 888 F.3d 418 (9th Cir. 2018); *see supra* Section I.A.

96. Paulina Julia Perkal, *Monkey business finally settled: the ‘monkey selfie’ disputes*, KLUWER COPYRIGHT BLOG (Feb. 5, 2018), <http://copyrightblog.kluweriplaw.com/2018/02/05/monkey-business-finally-settled-monkey-selfie-disputes/>.

97. U.S. COPYRIGHT OFFICE, *supra* note 2, § 306.

creativity would best suit a tradition that inextricably intertwines the concepts of originality with authorship? This Note grapples with these questions to preclude a similar in-court outcome for AI generated works.

III. ORIGINALITY

The U.S. Compendium of Copyright Practices, citing *Feist*, states, “originality is ‘the bedrock principle of copyright’ and ‘the very premise of copyright law.’”⁹⁸ To qualify for copyright protection, a work must be original to the author, which means that the work must be “independently created by the author,” and it must possess “at least some minimal degree of creativity.”⁹⁹ This next Section shows that works by AI satisfy both elements by the nature of the process of output and by the underlying policies of copyright law.

A. INDEPENDENTLY CREATED

The threshold for independent creation is very low. Works are independently created so long as they are not literally copied.¹⁰⁰ A work of authorship may be original, even though it is not novel, inventive, or if “it closely resembles other works.”¹⁰¹ This low threshold allows a full gamut of protectable works while also setting parameters for infringement (direct copying).¹⁰² Generative AI’s purpose is to create novel outputs based on prompts, its learned process of recognition, and elements of randomness.¹⁰³ This process of generation usually means that each generated work can be considered independently derived. If the source material is not incorporated but is instead cut and pasted wholesale, the program is defective because it is not actually “generating” anything.¹⁰⁴

B. MODICUM OF CREATIVITY

Courts tend to distort the “modicum of creativity standard” as hinging on novelty or romantic abstraction. However, rather than artistic or inspired

98. U.S. COPYRIGHT OFFICE, *supra* note 2, § 308.

99. *Id.*

100. *Id.* § 308.1.

101. *Feist Publications, Inc. v. Rural Tel. Service Co.*, 499 U.S. 340, 345 (1991).

102. *Id.* (“The vast majority of works make the grade quite easily, [if] they possess some creative spark, ‘no matter how crude, humble or obvious it might be.’”).

103. *See* Rodley, *supra* note 15.

104. *See* Clark D. Asay, *Independent Creation in a World of AI*, FIU L. REV. (forthcoming 2020).

merit, the modicum of creativity requirement mainly functions as a bar to functional works.¹⁰⁵

Case law has traditionally considered creativity an individualistic and purely internal activity. Requisite creativity is characterized as the “fruits of intellectual labor” that “are founded in the creative powers of the mind”¹⁰⁶ and the “original intellectual conceptions of the author.”¹⁰⁷ This conception is rooted largely in the idea of the romantic Jeffersonian author who, through sheer intellect, creates original ideas out of nothing.¹⁰⁸ Works produced by a romantic author are by nature inherently original and creative, having sprung from a creative mind. Foucault scholars argue that this mythic figure is so central to our understanding about creativity that it obscures the collective and cumulative nature of creativity and misrepresents the “actual processes of cultural production.”¹⁰⁹ With this mythos on center stage, it becomes immeasurably difficult for courts and laypeople to detach the notion of creativity from popular representations. The image of the creative spark hitting the mind of the author like lightning such that words (or paint, sculpture, etc.) come pouring out like the final scene of Shakespeare in Love (1998) is ubiquitous and lauded.¹¹⁰ In actuality, Shakespeare should be considered the anti-romantic author, since few of his plots are original.¹¹¹ It is the reality of cumulative creativity, rather than the fantasy of the romantic author, that has driven art. But if creativity in copyright remains tied to the idea of romantic authorship, an AI work could not be considered original, since it is inherently algorithmic.

Another stream of cases presents the concept of originality as a stand-in for market value.¹¹² Setting aside artistic merit or romantic authorship, in 1839

105. See Karjalam, *supra* note 9, at 201; *Feist*, 499 U.S. at 346 (creating the standard that creativity is a low bar that works to filter out utilitarian works like an alphabetized phone book).

106. Trade-Mark Cases, 100 U.S. 82, 94 (1879).

107. Burrow-Giles Lithographic Co. v. Saronby, 111 U.S. 53, 58 (1884).

108. Oren Bracha, *The Ideology of Authorship Revisited: Authors, Markets, and Liberal Values in Early American Copyright*, 118 YALE L.J. 186, 188 (2008) (investigating the role of romantic authorship as a pervasive myth that is cyclically invoked or diminished depending on the interests at stake).

109. Bridy, *supra* note 24, at 4.

110. SHAKESPEARE IN LOVE (Universal Pictures 1998) (“A blank page. A hand is writing: TWELFTH NIGHT. We see Will sitting at his table . . . Will looks up from the table. ‘And her name will be . . . ‘Viola.’ He looks down at the paper, and writes: ‘Viola’”).

111. J. M. Pressley, *Shakespeare’s Source Material*, SHAKESPEARE RES. CTR., <http://www.bardweb.net/content/ac/sources.html> (last visited Dec. 13, 2019).

112. See Bracha, *supra* note 108, at 203 (In *Emerson v. Davies*, “Justice Story turned to the market as the sole arbiter of value”). See also Peter Jaszi, *Toward a Theory of Copyright: The*

and 1845 Justice Story decided cases based on a more practical understanding of the interdependent and cumulative nature of creativity.¹¹³ He relied on “the market as the only criterion for assessing value.”¹¹⁴ Justice Story saw that works imitate and build on each other. A work did not have to add anything new or culturally significant to be beneficial and protectable. If it was in demand, it was worthwhile.¹¹⁵

Justice Holmes was likewise unconcerned with romantic authorship. In *Bleistein v. Donaldson Lithographing*,¹¹⁶ Holmes affirmed that courts have no role in making aesthetic judgements. Using content neutrality and market value, the Court found that copyright had no threshold requirement of objective aesthetic value.¹¹⁷ Holmes focused not on the author or the merit of the advertisement, but on how the approval of the public eye and aesthetic elements placed them under statutory protection.¹¹⁸ *Bleistein* shows that works need not be particularly novel or skillful to be worthy of incentivization and protection.

As illustrated above, courts have competing views of originality. There remains a split on how and where creativity arises, and how important it is to copyright. Precedent like *Feist*, *Torab-Soft*, and Justice Story’s cases suggests AI-generated works meet the creativity bar. The demand for AI systems like Endel and Deep Dinosaur demonstrate enough market interest to satisfy Justice Story’s and Justice Holmes’ tests for creativity.¹¹⁹ Likewise, the choices AI systems make when generating content meet *Feist* and *Torab-Soft*’s standards for creativity in authorship.¹²⁰ Unlike an alphabetical phonebook or code-

Metamorphoses of Authorship, 1991 DUKE L.J. 455, 481–85 (1991) (examining cases in which “the [a]uthor vanishes” in the commercialization of cultural production).

113. Bracha, *supra* note 108, at 204. See, e.g., *Gray v. Russell*, 10 F. Cas. 1035 (C.C.D. Mass. 1839) (No. 5,728); *Emerson v. Davies*, 8 F. Cas. 615 (C.C.D. Mass. 1845) (No. 4,436).

114. Bracha, *supra* note 108, at 204.

115. *Id.*

116. 188 U.S. 239 (1903) (holding that commercial lithographs are copyrightable beyond their mere commercial value).

117. Bracha, *supra* note 108, at 200.

118. See Jaszi, *supra* note 112, at 482.

119. See, e.g., Andrew Liszewski, *A Neural Network Turned a Book of Flowers into Shockingly Lovely Dinosaur Art*, GIZMODO (June 19, 2017, 11:27 AM), <https://gizmodo.com/a-neural-network-turned-a-book-of-flowers-into-shocking-1796221045> (“The estate of M.C. Escher may have just lost its lucrative stranglehold on the dorm room poster market thanks to . . . a deep learning algorithm.”); *Endel.io*, *supra* note 33 (“[Endel] will be introduced to a larger audience through the extensive reach of the Arts Music division’s marketing and distribution resources . . . with investors including Amazon Alexa Fund, Avex Inc., Major Lazer’s Jillionaire, Plus 8 Equity Partners, Kima Ventures, Impulse Ventures, and world-famous DJ La Fleur.”).

120. See *supra* Section II.B; see also Karjalam, *supra* note 9, at 172.

finder where the end-result is necessarily the same, generative AI incorporates randomness into the choices given by its parameters so that the outcome is not purely a function of its form.¹²¹

At the heart of the matter is whether courts and legislatures will make policy changes to recognize AI-generated works as creative as they have in the past. In the psychography cases, courts found that a “non-human source should not be a bar to copyright.”¹²² In signing photographs into copyright law in the wake of the Civil War, Congress and President Lincoln made a policy judgement that photography itself, as a medium and a means of providing societal value, should be protected. If not simply for more floral dinosaurs, encouraging artistic AI will produce innumerable discoveries and solutions much as photography has.¹²³ And, as some courts acknowledge, “all creativity is inherently algorithmic and that works produced autonomously by computers are therefore less heterogeneous to both their human counterparts and existing copyright doctrine.”¹²⁴ The basic story of Romeo and Juliet has been done many times over, and yet, as *Nichols v. Universal Pictures Corp.* teaches us, building on generalized ideas and clichés are nonetheless protectable and encouraged.¹²⁵ As courts have similarly noted, most of Shakespeare’s plots came directly from existing stories—like Deep-speare, he learned and practiced on them to produce recognizably creative works.¹²⁶

121. See Rodley, *supra* note 15.

122. Penguin Books U.S.A., Inc. v. New Christian Church of Full Endeavor, Ltd., 96 Civ. 4126 (RWS), 2000 U.S. Dist. LEXIS 10394 (S.D.N.Y. July 21, 2000).

123. See *Science Photography*, SCITABLE BY NATURE EDUCATION (2014) <https://www.nature.com/scitable/spotlight/science-photography-10285104/> (“Careful observation of evidence is the heart of modern scientific method; photography has always been valued as an objective technique of observation, freed from the potential for human error implicit in the older method of sketching experimental observations. Just as important, photography can gather data that can’t be detected or processed by the human eye. Using technology that captured the scattered path of invisible x-ray beams, for example, Rosalind Franklin in 1952 was able to reveal the precise structure of intertwined DNA molecules—what we now recognize as the double helix.”); Demis Hassabis, *On AI’s Potential*, ECONOMIST (Nov. 21 2019), <https://worldin.economist.com/article/17385/edition2020demis-hassabis-predicts-ai-will-supercharge-science> (“By deepening our capacity to ask how and why, AI will advance the frontiers of knowledge and unlock whole new avenues of scientific discovery, improving the lives of billions of people.”).

124. Bridy, *supra* note 24, at 2.

125. 45 F.2d 119, 122 (2d Cir. 1930) (holding no infringement where the same stock characters and basic plot were used). See also *Boucicault v. Fox*, 3 F. Cas. 977, 982 (C.C.S.D.N.Y. 1862) (No. 1,691); *Emerson v. Davies*, 8 F. Cas. 615 (C.C.D. Mass. 1845) (No. 4,436).

126. *Boucicault*, 3 F. Cas. at 982 (“[T]he plays of Shak[e]spere are framed out of materials which existed long before his time, and were gathered by him” such that the skill and judgment in the selection and exposition constitute the basis of his reputation.).

Questioning the value or artistic worth of AI-produced work undercuts copyright law's goal of encouraging the creation of art. Like any other original expressive work, AI works introduce new ideas and aesthetics. They also provide society with an insight into the most pervasive patterns of human creativity.¹²⁷ Humans create AI as a way of furthering creative thought and demonstrating different or non-human ways of creative expression. From the “enjoyably strange” AI generated short film *Sunspring*¹²⁸ to the “shockingly lovely” Deep Dinosaur art,¹²⁹ there already exists widespread public recognition and demand for AI generated works.

IV. AUTHORSHIP

Like creativity, authorship's reliance on the role of the romantic author has created discomfort with recognizing machine produced works. The 2017 update to the Compendium on Copyright Practices repeatedly makes very clear that a human author must have created the work to be copyrighted.¹³⁰ However, prior to *Naruto* and the 2017 edition of the Compendium, authorship was always impliedly, but not explicitly, human.¹³¹ Specifically, that definition states, “the creator of the original expression in a work is its author. The author is also the owner of copyright unless there is a written agreement . . . In cases of works made for hire, the employer . . . is considered to be the author.”¹³² Copyright law does not distinguish between individuals, corporations, or metaphysical beings for the incentivization of aesthetic works. AI can and should fit into this regime. The next Section explains how lessons from the works made for hire doctrine can ground AI authorship as a viable concept.

A. LESSONS FROM THE WORKS FOR HIRE REGIME

The works made for hire (WMFH) regime is an exception to the rule that only the author can rightfully claim copyright. If a work is made for hire, an

127. See, e.g., BENJAMIN, SUNSPRING (an AI produced science fiction screenplay that repeats the line “I don't know”).

128. Michael Nordine, *Watch 'Sunspring,' a Short Sci-Fi Film Written by an Artificial Intelligence Algorithm*, INDIEWIRE (June 9, 2016, 3:10 PM), <https://www.indiewire.com/2016/06/watch-sunspring-sci-fi-artificial-intelligence-ai-written-1201687033/>.

129. Liszewski, *supra* note 119.

130. U.S. COPYRIGHT OFFICE, *supra* note 2, §§ 302, 306, 313.2, 803.5(C), 808.7(C) (wherever authors are mentioned, the human author requirement is reiterated: the cited sections are the human authorship requirement itself as well as examples of where it arises throughout the Compendium).

131. See *Definitions*, U.S. COPYRIGHT OFFICE, <http://www.copyright.gov/help/faq-definitions.html> (last visited Feb. 11, 2020).

132. *Id.*

employer is considered the author in which initial copyright vests even if an employee actually created the work.¹³³ Early courts decided WMFH cases on the basis of implied intent rather than who the original author was.¹³⁴ Rather than having copyright vest first in the author and then be subsequently assigned, the initial allocation of copyright between parties was “implied in the [employment] relationship between them.”¹³⁵ Those who procured the work (like book commissioners) or orchestrated its creation (like a theater owner) were granted copyright on the basis of the nature of the relationship.¹³⁶ By the nineteenth century, courts evolved their rationale. The key consideration was “the degree of involvement of the employer’s representatives in the creative process [and] the supervision exercised or in the expenditure undertaken by it.”¹³⁷ In effect, this assumed an employer had ownership over the employee’s work unless the employee could point to an explicit, contractual assignation of ownership.¹³⁸ Rights no longer vested in the artist by the nature of their romantic authorship. Instead, the work became an entity separated from the employee’s labor under the understanding that their artistry was essentially a tool to be used by the whims of the employer.¹³⁹

WMFH presents several important consequences for an AI copyright regime. The policy rationale behind WMFH is to incentivize employers and grant them control over works made on their behalf.¹⁴⁰ Like a Hollywood studio, an AI programmer invests time and resources to develop protectable systems. This investment should be rewarded in kind. Programs like Deep Dinosaur and programmers like Chris Rodley serve the general copyright goals of innovation and culture and should be protected accordingly.

Further, corporations are non-human entities in which copyright vests. Consequently, WMFH strengthens the psychography cases’ conclusion that

133. U.S. COPYRIGHT OFFICE, CIRCULAR 9: WORKS MADE FOR HIRE 1 (2012).

134. Bracha, *supra* note 108, at 253.

135. *Id.*

136. *Id.* at 252–54 (examining Keene v. Wheatley, 14 F. Cas. 180 (C.C.E.D. Pa. 1861) (No. 7644) and Lawrence v. Dana, 15 F. Cas. 26 (Clifford, Circuit Justice, C.C.D. Mass. 1869) (No. 8136)).

137. *Id.* at 254.

138. *Id.* at 255.

139. See Robert C. Denicola, *Ex Machina: Copyright Protection for Computer-Generated Works*, 69 RUTGERS U.L. REV. 251, 276 (2016) (“William Patry, in his treatise on copyright law, views the employer as author rule as an example of an ‘instrumental approach’ to copyright authorship.”)

140. Shlomit Yanisky-Ravid, *Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era—The Human-Like Authors are Already Here—A New Model*, 2017 MICH. ST. L. REV. 659, 711 (2017).

copyright can subsist in a non-human work.¹⁴¹ By divorcing the artist from their presupposed copyright, WMFH creates an important distinction between the author-in-law, with whom the rights vest, and the author-in-fact, who undertook the creative steps to produce the work.¹⁴² This distinction is an important and major exception to copyright resting with the actual artist. In so doing, it is also as a means of recognizing the creator of a work without allocating copyright to them. As a legal fiction, it allows companies to maximize the profitable labor of their employees by appealing to the need to protect investment and created “value” as the basis of copyright.¹⁴³ This adaptation mirrors the evolution of originality as not being intrinsically meritorious by reason of romantic authorship, but as a representation of subjective market value over which objective courts have limited, if not arbitrary, purview.

V. APPROACHING THE FUTURE: MODELS AND RECOMMENDATIONS FOR THE TREATMENT OF ARTISTIC AI COPYRIGHT

The current state of U.S. copyright law for AI generated works is uncertain and needs to be definitively clarified to establish the rights and limits of a rapidly expanding industry. As it stands, “copyright law is not currently structured to accommodate the particular authorship matrix of people-who-write-programs-that-make-art.”¹⁴⁴ Although the Compendium and its human-authorship requirement does not have the force of law, it nonetheless reveals the attitudes and practices of the Copyright Office and is taken to “provid[e] expert guidance” to courts.¹⁴⁵ Courts are unlikely to follow *Torab Soft’s* reasoning and extend copyright to the software owner where the software does the “lion’s share” of the work. Rather, relying on *Naruto* and the Compendium,

141. The Ninth Circuit in *Naruto v. Slater* noted that, although courts have allowed corporations to sue, corporations “are formed and owned by humans; they are not formed or owned by animals.” Nevertheless, this distinction made by the court only serves to illuminate how corporations under the WMFH regime are interpreted as non-human human entities nonetheless entitled to legal rights. 888 F.3d 418, 426 n.9 (9th Cir. 2018). See Bridy, *supra* note 24, at 26 (“The work made for hire doctrine is a more fitting framework within which to situate the problem of AI authorship because it represents an existing mechanism for directly vesting ownership of a copyright in a legal person who is not the author-in-fact of the work in question.”).

142. See Bridy, *supra* note 24, at 26 (“the work made for hire doctrine acknowledges a disidentity between the author-in-fact (the employee or contractor) and the author-in-law (the employer or other person for whom the work was made)”).

143. Bracha, *supra* note 108, at 260–61.

144. Bridy, *supra* note 24, at 22.

145. U.S. COPYRIGHT OFFICE, *supra* note 2, at 1.

a court would likely deny copyright protection for an AI-generated work. While U.S. programmers are without an avenue to rights through litigation, countries such as Hong Kong (SAR),¹⁴⁶ India,¹⁴⁷ Ireland,¹⁴⁸ New Zealand,¹⁴⁹ and the United Kingdom¹⁵⁰ allow AI-produced works to be protected.¹⁵¹ This discrepancy makes clear that the United States needs to address the status of artistic AI works in the near future.

A. A BRIEF OVERVIEW OF PROPOSED MODELS OF AI COPYRIGHT OWNERSHIP BY U.S. LEGAL SCHOLARS

Among U.S. legal scholars, the discourse on copyrightable AI works is governed by four main models. Each differently reflects concerns of how to place AI within authorship. Each tries to speak to whether the work is: (1) copyrightable in the first place (if non-humans can be creative or “original”); (2) whether and how authorship can vest in a non-human; and (3) the administrability of the proposed changes. To evaluate the models, this Section asks three questions that bear on the efficacy and rationales: Who is this meant to incentivize? Is there a need for this model? How would this affect the market for these works?

1. *Works Remain in the Public Domain*

Under the current copyright regime, AI works are in the public domain. Even though it is the official status quo, enforcing it would likely mean heightened scrutiny on behalf of the Copyright Office to ensure that programmers are not copyrighting AI generated works, as they currently are. Proponents of this model claim that incentive is maintained in two ways.¹⁵² First, because programmers can register copyright on the AI software itself, there is no logical reason for furthering the stream of ownership down to the

146. Hong Kong Copyright Ordinance, (1997) Cap. 528, 40, 44, §§ 11(3), 17(6).

147. Copyright (Amendment) Act, 1994, No. 38, Acts of Parliament, 1994 § 2(vi) (India).

148. Copyright And Related Rights Act 2000 (Act No. 28/2000) (Ir.) § 21(f), *available at* <http://www.irishstatutebook.ie/eli/2000/act/28/section/21/enacted/en/html#sec21>, *archived at* <https://perma.cc/XWA4-D7QP>.

149. Copyright Act 1994, s 5(2)(a) (N.Z.).

150. Each country’s statutory language either replicates or closely matches the United Kingdom’s, where “[i]n the case of a literary, dramatic, musical, or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken,” and only differ in the term granted—either seventy or fifty years from the date of creation. Copyright, Designs and Patents Act 1988, c. 1, § 9(3), 12(2) (Eng.).

151. Andres Guadamuz, *Artificial intelligence and copyright*, WIPO MAG. (Oct. 2017), https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html.

152. Ana Ramalho, *Will Robots Rule the (Artistic) World? A Proposed Model for the Legal Status of Creations by Artificial Intelligence Systems*, 21 J. INTERNET L. 11, 21–22.

AI outputs.¹⁵³ However, much like a manufacturer makes money by the product, not the production line, a programmer creating an artistic AI would be hard-pressed to monetize the program itself. The public domain proponent's response is that the creators of the AI can also be incentivized through the right to dissemination.¹⁵⁴ By treating owners of the AI as a publisher, AI generated works would exist behind a paywall and allow the existing structure of AI copyright to remain unchanged.¹⁵⁵ Reviews of the efficacy of paywalls are mixed, but the general consensus is they are only sustainable if the provider has a high level of reputation and uniqueness of content.¹⁵⁶ Proponents further argue that this gap may be filled by sites like Patreon, an "online tip jar" where artists crowdfund directly from fans either through one-time donations or subscriptions.¹⁵⁷

Theoretically, this model would work best for hobby artists like Chris Rodley, whose Deep Dinosaur and iterations thereof would likely bring in revenue for the high-resolution printable versions.¹⁵⁸ However, if those works are uploaded and available for download on another site, Rodley and others like him would be unable to enforce an action against the pirating site, since the works are technically in the public domain. This could incentivize a generation of AI works that are impossible or difficult to copy, such as experiential or constantly changing outputs. More likely though, with the knowledge that the free version will be uploaded elsewhere, fewer patrons would pay and market incentivization would decline. Additionally, for companies like Endel, who seek to monetize their investment via streaming services, this model would make that impossible, since streaming licensing is

153. *Id.*

154. *Id.*

155. *Id.*

156. See Kristen Senz, *Are Paywalls Saving Newspapers?*, HARVARD BUS. SCH. WORKING KNOWLEDGE (July 8, 2019), <https://hbswk.hbs.edu/item/are-paywalls-saving-newspapers>; see also Mike Masnick, *The Media's Paywall Obsession Will End In Disaster For Most*, TECHDIRT (May 8, 2018, 9:37 AM), <https://www.techdirt.com/articles/20180506/11501539779/medias-paywall-obsession-will-end-disaster-most.shtml> ("It's not that we think that paywalls are somehow "bad," but that (1) for most publications, they won't actually work and (2) they are quite frequently counterproductive.").

157. Adi Robertson, *Inside Patreon, The Economic Engine Of Internet Culture*, THE VERGE (Aug. 3, 2017, 11:36 AM), <https://www.theverge.com/2017/8/3/16084248/patreon-profile-jack-conte-crowdfunding-art-politics-culture>.

158. See, e.g., Phil Wang, *Phil Wang is creating This Person Does Not Exist*, PATREON, <https://www.patreon.com/lucidrains> (last visited Feb. 11, 2020) (the creator of thispersondoesnotexist.com currently has seventeen patrons at \$37 dollars a month as of February 2020).

predicated on copyrightability.¹⁵⁹ Without this assurance, large companies in those fields, like Warner Music, would be disincentivized to buy into the new industry. This would shift the market away from investment in music and video media, since those two areas are largely dominated by streaming services.

2. *AI as Both the Author and Owner*

The second model would have the AI be both owner and author. The need for this model is unclear because incentives are meant to entice human decision makers. Proponents are often thinking about the distant future in which robots and androids may have fully vested rights and possess the discretion over whether to produce future works.¹⁶⁰ But until the day that a robot expresses a desire for copyright on its works, this model only incentivizes programmers more interested in far-reaching philosophical and legal implications of granting non-human actors human rights than monetizing their creations.

This model would have some bizarre impacts on the market. It does not deal with issues of liability. It is also unclear how programmers who do seek incentive would reap the rewards of investment if all interests vest in the AI itself. The only logical element to come out of this model would be a fixed term limit, since AI do not have “lives.” Ultimately, it would likely chill investment by being even more unmanageable and ambiguous than the current regime.

3. *Modifying the Definition of an “Employee” Under the Work for Hire Regime*

Popular among American scholars is a modified WMFH regime. Overall, this model is not likely to chill investment, since it establishes all the copyright protections of a traditional work. This proposal essentially modifies the definition of an “employee” under 17 U.S.C. § 101 to include AI.¹⁶¹ As in traditional WMFH, copyright would then vest in the human or corporate entity that procured the work. Some scholars assume that the programmer is the employer,¹⁶² but others take the “person for whom the work was prepared” to mean either the programmer or the end user.¹⁶³ When attributing ownership

159. See *Music Licensing Modernization Act*, U.S. COPYRIGHT OFFICE, <https://www.copyright.gov/music-modernization/115/> (last visited Feb. 11, 2020).

160. See Andrew J. Wu, *From Video Games to Artificial Intelligence: Assigning Copyright Ownership to Works Generated by Increasingly Sophisticated Computer Programs*, 25 AIPLA Q.J. 131 (1997).

161. Hristov, *supra* note 5, at 447.

162. See, e.g., *id.* at 445.

163. 17 U.S.C. § 201(b) (2018).

to the programmer, the incentive clearly rewards the work done to create the AI, and also ensures that liability for infringement falls on the person best placed to prevent it.

But some scholars, like Professor Shlomit Yanisky-Ravid, suggest that when users interact with generative AI, they should be considered the owners of the works.¹⁶⁴ For tool-like AI, this makes sense, because the user is doing the lion's share of the work. But these works can already receive copyright protection because they involve substantial human creativity.¹⁶⁵ When it comes to truly generative works, it is unfair to attribute copyright to a user. Unless mitigated through licensing or contract, an unaware user clicking through a site like *thispersondoesnotexist.com*¹⁶⁶ could suddenly be on the hook for an infringing work that they had no part in producing, other than clicking “generate.”

This model also acknowledges the issue of protection term length. A traditional work is protected by copyright for the life of the author plus seventy years.¹⁶⁷ The term of copyright protection of a work made for hire is ninety-five years from the date of publication or one hundred twenty years from the date of creation, whichever expires first.¹⁶⁸ While still an excessively long period amount of time for AI works to be protected, this explicit limit nonetheless speaks to fairness within the copyright system and ensures works eventually enter the public domain.

This model aims to quickly address protecting AI works using an existing copyright mechanism. However, the Supreme Court has suggested that the WMFH doctrine is very limited in scope, and applies only where “Congress has expressed a clear and explicit intent to override section 102.”¹⁶⁹ Because implementing such a substantial change to the doctrine would require new legislation, it seems ineffective to try and push AI into a regime that is primarily about assignation of agency.¹⁷⁰ As of yet, AI are not sentient and do not have agency—so it is incongruous to incorporate them in a regime that requires a

164. Yanisky-Ravid, *supra* note 140, at 707 (“[W]e should view AI systems as working for the users, and hence the users should bear accountability for the systems’ production, in addition to the benefits thereof.”).

165. *See supra* Section II.B.

166. THIS PERSON DOES NOT EXIST, <https://thispersondoesnotexist.com/> (last visited Feb. 11, 2020) (a web-based GAN that generates a new human face each time the site is refreshed).

167. U.S. COPYRIGHT OFFICE, WORKS MADE FOR HIRE, *supra* note 133, at 3.

168. *Id.*

169. Yanisky-Ravid, *supra* note 140, at 715.

170. *See* U.S. COPYRIGHT OFFICE, WORKS MADE FOR HIRE, *supra* note 133, at 2 (devoting a section to “Agency Law”).

test to determine whether the work was actually made in an employment context and the employee was an “agent.” An AI and its subsequent works will always be attributable as a creation to an individual or a team. Instead of using an existing regime as a red herring to distract legislators from the actual changes to the law they would implement, it is worthwhile to explicitly and directly create a carve-out for AI works.

4. *Applying Existing U.K. Law to the United States*

Finally, some proponents advocate for the wholesale adoption of current U.K. copyright law, which treats computer-generated works as if they originated directly from the programmer. That law states that “[i]n the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken.”¹⁷¹ The United Kingdom’s Copyright Act goes on to define a computer-generated work as one that “is generated by computer in circumstances such that there is no human author of the work.”¹⁷² This provision creates an exception to human authorship by recognizing the time and resources that go into creating a generative AI, even if the actual work is undertaken by the machine.¹⁷³ Additionally, when it comes to the copyright term, the United Kingdom limits computer generated works to fifty years from the end of the year of creation.¹⁷⁴

Like the modified WMFH regime, this model grants full copyright protection and would therefore not chill investment. Because this regime fits how the market is functioning now in terms of programmers putting their names on AI-generated works,¹⁷⁵ it would not change applicable streams of revenue. The added benefit of a shortened term limit would ensure that programmers do not benefit for longer than is fair. Critics of this model note the ambiguity of the person who made the “arrangements necessary” for the work to be generated.¹⁷⁶ Depending on the interpretation, this could be the programmer or the user of the program. If the engineers or programmers of the AI are taken to be the copyright owners, return on investment is guaranteed and liability rests fairly on the creators.¹⁷⁷ However, this should be

171. Copyright, Designs and Patents Act 1988, c. 1, § 9(3) (Eng.).

172. *Id.* c. 10, § 178.

173. Guadamuz, *supra* note 151, at 4.

174. Copyright, Designs and Patents Act 1988, c. 1, § 12(7) (Eng.).

175. *See, e.g.,* ENDELIO, *supra* note 35.

176. *See* Bridy, *supra* note 24, at 27.

177. Pratap Devarapalli, *Machine Learning to Machine Owning: Redefining the Copyright Ownership from the perspective of Australian, U.S., UK and EU law*, 40 EUR. INTELL. PROP. REV. 722, 727 (2018).

clearly distinguished within the text of the law to limit disputes arising out of ambiguity.

B. PROPOSED MODEL: AI AS THE AUTHOR-IN-FACT AND THE PROGRAMMER AS THE AUTHOR-IN-LAW

This Note's proposed model lists the AI as the author-in-fact and the programmer as the author-in-law. This distinction serves to explicitly indicate that AI generated works will be treated differently than traditional copyrightable works and allows the Copyright Office to easily delineate between the two. This model would include a term limit—either fifty years to meet international standards¹⁷⁸ or fewer depending on what is deemed equitable by the legislature. For fair allocation of benefits and risks, the programmer, not the user, is explicitly listed as the author. Thus, the person reaping rights-based incentives also faces liability if their program is faulty. This model specifically incentivizes creators of generative AI who would be unable to monetize their software without monetizing the output.

Creators are already copyrighting their AI generated works and profiting off them. But there is no mechanism for enforcing the right against infringement nor does the term of life of the programmer plus seventy years accurately reflect the payoff of time and resources for the process of generation following the initial creation of the AI. Further, any legal challenge against a generated work would likely lead to its release into the public domain, as implied in *Naruto v. Slater*. A definitive regime needs to be put into place to clearly delineate the bounds of what rights are and are not applicable to AI-generated works. As proponents of the WMFH model suggest, it is expedient to rely on ties to existing U.S. copyright law. This model uses a legal fiction of authorship parallel to the one put forth by WMFH, in which copyright vests as a matter of law in someone who is not the author-in-fact. As with the WMFH and U.K. models, this proposed model would ensure that the market continues to encourage creators of AI, while the term limit reflects that each work is less impactful as a human-created work.

Finally, this modification would be made directly to the Copyright Office's "Copyrightable Authorship" section. Legislation is needed to make any of the changes listed in the above models (excepting public domain). It is unproductive for the legislature to decide on an adequate or easy-sounding, but problematic option, such as WMFH.¹⁷⁹ An independent clause on AI

178. See *supra* notes 146–50.

179. See, e.g., Hristov, *supra* note 5, at 452–53; see also Timothy L. Butler, *Can a Computer be an Author—Copyright Aspects of Artificial Intelligence*, 4 HASTINGS COMM. & ENT. L.J. 707, 741–

would not only clarify the distinctiveness of AI as a copyright matter, it would also speak to the willingness to address technological revolution head-on, in the footsteps of President Lincoln and Justice Holmes.

VI. CONCLUSION

It is important to encourage creative, not just utilitarian, AI. Artistic AI not only develops the full potential and breadth of AI but provides a point of insight into what creativity means, how humans reflect on creative works, and how alternative creative sources can drive new understandings and innovations. As AI-as-an-industry and source of creative works continues to expand, the need for clarification and a modified regime is pressing. A patch fix like adding “human” to authorship requirements or modifying the definition of “employee” to slip AI into an ill-fitting regime is inefficient. The coming update needs to keep the policies and historical trajectories behind copyright in mind. Congress and the courts have altered the doctrine and requirements in the face of new technologies before—this Note and its proposed changes argues for a continuing willingness to adapt the law to technological advances. While AIs do not dream of copyright protection, their creators and proponents do. They need a copyright regime that adequately reflects and delineates the protection of that dream.

42 (1982) (explaining that the contractual nature of WMFH imposes limitations on redefining the meaning of “employee”).

THE BREWING BATTLE: COPYRIGHT VS. LINKING

Marta Rocha[†]

I. INTRODUCTION

Imagine going to an online article, reading a citation of a tweet by President Trump, but not being able to see the tweet embedded within the article itself. Instead, you would have to navigate to Twitter's website to confirm that the tweet actually exists and that it came from the President himself. Next, imagine attempting to find a picture on the internet without Google Images showing an embedded preview of that image.¹

These tasks might seem rather unsurmountable without the convenience of previewing embedded images. Those are just two examples of what the internet would be like if copyright law rendered embedding illegal by labeling it an infringement of copyright holders' rights. Over the past few years in the United States and the European Union, judicial decisions and newly-passed laws have threatened free access to information on the World Wide Web by potentially limiting embedding practices. Such limitations may create a slippery slope to prohibiting other types of linking, which would severely curtail the ease of finding information on the internet for an average user.

This Note ultimately argues that we should aim to preserve content embedding on the internet in its current form. The negative consequences of stringent copyright protections in linking significantly outweigh their likely positive effects. Banning embedding on the internet will likely skew the incentive-access balance,² critically inhibit competitive practices on the internet, upset the settled expectations of the web and potentially, in the long term, lead to creation of "walled gardens."³ In order to account for the economic incentives currently given by copyright law to copyright holders, this Note also proposes a licensing scheme that would ensure compensation to

DOI: <https://doi.org/10.15779/Z38B56D54Q>

© 2020 Marta Rocha.

[†] Marta Sylwia Rocha, J.D. UC Berkeley Law, B.S. UW Madison.

1. According to Ben Ling, then Google's Director of Search Products, in 2010 Google Images had hundreds of millions of users. MG Siegler, *Google Image Search: Over 10 Billion Images, 1 Billion Pageviews A Day*, TECHCRUNCH (July 20, 2010), <https://techcrunch.com/2010/07/20/google-image-search/>.

2. Intellectual property protections assume a balance between incentives created through exclusive rights for authors and access to creative works that will spur further creation of content. *See infra* Section V.A.2.b).

3. *Escaping the Walled Gardens in the Clouds*, TECH-FAQ, <http://www.tech-faq.com/escaping-the-walled-gardens.html> (last visited Feb. 17, 2020).

authors, creators and copyright holders. Such a scheme would fill a void of economic incentives that copyright has not been able to fill.

Due to the magnitude of these potential negative consequences, this Note concludes with three recommendations. First, the legislature should address existing gaps in copyright law and allow an exception from copyright infringement for embedding on the internet. Second, the legislature should create a licensing system that will allow for economic reimbursement of authors and copyright holders, without putting a strain on the existing reference techniques.⁴ Finally, the Note advocates for courts to recalibrate how they consider the goals of copyright to better reconcile them with the need for an open and thriving internet.

To fully understand why such a licensing scheme is more desirable than the ever-expanding copyright regime, this Note lays out the necessary background on the interaction of copyright law and linking on the internet. More specifically, Part II of this Note discusses the goals of the internet and copyright law. It fleshes out the different motivations behind copyright protections both in the United States and Europe. Part III provides the necessary technological understanding of different types of linking on the internet. Part IV shows how these different types of linking are treated by the courts in different circuits in the United States and in light of the newly European Directive on Copyright in the Digital Single Market. Part V shows that the goals of the internet and copyright law can and should be reconciled, at least in the United States, and that such a reconciliation would have to account for the long-ignored access to information requirement of intellectual property protections.

II. THE INTERNET, COPYRIGHT, AND THEIR FUNDAMENTAL GOALS

A. THE GOALS OF THE INTERNET

The internet permeates all aspects of our lives as we know them today. Everything including banking, shopping and communicating depends on our access to the web. Just as the internet is integral to our everyday existence, free access to information on the web is fundamental to the internet's structure. The history of the internet is infused with the idea of access, both to technological information critical to its evolution (such as protocols), as well

4. "Reference techniques" is a phrase used throughout this Note to refer to linking, embedding, and other technological ways to reference materials on the internet, which allow for easy navigation on the internet.

as information for everyday users.⁵ One of the internet's visionaries, J.C.R. Licklider, conceived of the internet as "a globally interconnected set of computers through which everyone could quickly access data and programs from any site."⁶ The continual success of the internet and its development draws from its openness. The internet's interconnectivity is what facilitates that openness. And the ability to reference other sources through linking techniques is the main facilitator for that interconnectivity.

Before the existence of search engines, the internet was difficult to navigate.⁷ The information was available on the web, scattered across servers around the world.⁸ But it was difficult to locate.⁹ By indexing each site across the internet, search engines made information more easily accessible.¹⁰ Search engines helped forge a link between the internet's open information and the users searching for it.¹¹ Linking and other reference techniques quickly became one of the internet's key features that made the enormous amount of information not only available, but more importantly, discoverable to everyday users.¹²

The latest developments in copyright protection and linking in the United States and Europe, however, threaten to severely impede such reference techniques.¹³ More stringent copyright protections on the internet pose a threat to linking techniques as they exist today, thus undermining the way internet users find information. The destruction of fundamental reference techniques threatens the basic navigability of the web. If certain information becomes virtually unavailable due to how difficult it is to find, the internet will divide into silos. These silos, commonly referred to as "walled gardens,"¹⁴ would limit end users to only access particular type of information based on their starting point. Alternatively, the information, although theoretically accessible, would be rendered virtually undiscoverable.

5. See Bary M. Leiner, Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard Kleinrock, Daniel C. Lynch, Jon Postel, Larry G. Roberts & Stephen Wolff, *A Brief History of the Internet*, 39 ACM SIGCOMM COMPUTER COMM. REV. 22 (Oct. 2009).

6. *Id.* at 23.

7. Steven Vaughan-Nichols, *Before Google: A history of search*, ENTERPRISE.NXT (Mar. 24, 2017), <https://www.hpe.com/us/en/insights/articles/how-search-worked-before-google-1703.html>.

8. *Id.*

9. *Id.*

10. *Id.*

11. *Id.*

12. *Id.*

13. See *infra* Part IV for the discussion of the developments.

14. See *supra* note 3.

To understand the potential repercussions of strict copyright enforcement on linking, one must understand the basics of the technology and the legal background of these issues. To provide a holistic picture and understanding of the topic, the sub-sections below discuss how copyright law governs linking, the ideas behind linking, and how different jurisdictions have dealt with issues of copyright infringement in this context.

B. THE GOALS OF COPYRIGHT LAW

Some of the goals of copyright law and the internet seem fundamentally irreconcilable. The internet's openness and free access to information seems to go against copyright law as it been recently interpreted by the courts. The courts have become too preoccupied with copyright's grants of exclusive rights of ownership to the authors, while forgetting the second part of the equation: providing access to the public. According to the U.S. Constitution, the goal of all intellectual property protections is "to promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries."¹⁵

Over time the court's interpretation of "progress" as pertaining to copyright has evolved to become synonymous with incentivizing content creation.¹⁶ Arguably, more content does not necessarily guarantee the quality of the content that promotes progress of the arts, just like economic incentives themselves do not guarantee the quality.¹⁷ However, in *Bleistein v. Donaldson Lithographing Co.*, the Supreme Court settled the matter by stating that copyright protections should not rest on the evaluations of quality of particular piece of work.¹⁸ The Supreme Court noted that "[i]t would be a dangerous undertaking for persons trained only to the law to constitute themselves final judges of the

15. U.S. CONST. art. I § 8, cl. 8.

16. See *Bleistein v. Donaldson Lithographing Co.*, 188 U.S. 239, 251–52 (1903) ("It would be a dangerous undertaking for persons trained only to the law to constitute themselves final judges of the worth of pictorial illustrations, outside of the narrowest most obvious limits.").

17. See Diane Leenheer Zimmerman, *Copyrights as Incentives: Did We Just Imagine That?*, 12 THEORETICAL INQUIRIES L. 29, 34 (2011) (arguing that economic incentives provided by copyright protections do not actually promote creativity as "large numbers of creators have little hope of ever gaining the economic rewards promised by the intellectual property regime"); Orrin G. Hatch & Tomas R. Lee, "To Promote the Progress of Science": *The Copyright Clause and Congress's Power to Extend Copyrights*, 16 HARV. J.L. & TECH. 1, 3 (2002) ("This understanding — which views "progress" as encompassing not only an increase in quantity or quality of works, but also an improvement in the dissemination and preservation of works already in existence — finds support in founding-era usage of the constitutional language, in the structure of the Constitution, and in the historical exercise of the copyright power.").

18. *Bleistein*, 188 U.S. at 251–52.

worth of pictorial illustrations.”¹⁹ Accordingly, courts’ interpretations of progress have evolved to become synonymous with incentivizing content creation, without regard to the quality of the work itself.²⁰ Therefore, the overarching goal of U.S. copyright protections can be viewed as incentivization for continual production of content (whether measured by quantity or quality).

Regardless of its quantity or quality, content is the primary motive behind intellectual property protections, it is critical that more people should have access to art and information. The internet provides users with the ability to easily access the “sciences and useful arts,” thus fulfilling one of the constitutional principles.²¹ The connectedness of the internet is therefore critical to fulfilling at least one of the underlying constitutional principles of intellectual property protections.

Although economic incentives are likely to always be a mechanism for spurring the production of arts,²² they should not be the sole goal of copyright protections. Rather, copyright laws should aim to “promote” progress in the arts and sciences,²³ not protect the economic interest of the artists.²⁴ An equilibrium between those goals, however, can be reached by providing economic incentives that do not put limits on the public access to information.²⁵ Over the years, protecting the artist and incentivizing content creation have become so inherently intertwined they are often treated as one and the same. The internet, however, has made it possible to distinguish those goals.

On the other hand, copyright law in the European Union focuses on the rights of the creator to control how their works are used. This focus can be detrimental to linking because it tends to protect copyright at all costs.²⁶ The E.U. laws harmonize intellectual property protections across the E.U. member states, although each country continues to possess its own intellectual property

19. *Id.*

20. *See, e.g.,* *Drop Dead Co. v. S. C. Johnsons & Son, Inc.*, 326 F.2d 87, 92 (9th Cir. 1963); *Ansehl v. Puritan Pharm. Co.*, 61 F.2d 131, 134 (8th Cir. 1932).

21. *See* U.S. CONST. art. § 8, cl. 8.

22. Although authors might be willing to produce their work for free given other incentives, the production of movies, or creation of online newspaper content (as just two examples) would be impossible without financial backing that allows the organizations to pay its staff and keep the lights on.

23. U.S. CONST. art. I § 8, cl. 8.

24. *See* Oren Bracha, *The Ideology of Authorship Revisited: Authors, Markets, and Liberal Values in Early American Copyright*, 118 YALE L.J. 186, 203 (2008) (in which “Justice Story turned to the market as the sole arbiter of value” in *Emerson v. Davies*, 8 F. Cas. 615 (C.C.D. Mass. 1845)).

25. *See infra* Section V.C.

26. *See* International Convention for the Protection of Literary and Artistic Works, Sept. 9, 1886, 25 U.S.T. 1341; 1161 U.N.T.S. 3 [hereinafter *Berne Convention*].

protections.²⁷ Nonetheless, E.U. copyright law continues to embrace the strong presence of authors' rights in copyrightable material, despite copyright's growing economic importance.²⁸ A prime example is the protection of authors' moral rights—a set of rights that, in part, protect the authors' right of attribution and prevent the artwork from derogatory treatment.²⁹

While U.S. intellectual property protections focus mainly on promoting content creation, both the United States and Europe rely on economic incentives to motivate content creation. But in the United States, the separation of economic incentives and content creation seems possible, by providing other motivations for creation.³⁰ The European Union's focus on authors' moral rights makes such a separation almost impossible, because it considers the exclusive rights of the author one of the primary goals of copyright protections, emphasizing the need to reward authors for their efforts in creating their works.³¹

III. TECHNICAL BACKGROUND ON LINKING

Armed with a basic understanding of the underlying goals of the copyright law both in the United States and the European Union, this Part discusses different forms of linking on the internet. These differences are critical to understanding why copyright jurisprudence has primarily focused on embedded links rather than ordinary or deep links.

For the purposes of this Note, the linking on the internet can be categorized into three basic types: ordinary, deep, and inline (also known as embedding).³² The goal of all three types is to direct users to material that is pertinent, but not necessarily hosted on the website they are browsing at the

27. EU COPYRIGHT LAW: A COMMENTARY 8 (Irina Stamatoudi & Paul Torremans eds., 2014).

28. *Id.* at 15.

29. Jonathan Bailey, *U.S. vs. Europe: Moral Rights*, PLAGIARISM TODAY (June 12, 2006), <https://www.plagiarismtoday.com/2006/06/12/us-vs-europe-moral-rights/>.

30. See Margaret Chon, *Postmodern Progress: Reconsidering the Copyright and Patent Power*, 43 DEPAUL L. REV. 97, 105 (1993) (stating that the Supreme Court has recognized that “provision of incentives to authors and investors will not always coincide with the underlying objectives of [incentivizing progress]”).

31. In Europe, the protection of authors moral rights (rights that include the right to the integrity of the work) is incorporated into Article 6bis of the Berne convention. Berne Convention, *supra* note 26.

32. W3C Consortium, “*Deep Linking*” in the World Wide Web, W3C (Tim Bray ed., Sept. 11, 2003), <https://www.w3.org/2001/tag/doc/deeplinking.html>; Richard Stim, *Linking, Framing, and Inlining*, NOLO, <https://www.nolo.com/legal-encyclopedia/linking-framing-inlining-30090.html> (last visited May 5, 2020).

moment.³³ I refer to the website that is embedding material as the “embedding website,” and the website being linked to as the “source website.”

Ordinary and deep links appear very similar to the end user.³⁴ Both look like interactive text within a website.³⁵ When the user clicks an ordinary or deep link, the link sends the user to the source website.³⁶ The key difference between the two lies in where the user is directed.³⁷ Deep links direct the user to a page other than the home page of the source website, while ordinary links send the user to the source website’s homepage.³⁸

For example, a link that directs a user to a specific product on the Amazon website, instead of directing them to www.amazon.com, is a deep link. Deep linking is more controversial in the world of copyright law, as it often bypasses the main pages of the source website that might include advertisements or other ways of earning money.³⁹ It is however, fundamental to finding information, as it allows the website creator to direct the user to the exact content they are referencing.⁴⁰ This saves the user the trouble of trying to find that same information on the source website. Imagine a user who finds a list on a website of “Top 30 best Christmas gifts from Amazon for 2019” but every time they click a link from that website, it directs them to the Amazon home page, instead of to the exact item they were interested in. Chances are they will quickly give up on such a list of ordinary links, and instead move onto a list that contains deep links to the specific products that they’re interested in purchasing. The information provided by such a list will be inefficient at providing access to the desired information.

The third type of linking is by far the most controversial in the copyright world. Unlike the other forms of linking, inline linking, commonly referred to as embedding, does not redirect the user to a different webpage to view the copyrighted content. Instead, inline links embed the copyrighted material from another website into the web page that the user is viewing.⁴¹ For example, when a user views pictures on Google Images, those pictures are embedded in

33. See Richard Stim, *supra* note 32.

34. See W3C Consortium, *supra* note 32.

35. *Id.*

36. *Id.*

37. *Id.*

38. *Id.*

39. See, e.g., *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 816 (9th Cir. 2003) (first decision in United States pertaining to deep linking of copyrightable material where plaintiff argued that deep links of images caused them economic harms).

40. See W3C Consortium, *supra* note 32.

41. Ross Bagley, *Rules on Copyright Infringement for Inline Linking Developing in the United States and Abroad*, IPWATCHDOG (Nov. 12, 2019), <https://www.ipwatchdog.com/2019/11/12/rules-copyright-infringement-inline-linking-developing-united-states-abroad/id=115691/>.

Google's website. It is only after the user clicks on one of the images multiple times that they get redirected to the website where the image originated. Those images are thus not only embedded but also function as deep links to the source pages. Search engines are not the only services heavily reliant on embedding. News articles, hobby websites like Pinterest and Reddit, and coding repositories such as Github also rely on the ability to embed content on their sites.

For the convenience of legal distinction this Note further divides Embedding into two subcategories. Although at a first glance these subcategories do not appear any different to the end user, they are fundamental to understanding how copyright law has been applied differently across jurisdictions in the United States.⁴² The copyrighted materials in both types of embedding appear to be part of the website. However, the difference lies in who is hosting the material.⁴³ From the example above, the picture shown on Google Images can be either linked to via the source website ("image source linking"), or it can be copied to the server of the embedding website, which the user is viewing ("server storage embedding").⁴⁴

This can be analogized to the Mona Lisa in the Louvre Museum. Imagine another museum wanted to display the Mona Lisa without stealing it or renting it from the Louvre. In the digital world they could embed the picture via one of the two options: server storage embedding or image source linking. The second museum could image source link by setting up cameras in the Louvre and live streaming the painting into their own museum. Thus, the embedding museum would never have to make a copy of the Mona Lisa. The embedding museum would simply create a direct link to the painting in its original form.

As a second option, the embedding museum could create a copy of the painting, put it in their basement and live stream the painting onto their display. Creating a copy of the work is analogous to server storage embedding. Although this type of linking is relatively rare, and the difference is imperceptible to the viewer, it has created a tension in the U.S. courts about copyright infringement.⁴⁵ While content source linking is a popular practice (for example, tweets embedded in news articles), server storage embedding is virtually unused. It was essentially eradicated by current legal regimes, as it

42. *See infra* Section IV.B.

43. *See infra* Section IV.B.2.

44. Both "image source linking" and "server storage embedding" are not technical terms, but rather terms created in this Note for ease of understanding and explaining current legal distinctions.

45. *See infra* Section IV.B.1.

constitutes copyright infringement.⁴⁶ This is also a key differentiation between embedding as it happens on the internet today, and illegally downloading and reposting copyrightable materials. While server storage embedding and source linking at a first glance seem indistinguishable to the end user, only image source links provide deep linking functionality that links back to the source website.

This Note focuses on the importance of protecting content source linking (from now on also referred to as embedding) as it stands today. Unlike server storage embedding, content source linking continues to serve as a deep link for the source website. The server storage embedding does not fulfill that function, because it does not have the ability to send the end user back to the source website. Unlike server storage embedding, image source linking allows the user to find the origins of the embedded materials, thus allowing the source website to obtain ad revenue. But what does embedding offer that cannot be substituted by other forms of linking?

There are three major advantages that embedding offers over any other type of linking on the internet. All these advantages benefit the end users, but some of them also provide benefits to authors, copyright holders, and website providers. The first and most obvious benefit of embedding is the convenience it provides for the user. Embedded materials provide only pertinent pieces of the source website included in the embedding website that the user is reading. That allows the user to easily review and comprehend all the information, without ever having to open a separate window. Embedding therefore saves users time and prevents information overload.

Second, links are easier to break—meaning the source page changes and the established link no longer points to the correct page—than embedded material, because the embedded information is static, meaning it will not change after embedding without additional action from the embedding website.⁴⁷ Therefore, links are more likely to incur higher maintenance costs for the website providers. Additionally, if a link breaks, the copyrighted material is not receiving as many views as it would if it was embedded, providing both better access for the end user and exposure for the author.

46. See *infra* Section IV.B.1.

47. Jeyakanth, *Difference Between Linking and Embedding*, DIFFERENCEBETWEEN.NET, <http://www.differencebetween.net/technology/software-technology/difference-between-linking-and-embedding/> (last updated Aug. 1, 2017).

Finally, embedding content can increase engagement, which is beneficial for authors, users, and website providers.⁴⁸ Embedding makes posts stand out to the user and provides clarity for the purpose of the embedded content.⁴⁹

IV. CURRENT INTERPRETATIONS OF LINKING AND COPYRIGHT LAW

Copyright law in America is constantly evolving.⁵⁰ It is defined in the Copyright Act of 1976 and its subsequent amendments, including the Digital Millennium Copyright Act, the Semiconductor Chip Protection Act of 1984, the Vessel Hull Design Protection Act, and others.⁵¹ The continuous changes and amendments to copyright law show that the legislature recognizes that copyright law does not adapt well to unprecedented technologies, and therefore cannot be effective while remaining static. The latest addition to copyright protection, the Music Modernization Act (MMA), further demonstrates the continuing need to adapt old standards to newly developing technologies.⁵² This Part focuses on the details of copyright infringement and how it relates to linking and embedding. It addresses the recent circuit split in the United States, and the Tax Link Directive from the European Union.

In the United States, copyright law protects “original works of authorship,” “fixed in a tangible medium.”⁵³ This includes music, literary works, pictures, graphics, sculptures, audio-visual works, sound recordings and architectural works.⁵⁴ Copyright protects the expression of ideas, but not the ideas themselves.⁵⁵ It also grants authors, for a limited time, a set of exclusive rights to reproduce, prepare derivative works, distribute copies, publicly perform, and publicly display their works.⁵⁶ Copyright is automatically given to the author of an original work without any registration.⁵⁷ If the copyright

48. BigCommerce Essentials, *What is embedding and when to do it on Facebook and Twitter*, BIGCOMMERCE, <https://www.bigcommerce.com/ecommerce-answers/what-is-embedding/> (last visited Mar. 17, 2020).

49. *Id.*

50. For example, see the newest advancement in the copyright realm, the Music Modernization Act. *See* Orrin G. Hatch-Bob Goodlate Music Modernization Act, H.R. 1551, 115th Cong. (2018).

51. Copyright Act of 1976, 17 U.S.C. §§ 101–1332 (2012).

52. *See* 17 U.S.C. § 115.

53. *Id.* at § 102.

54. *Id.*

55. *Baker v. Selden*, 101 U.S. 99 (1879); *see also* *CDN Inc. v. Kapes*, 197 F.3d 1256, 1261–62 (9th Cir. 1999).

56. 17 U.S.C. § 106.

57. United States Copyright Office, *Copyright in General*, COPYRIGHT.GOV, <https://www.copyright.gov/help/faq/faq-general.html> (last visited Feb. 7, 2020).

holder's copyrights are infringed, the Copyright Act allows for injunctions,⁵⁸ equitable relief,⁵⁹ damages,⁶⁰ attorney's fees,⁶¹ and criminal penalties.⁶²

Copyright protections are not boundless, however. They are limited by such concepts as the first sale doctrine,⁶³ fair use,⁶⁴ various safe harbors,⁶⁵ and many more.⁶⁶ The fair use doctrine is an affirmative defense to infringement that is especially relevant to the internet. Courts consider the following four elements when determining whether a would-be infringer's use of copyrighted material was "fair": (1) the purpose of use was not of commercial nature; (2) the nature of the copyrighted work used was more factual than imaginative; (3) the amount of copyrighted material used is not disproportionate; and (4) the effect on the copyrighted materials potential market is minimal.⁶⁷ When considering the fair use elements, courts also focus on whether the use of copyrighted material was transformative.⁶⁸ This approach virtually eliminates the possibility for embedding to constitute fair use, as embedding displays the information in an unaltered form from its original source, and although the use can sometimes be transformative,⁶⁹ that is unlikely to happen often.

A. COPYRIGHT INFRINGEMENT AND ITS APPLICATION TO LINKING AND EMBEDDING

The main controversy at the intersection of copyright law and linking is whether linking infringes the copyright holder's right to exclusive control over their works. To prevail in an infringement lawsuit, the plaintiff has to show

58. 17 U.S.C. § 502 (allowing for both preliminary and permanent injunctions).

59. On top of injunctions, the Copyright Act also authorizes a seizure order, which allows for impoundment of all the copies of the infringing product. *See* 17 U.S.C. § 502.

60. 17 U.S.C. § 504 (allowing the author to recover for either actual damages and defendant's profits, or statutory damages).

61. 17 U.S.C. § 505.

62. The Copyright Act provides for criminal penalties for "willful and for profit" copyright infringement. *See* Copyright Act of Jan. 6, 1897, ch. 4, 29 Stat. 481.

63. 17 U.S.C. § 109.

64. *Id.* § 107.

65. For example, the Digital Millennium Copyright Act (DMCA), 17 U.S.C. § 512 provides a safe harbor from third-party liability for online service providers (OSPs).

66. For example, built into the Copyright Act are exceptions for public broadcasters, libraries, and software backup. 17 U.S.C. §§ 108, 110, 117.

67. Stanford University Libraries, *Measuring Fair Use: The Four Factors*, STANFORD UNIVERSITY, <http://fairuse.stanford.edu/overview/fair-use/four-factors/> (last visited Feb. 7, 2020).

68. Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV 1105, 1111 (1990).

69. *Kelly*, 336 F.3d at 818.

under 17 U.S.C. § 106 that their exclusive rights in a copyrighted work were violated.⁷⁰

Central to this debate is § 106(5)⁷¹ which grants the copyright owner an exclusive right to “display the copyrighted work publicly.”⁷² The concept of publicly displaying the work is defined in 17 U.S.C. § 101, and requires a showing of the work (1) “at a place open to the public,” or by (2)

transmit[ting] or [. . .] communicat[ing] a performance or display [. . .] 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.⁷³

So, in the Mona Lisa example in the Section above,⁷⁴ live-streaming the paintings would fulfill the public display requirement of 17 U.S.C. § 106.

To further understand the relationship between copyright infringement and linking, the following subsections provide an overview of potential defenses to internet copyright infringement, and how each type of linking has been treated in the copyright infringement jurisprudence. The first Subsection illustrates the potential repercussions of strengthening or weakening copyright protections. It shows that the protections currently offered by copyright law are inadequate. The second Subsection discusses which types of linking practices U.S. courts are most likely to consider copyright infringement.

1. Potential Defenses to Copyright Infringement

Although the current copyright regime provides some level of protection to linking practices on the internet, this Section shows that those protections are neither sufficient nor efficient at protecting fundamental reference techniques. The main controversy with online linking is whether it violates the copyright holders’ right to control public displays of their works. This right is violated when the author can show that their exclusive rights in a copyrighted work were violated.⁷⁵ Currently, there exist two main defenses that could be utilized to avoid liability for embedded copyrighted content: (1) the DMCA’s § 512 safe harbor defense and (2) licenses (both express and implied).⁷⁶

70. 17 U.S.C. § 106 (2012).

71. *See infra* Section IV.B.

72. 17 U.S.C. § 106.

73. *Id.* § 101.

74. *See supra* Part III; 17 U.S.C. § 106.

75. 17 U.S.C. § 106.

76. *Id.* § 512.

In their article “Embedding Content or Interring Copyright: Does the Internet Need the ‘Server Rule?’” Jane C. Ginsburg and Luke Ali Budiardjo lay out these two defenses, arguing that they could be utilized to protect embedding from infringement.⁷⁷ They argue that the aforementioned defenses could protect would-be infringers, minimizing the impact of a change in how copyright protections on the internet are viewed by the courts.⁷⁸ According to their argument, potential infringers could shield themselves from liability using these two defenses, leaving the internet unable to utilize linking to a large extent as we know it today.⁷⁹ But because their interpretations are vulnerable to the whims of the courts, neither of these two defenses seems very promising at protecting embedding. Even if § 512 safe harbors or licenses were promising defenses, relying on them to justify embedding is undesirable in the long run.

DMCA § 512 limits the liability of internet service providers for online copyright infringement.⁸⁰ Although DMCA § 512 does not absolve online service providers of any copyright infringement liability, it creates safe harbors based on the type of services a website provides.⁸¹ A website provider can receive a safe harbor protection from copyright infringement as an intermediary if they fulfil certain requirements of the DMCA.⁸² Amongst other things, these requirements demand that ISPs have no knowledge or financial benefit from the infringing activity and that they comply with any takedown notice served by the copyright holder.⁸³ This provides no room for protection for embedding materials on the internet, especially if the embedding activity cannot be separated from the advertising activity of the embedding website. Therefore, the content distributors could not seek safe harbor under DMCA

77. See generally Jane C. Ginsburg & Luke Ali Budiardjo, *Embedding Content or Interring Copyright: Does the Internet Need the “Server Rule”?*, 42 COLUM. J. L. & ARTS 417 (2019). For more information on the server rule, see *infra* Section IV.B.1.

78. *Id.* at 474.

79. *Id.*

80. 17 U.S.C. § 512(l).

81. 17 U.S.C. §§ 512(a)–(d).

82. The website service provider, for example, must:

adopt[] and reasonably implement[], and inform[] subscribers and account holders of the service provider’s system or network of, a policy that provides for the termination in appropriate circumstances of subscribers and account holders of the service provider’s system or network who are repeat infringers; and accommodate[] and do[] not interfere with standard technical measures.

17 U.S.C. §§ 512(i)(1)(A)–(B). Additionally, the online service provider is also required to (1) not receive a financial benefit from the infringing activity; (2) not have knowledge or awareness of the infringing material; (3) upon receiving notice of infringement act “expeditiously to remove or disable access to” the infringing material. 17 U.S.C. §§ 512(c)(1)(A)–(E).

83. 17 U.S.C. §§ 512(c)(1)(A)–(E).

§ 512(c)–(d)⁸⁴ if they knowingly embedded copyrighted material or they did not comply with the takedown request from the copyright holder.⁸⁵ Additionally, in order for this defense to provide the appropriate breadth of immunity⁸⁶ to leave the reference techniques widely available, the courts would still have to determine: (1) if every type of linking was meant to be protected under DMCA § 512, and (2) what constitutes a “service provider.”⁸⁷

If the courts deemed that embedding cannot be protected under the safe harbor created by DMCA § 512, the defense could not be used in the vast majority of linking-copyright infringement cases. On the other hand, if the courts answered the first question affirmatively, while narrowly defining the term service provider, the defense would only be available to large providers of online services, while exposing smaller internet sites and their owners to potential copyright liability.⁸⁸ The availability of this defense therefore relies on a judicial decision on whether the defense is applicable in these types of cases and to what degree. Leaving this up to courts that currently cannot agree on how to treat law on the internet would likely lead to a circuit split, without addressing the underlying issues. And because the internet is so fundamental to individuals’ everyday lives, the rights of access to information on the web should be consciously determined by the legislature, rather than left to a piecemeal evolution in the courts.

The second potential defense is express and implied licenses. In order to be covered by licenses, a website provider would have to obtain express written or implied permission to do something.⁸⁹ An express licensing system would allow the copyright owners to determine what type of sharing they are willing to expose their content to. Although some websites already account for express licenses in their Terms of Service,⁹⁰ a new licensing system that would account for display rights would still have to be developed. An implied license to embed materials, although particular to each site responsible for content

84. DMCA § 512(d) limits liability “. . . for infringement of copyright by reason of the [service] provider referring or linking users to an online location containing infringing material or infringing activity, by using information location tools, including a directory, index, reference, pointer, or hypertext link . . .” 17 U.S.C. § 512.

85. Ginsburg, *supra* note 77, at 474.

86. An appropriate breadth of immunity is one that would keep the information on the internet at least as easily accessible as it is today.

87. 17 U.S.C. § 512.

88. Ginsburg, *supra* note 77, at 453–54.

89. *See, e.g.,* United States v. Univis Lens Co., 316 U.S. 241, 249 (1942) (implying a license as “an incident to the purchase of [an article]”).

90. *See Terms of Service*, TWITTER, <https://twitter.com/en/tos> (last visited Feb. 9, 2020) (“. . . you grant [Twitter] a worldwide, non-exclusive, royalty-free license . . . to use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute such Content . . .”).

creation, could be utilized by the defendant in copyright infringement if a copyright license could “be implied from conduct.”⁹¹

Two potential types of conduct could imply a license on the internet: (1) posting content on the internet, or (2) not opting-in to using currently available technology to block linking.⁹² The courts, however, are unlikely to embrace the first conduct as a standard due to the broadness of its implications. If courts were to embrace such a standard, anytime anyone posted copyrighted work on the internet, the copyright holder would lose their copyright protections. The latter standard, although more likely to be adopted, would cause a dramatic shift in copyright by creating an opt-in, rather than automatic, system for copyright protections. Currently, copyright is granted automatically.⁹³ But by requiring the authors to affirmatively protect their work on the internet, the copyright protections could only be automatically lost, not granted.

These defenses are not likely to provide the necessary amount of coverage to truly preserve reference techniques for several reasons. First, relying on DMCA safe harbors likely requires adjudication by the courts, which will put new strain on the already congested judicial system.⁹⁴ Second, the current system creates an institutional advantage for large, established companies who have the funds to defend or pursue these kinds of suits.⁹⁵ Third, although licensing systems do not necessarily require adjudication, the negotiation of individual express licensing systems between each ISP and copyright holder would simply be impractical. Overall, it is better to adapt the law to current practices on the internet, than to “Frankenstein” the existing law into something that will provide the protections necessary for the internet’s continued connectedness.

91. MELVILLE NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 10.03(A)(7) (2006).

92. Ginsburg, *supra* note 77, at 469–70.

93. COPYRIGHT.GOV, <https://www.copyright.gov/help/faq/faq-general.html> (last visited May 6, 2020).

94. Currently, reference techniques such as embedding are protected by the judicially created “server rule” which requires a copy to be made of a copyrighted work, to constitute infringement on the internet. Therefore, the adjudication necessary to settle DMCA safe harbors claims would create an additional set of cases for the court to address, which currently does not fall into that realm. *See infra* Section IV.B.1.

95. *See, e.g.,* Adrian Covert, *Apple and Google Spent More Money on Legal Fees Than R&D Last Year (And Google Apparently Thinks Apple Wants It That Way)*, GIZMODO (Oct. 8, 2012), <https://gizmodo.com/apple-and-google-spent-more-money-on-legal-fees-than-r-5949909>; Jeremy Horwitz, *Apple’s former top lawyer: \$1 billion budget enabled high-risk strategies*, VENTUREBEAT (June 10, 2019), <https://venturebeat.com/2019/06/10/apples-former-top-lawyer-1-billion-budget-enabled-high-risk-strategies/>.

2. *How Are Different Types of Linking Treated?*

Generally, U.S. courts have held that simple and deep linking do not constitute any copyright infringement.⁹⁶ In *Ticketmaster v. Microsoft*, Ticketmaster sued Microsoft for bypassing their home pages with deep links, arguing that such an action diluted its value.⁹⁷ Microsoft brought up several defenses, including that Ticketmaster had breached an accepted internet custom that allows linking to other people's sites.⁹⁸ The case eventually settled on confidential terms. A later, similar case, *Ticketmaster Corp. v. Tickets.com, Inc.*, picked up the issue. There, the district court had to determine whether the defendant's deep linking constituted an unauthorized public display of the plaintiff's event pages.⁹⁹ The Ninth Circuit held in favor of deep linking, while pointing out that copyright does not protect facts.¹⁰⁰

While deep and ordinary links redirect the end user to the source website, clearly letting the user know that they are being shown content from another site, embedding makes the information appear as part of the embedding website. Thus, often the embedded information does not appear any different than non-embedded information. It is precisely because of this inability to distinguish between linking and source website information that the majority of jurisprudence in the United States focuses on embedding.

B. POTENTIAL EMBEDDING CIRCUIT SPLIT

Although copyright law in the United States has undergone many changes, it has largely failed to adapt to the internet age. Unable to find consensus, courts developed two approaches for addressing the interaction of copyright infringement with one of the internet's most central tools. The Ninth Circuit and district courts in the Second and Fifth Circuits are at odds with each other, which is likely to lead to a circuit split. Sooner or later the Supreme Court or the legislature will have to resolve these differences. Over the years, the Ninth Circuit embraced the so called "server test,"¹⁰¹ while the district courts in other circuits recently advocated for the "incorporation test."¹⁰² The two tests are

96. See *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 816 (9th Cir. 2002).

97. *Microsoft's Link to Ticketmaster Site Spurs Trademark Lawsuit*, Computer & Online Industry Litigation Reporter, at 24087 (May 6, 1997), <https://cyber.harvard.edu/metaschool/fisher/linking/linking/link3.html>.

98. *Id.*

99. *Ticketmaster Corp. v. Tickets.Com Inc.*, 2 F. Appx 741, 741 (9th Cir. 2005).

100. See *Feist Publications, Inc v. Rural Telephone Service Co.*, 499 U.S. 350, 358 (1991) (holding that addresses and telephone numbers do not have copyright protection, because they are mere facts).

101. See *infra* Section IV.B.1.

102. See *infra* Section IV.B.2.

diametrically opposite. In the competition for survival between the two tests, the server test should persevere, as it fits already settled expectations about internet linking.

1. *Ninth Circuit and the Server Test*

The Ninth Circuit's server test is a bright line rule for copyright infringement in linking. In *Kelly v. Arriba Soft*, the Ninth Circuit ruled that deep linking from thumbnails was not infringement because it was highly transformative and therefore fell within the purview of fair use.¹⁰³ In *Perfect 10 vs. Amazon*, the Ninth Circuit again considered whether an image search engine's use of thumbnails could be considered fair use.¹⁰⁴ Once again, the court found the use of thumbnails highly transformative, thus finding fair use.¹⁰⁵

In *Perfect 10*, the court also addressed embedding by creating the heavily-relied-upon "server test," which states that embedding does not violate the copyright holder's right to public display unless the copyrighted material was stored on the server of the embedding website.¹⁰⁶ Thus, in order to show infringement of their public display right under the server test, the plaintiff would have to show that server storage embedding took place (akin to making a copy of Mona Lisa before live streaming it).¹⁰⁷ The server test has also been embraced by the Seventh Circuit in *Flava Works vs. Gunter*,¹⁰⁸ and was largely accepted as the law of the land in the United States until 2017.¹⁰⁹

2. *District Courts and the Incorporation Test*

Since the ruling in *Perfect 10* came down in 2007, most of the internet was built around the server test.¹¹⁰ However, recently district courts in both the Fifth and Second Circuits rejected the server test to some degree. In 2017, the Northern District of Texas rejected the server test in *Leader's Institute, LLC v.*

103. *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 818 (9th Cir. 2003).

104. *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1163 (9th Cir. 2007).

105. *Id.* at 1165.

106. *Id.* at 1159.

107. *Id.*

108. *Flava Works, Inc. v. Gunter*, 689 F.3d 754, 756 (7th Cir. 2012).

109. Jeffrey Neuburger, *New York Court Rebuffs Ninth Circuit's Copyright "Server Test." Finds Embedded Tweet Displaying Copyrighted Image to Be Infringement*, PROSKAUER (Mar. 2, 2018), <https://newmedialaw.proskauer.com/2018/03/02/new-york-court-rebuffs-ninth-circuits-copyright-server-test-finds-embedded-tweet-displaying-copyrighted-image-to-be-infringement/>.

110. For example, YouTube allows the video creators that post to its website to prevent their videos from being embedded. *Restriction embedding*, YOUTUBE, <https://support.google.com/youtube/answer/6301625?hl=en> (last visited Feb. 9, 2020).

Jackson by distinguishing *Perfect 10* and rejecting the idea that a copy of copyrighted material is required for public display.¹¹¹

The Southern District of New York likewise rejected the server test in favor of the so-called “incorporation test.”¹¹² In *Goldman v. Breitbart*, Goldman posted a photo of Tom Brady to his Snapchat.¹¹³ One of Goldman’s friends then posted the photo online, and it went viral as Twitter users began reposting it.¹¹⁴ The tweets which included the photo were picked up by some news organizations and embedded in their online stories.¹¹⁵ Goldman sued nine of those news organizations claiming copyright infringement.¹¹⁶

The court ruled in favor of the photographer, stating that “liability should not hinge on invisible, technical processes imperceptible to the viewer.”¹¹⁷ The court did not address the fact that embedding creates a deep link to the source website, which is clear to the user looking for the source. The “incorporation test” would find infringement anytime a website incorporates or embeds the copyrighted material without a license from the copyright owner. This broad sweeping ruling can “radically change linking practices, and thereby transform the Internet as we know it.”¹¹⁸ The Second Circuit thus far has denied an interlocutory appeal.¹¹⁹ At the time of writing it is unclear whether the issue will be taken up on appeal.¹²⁰

The server and incorporation tests are radically different and clearly in juxtaposition to each other. The Southern District of New York’s approach protects copyright holders’ right of exclusive ownership. It also renders

111. *Leader’s Inst., LLC v. Jackson*, No. 3:14-CV-3572-B, 2017 WL 5629514 (N.D. Tex. Nov. 22, 2017).

112. Eric Goldman, *In-line linking May Be Copyright infringement - Goldman v. Breitbart News*, TECH & MKT. L. BLOG (Feb. 16, 2018), <https://blog.ericgoldman.org/archives/2018/02/in-line-linking-may-be-copyright-infringement-goldman-v-breitbart-news.htm> [<https://perma.cc/Y25K-2Y64>].

113. *Goldman v. Breitbart News Network, LLC*, 302 F. Supp. 3d 585, 586–87 (S.D.N.Y. 2018).

114. *Id.*

115. *Id.*

116. *Id.*

117. *Id.* at 595.

118. Brief for the Electronic Frontier Foundation and Public Knowledge as Amicus Curiae Supporting Defendants at 12, *Goldman v. Breitbart*, 302 F.Supp.3d 585 (S.D.N.Y. 2018) (No. 17-CV-3114).

119. Adam R. Bialek, *SCOTUS Showdown Will Have to Wait as Second Circuit Denies Petition to Review SDNY Rejection of Server Test for Copyright Infringement*, NATIONAL LAW REVIEW (July 19, 2018), <https://www.natlawreview.com/article/scotus-showdown-will-have-to-wait-second-circuit-denies-petition-to-review-sdny>.

120. Leila Knox, *Second Circuit Punts on In-Line Linking Appeal*, JDSUPRA (Aug. 15, 2018), <https://www.jdsupra.com/legalnews/second-circuit-punts-on-in-line-linking-39813/>.

embedding without a license illegal on the internet, which will result in dramatic changes to the web's structure.¹²¹ In contrast, the Ninth Circuit's approach allows the flexibility necessary to preserve embedding on the internet. The server test also accommodates settled expectations both of consumers and companies whose businesses rely on embedding to quickly display all necessary information.

C. LINKING AND EMBEDDING IN EUROPE

Although the United States and the European Union try to fulfil different goals through their intellectual property protections,¹²² Europe has moved in a very similar direction to the recent district court decisions in the Second and Fifth Circuits. The newly passed Directive on Copyright in the Digital Single Market attempts to enlarge and strengthen the rights of copyright holders to the detriment of linking on the web.

1. *Directive (EU) 2019/790 Background*

On April 17, 2019, the European Parliament passed the Directive 2019/790 (the "Directive"), adopting stringent copyright protections that threaten internet linking.¹²³ The new law contains two articles which extend copyright protection for embedded works.¹²⁴ According to the European Council, the goal of the Directive is to broaden previous copyright protections by reducing the "value gap" between profits for copyright holders and internet platforms.¹²⁵

During the legislative process, the Directive has spurred much debate. The Directive's opponents included major tech companies, internet users, and human rights advocates, while the proponents included media organizations, newspapers and publishers.¹²⁶ The parties opposing the Directive were concerned with its potential to inhibit freedom of speech on the internet, and the possibility it turns large tech companies into the copyright police.¹²⁷ In an open letter to the European Committee in 2017, the opponents of the

121. *See supra* Section II.B.

122. *Id.*

123. Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 Relating to Copyright on the Digital Single Market, 2019 O.J. (L 130) 92.

124. *See supra* Section IV.C.2.

125. Council of the European Union Press Release, Copyright Rules for the Digital Environment (May 25, 2018).

126. *See* Letter from David Kaye, Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (June 13, 2018); *Protests greet Brussels copyright reform plan*, BBC NEWS (June 15, 2018), <https://www.bbc.com/news/technology-44482381#>; *Letter by 9 News Agencies* (Dec. 2017), <https://perma.cc/8T9N-X7PK>.

127. Palais des Nations, *supra* note 126.

Directive warned of the dangers that this new law could bring.¹²⁸ They were concerned about the creation of uncertainty that would incentivize the monitoring roles of ISPs and the potential communication blocking that ISPs would have to do, “if they are to have any chance of staying in business.”¹²⁹ At the same time, the Directive’s proponents argued it decreases the difference in earnings between copyright holders and ISPs.¹³⁰

The Directive came into force on June 7, 2019, and member states have until June 7, 2021, to adopt it within their borders.¹³¹ The actual effects of the Directive on the member states are yet unknown, but it is likely to turn European courts into a hotbed of linking- and human rights-related litigation.¹³²

2. *Articles 15 and 17*

Article 15 (Draft Article 11) gives print publishers a year-long direct copyright over “online use of their press publication by information society

128. The opponents included Human Rights Watch, Reporters Without Borders, Creative Commons, Electronic Frontier Foundation, Max Planck society, and many more. *See* Mallory Locklear, *Digital rights groups speak out against EU plan to scan online content*, ENGADGET (Oct. 17, 2017), <https://www.engadget.com/2017/10/17/digital-rights-groups-against-eu-scan-online-content/> (noting that mandates similar to Article 13 have been twice rejected by the Court of Justice before); *see also* Eileen Hershenov, *How the EU copyright proposal will hurt the web and Wikipedia*, WIKIMEDIA FOUNDATION (June 29, 2018), <https://blog.wikimedia.org/2018/06/29/eu-copyright-proposal-will-hurt-web-wikipedia/> (discussing how “[the] proposed new copyright package in the European Union is a threat to our fundamental right to freely share information,” and outlining the specific effects it might have on Wikipedia); Katie Collins, *Article 13: Europe’s hotly debated revamp of copyright law, explained*, CNET (Mar. 25, 2019), <https://www.cnet.com/news/article-13-europes-hotly-debated-eu-copyright-law-explained/> (describing how great the potential impact of the directive can be “[a]n organized campaign against Article 13 warns that it’d affect everything from memes to code, remixes to livestreaming”).

129. Open Letter from Civil Liberties Union for Europe and European Digital Rights et. al., to President Juncker et. al. (“Article 13 of the proposal on Copyright in the Digital Single Market include obligation on internet companies that would be impossible to respect without the imposition of excessive restrictions on citizens’ fundamental rights”).

130. *See* Council of the European Union Press Release, *supra* note 125.

131. Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 Relating to Copyright on the Digital Single Market, 2019 O.J. (L 130) 92, <https://perma.cc/9HAJ-RZP7>.

132. *See* Dawn C. Chmielewski, *YouTube CEO Susan Wojcicki Warns Against “Unintended Consequences” Of EU’s Article 13*, DEADLINE (Nov. 12, 2018), <https://deadline.com/2018/11/youtube-ceo-susan-wojcicki-warns-unintended-consequences-article-13-1202500444/>; Sam Forsdick, *European Parliament votes against ‘publisher’s right’ copyright law changes as Facebook warns of ‘unintended consequences,’* PRESSGAZETTE (July 5, 2018), <https://www.pressgazette.co.uk/european-parliament-votes-against-publishers-right-copyright-law-changes-as-facebook-warns-of-unintended-consequences/>.

service providers.”¹³³ Before the Directive went into effect, publishers in Europe had to rely on authors to assign them the copyright before they could invoke copyright protections.¹³⁴ In addition to limiting the protection to only last one year, Article 15 also includes exemptions for copying an “insubstantial” part of a work and copying required for scientific or academic research.¹³⁵ Both exceptions remain rather poorly defined and ambiguous.

Article 17 (Draft Article 13) requires for-profit “online content sharing service providers” to implement “effective and proportionate measures” to prevent the availability of unlicensed copyrighted material.¹³⁶ What constitutes “effective and proportionate measures”?¹³⁷ The article mentions removing unlicensed works “expeditiously” and demonstrating that the website provider made “best efforts” to prevent the future availability of such works.¹³⁸ Article 17 carves out an exception for private cloud storage services, non-profit online encyclopedias, and non-profit educational or scientific repositories.¹³⁹ But large parts of how this Article will be applied remain ambiguous similarly to the exceptions in Article 15.¹⁴⁰

As both Europe and the United States seem to move towards more stringent copyright protections for embedding on the internet, the following sections discuss potential ramifications of both strengthening and loosening those copyright protections.

V. ANALYSIS OF INTERPLAY BETWEEN COPYRIGHT LAW AND LINKING ON THE INTERNET

As shown above, at least some of the goals of copyright law and the internet seem to be different. The former prioritizes exclusive rights, while the latter promotes freedom to access content without restrictions. This Part focuses on evaluating the potential consequences of favoring either of these goals and shows that it is possible to reconcile them.

133. Council of European Union Press Release, *supra* note 125.

134. Policy Department for Citizens Rights and Constitutional Affairs, *Strengthening the Position of Press Publishers and Authors and Performers in the Copyright Directive*, at 15 (Sep. 15, 2017), [https://www.europarl.europa.eu/RegData/etudes/STUD/2017/596810/IPOL_STU\(2017\)596810_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2017/596810/IPOL_STU(2017)596810_EN.pdf).

135. *Copyright Rules for the Digital Environment*, *supra* note 125.

136. Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 Relating to Copyright on the Digital Single Market, 2019 O.J. (L 130) 92, <https://perma.cc/9HAJ-RZP7>.

137. *Id.*

138. *Id.*

139. *Id.*

140. *See supra* Section IV.C.2 (discussing Draft Article 11).

A. CONSEQUENCES OF STRINGENT COPYRIGHT PROTECTION IN LINKING ON THE INTERNET

If the United States chooses to embrace the emerging incorporation test rather than the Ninth Circuit's server test, the consequences for the internet are likely to be dire. Such a decision will essentially eliminate the ability to embed on the internet because it will create copyright infringement liability for embedding. This Section discusses how the likely positive impacts of such protections are significantly outweighed by their negative impacts.

1. *Positive Effects of Stringent Copyright Protections in Linking on the Internet*

The positive aspects of establishing stronger copyright protections on the internet include: (1) creating a clear vehicle for reimbursement of authors' creative efforts;¹⁴¹ (2) empowering authors to create a culture on the internet that respects their authorship rights;¹⁴² and (3) incentivizing continuous content creation by the authors.¹⁴³

Although the first two effects seem generally desirable, they do not fulfil the fundamental goals of intellectual property protections in the United States and should only be a secondary consideration.¹⁴⁴ Intellectual property protections such as copyright are meant to incentivize "progress in science and useful arts."¹⁴⁵ On the internet this can be achieved in two ways. First, internet can promote progress by encouraging the creation of more new content, focusing on the output quantity. Second, the internet can incentivize further creation by providing access to information that will spur the creation of new quality content (independent of quantity).

Both of these incentives can be encouraged on the internet without the strong enforcement of copyright holders' rights. The idea that the exclusive rights, granted by copyright protections, incentivize authors to create comes from the non-digital world, where each physical copy of individual pieces of work could have been sold for monetary gain.

This assumption, however, no longer holds true on the internet. Amongst other reasons, it's likely impossible to keep track of all the copies of one's own work to the same degree as in the non-digital world due to its sheer size and

141. SFWA, <https://www.sfwaw.org/other-resources/for-authors/writer-beware/copyright/> (last visited May 8, 2020).

142. Ginsburg, *supra* note 77, at 477.

143. James Campbell, *Authorship, Incentives for Creation, and Copyright in the Digital 21st Century*, 1, 6–7, <https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/meet.1450430168>.

144. *See supra* Section II.B.

145. U.S. CONST. art. I § 8, cl. 8.

openness. And the policing of one's copyrightable material can become extremely difficult and expensive.¹⁴⁶ Thus, although stringent copyright protections could protect copyright holder's rights to a certain degree, they are likely to only favor those copyright holders who have the means to police their own work. Furthermore, in the current copyright system, if a copyright holder is successful in policing their copyrighted material, they are still not economically rewarded unless they choose to pursue litigation.

Although economic incentives should not be the primary concern of copyright law, they are still critical to drive creation of quality art on the internet. Access to information, after all, is meaningless if no new works will continue to be created due to lack of funds. Publication houses, newspapers, film studios, and many others require an ability to collect money to keep their lights on, in order to continually produce new "sciences and useful arts."¹⁴⁷ While fully acknowledging the importance of economic compensation for the work that authors and copyright holders often put into their works, this Note suggests that enforcement of copyright should not be the way we achieve such compensation. Instead, copyright holders could rely on licensing schemes, similar to the ones seen in the music industry, to obtain monetary compensation for their works.¹⁴⁸

2. *Negative Effects of Stringent Copyright Protections in Linking on the Internet*

Stronger online copyright protections have few beneficial effects, but their negative consequences would be disastrous for many industries. Strengthening copyright protections on the internet would not only upset settled industry expectations and prioritize the needs of copyright holders over the public, it would largely favor established artists and ISP giants, and in the worst-case scenario create absolute barriers to access of information.

a) *Upsetting Settled Expectations*

Proponents of more stringent copyright protections worry that embedded links will drive down content creation by preventing copyright holders from

146. Once information is released on the internet it is difficult to control who has access to it (unless it's behind a paywall). If anyone can gain access to information anyone can post it anywhere on the web—monitoring such activity can be extremely time consuming. Some services help copyright holders with such tasks, but the user has to actively go search for copyright infringements. Furthermore, legal enforcement of copyright can also be expensive. See Yang Sun, *Copyright Law Enforcement in Online Environment*, (2015), <https://ssrn.com/abstract=2432987>.

147. *Id.*

148. See *infra* Section V.C.

being fully compensated for their work. But embedding has been a largely accepted practice since the decision in *Perfect 10* over a decade ago.¹⁴⁹ And content creation on the internet has continued to thrive.¹⁵⁰ Therefore, embedding practices on the internet have not stalled progress, but likely further facilitate it by giving easy access to inspiration for the end users. This trend shows that copyright protections for embedding are not necessarily needed to incentivize content creation on the internet, but just the opposite. By making material more difficult to find, they may act as a hurdle in the creative process.

b) Shifting the Incentive-Access Balance

To many, it might seem strange that authors would be incentivized to create content even without strong copyright protections. But the internet has exponentially expanded the potential economic markets available to copyrightable material, yet copyright law has not addressed this change.¹⁵¹ Intellectual property protections assume a balance between incentives created through exclusive rights for authors and access to creative works that will spur further creation of content (the “incentive-access balance”).¹⁵²

The internet profoundly shifted that balance by expanding the economic possibilities for copyrighted materials and at the same time creating an easier access to information for countless individuals.¹⁵³ By exponentially increasing the audience for works, the internet allows copyright holders and authors to make less compensation per each view of their work, while also exposing them to the benefits of new, previously unreachable, audiences.

Stringent copyright protections in linking will limit access to information, as copyrightable material will be more difficult to share. The restriction of embedding on the internet might seem insignificant due to the availability of deep and simple linking. But would a user view an embedded image, tweet or video if it required them to navigate to another website? The more information is buried, the less likely it is to reach the eyes of the end user. Thus, decisions to strengthen copyright holders’ exclusive rights often ignore the access side

149. See *supra* Section IV.B.1.

150. See Alice Klat & Jayant Bhargave, *Content democratization & How the Internet is fueling the growth of creative economies*, PWC GLOBAL (Jan. 5, 2017), <https://www.strategyand.pwc.com/gx/en/insights/content-democratization.html>.

151. COMM. ON INNOVATIONS IN COMPUTING AND COMM’N, FUNDING A REVOLUTION: GOVERNMENT SUPPORT FOR COMPUTING RESEARCH 180–81 (1999).

152. See *Twentieth Century Music Corp. v. Aiken*, 422 U.S. 151, 156 (1975) (“Creative work is to be encouraged and rewarded, but private motivation must ultimately serve the cause of promoting broad public availability of literature, music, and the other arts.”).

153. See COMM. ON INNOVATIONS IN COMPUTING AND COMM’N, *supra* note 151, at 169, 180.

of the balance that should be struck in intellectual property protections. This in turn diminishes the access to arts and sciences that is at the very heart of intellectual property law. Enforcing stringent copyright protections in embedding will once again push that balance in favor of the copyright holder, without accounting for the need to access to information, which is so crucial to stimulating future creativity.

Limiting embedding can also create dire consequences for authors. As mentioned above,¹⁵⁴ image source embedding as used by Google Images also serves as a deep link to the source website.¹⁵⁵ Without the availability of embedding, which at least provides some level of credit to the source website, users are more likely to turn to illegal downloads and uploads,¹⁵⁶ which would be harder for authors to track. And since the culture on the internet has largely ignored copyright holders' exclusive rights and instead facilitated free use of copyrightable material,¹⁵⁷ the average internet user is unlikely to be stopped by copyright protections. This is especially true in terms of images available on the internet, as they do not have the equivalent of Netflix for videos or Spotify for music, to which users can turn. In terms of copyrightable images, restricting embedding is equivalent to the removal of streaming services for music and would likely push us back to the time of Napster.¹⁵⁸

c) Working Against the “Little Guy”

The impact of banning embedding on the internet would not be limited to large companies that heavily rely on embedding practices, or authors that loathe them. Small businesses are also likely to be impacted due to the

154. See *supra* Part III.

155. If a user clicks on the image after expanding it, although the image is embedded, it takes the user to the website from which the image originates.

156. The internet from its conception has been used and often viewed as a tool for exercise of the freedom of speech, and granting access to fundamental rights, this created a culture of entitlement of access that users often rely on to justify to illegally download music or movies. See SULA WONG, EITAL ALTMAN & JULIA ROJAS-MORA, INTERNET ACCESS: WHERE LAW, ECONOMY, CULTURE AND TECHNOLOGY MEET 477 (vol. 55 issue 2 2011). The individuals are therefore likely to continue using copyrighted material for as long as they can get away from it.

157. See Kit Walsh, *Copyright Law Versus Internet Culture*, EFF (Feb. 24, 2017), <https://www.eff.org/deeplinks/2017/02/copyright-law-versus-internet-culture> (describing how internet culture depends on remixing copyrighted content via “fan fiction,” “Photoshop battles,” “fan music videos,” “redubbing and resubbing,” “reaction videos,” and more).

158. See Alex Suskind, *15 Years After Napster: How the Music Service Changed the Industry*, DAILY BEAST (July 12, 2017), <https://www.thedailybeast.com/15-years-after-napster-how-the-music-service-changed-the-industry>.

favoritism that already exists.¹⁵⁹ Smaller businesses or start-ups are far less likely to have legal teams or funds necessary to protect themselves from copyright infringement litigation.¹⁶⁰ Thus, stringent copyright protections around embedding would drive small businesses to avoid practices that could land them in court. That would ultimately make their large competitors more attractive to end users, because they would offer a wider range of services. In turn, the inability to use certain technologies by small businesses has the potential to stifle business growth and further propagate the monopolies held by large ISPs.

When discussing the need for more stringent copyright protections in embedding practices, many emphasize the need for small authors to “make a living.”¹⁶¹ But how is a small author who cannot make a living supposed to hire a decent lawyer to enforce their copyrights?¹⁶² Putting aside the fact that a lot of small authors are not themselves copyright holders and instead assign their rights to a third-party, the costs of such a lawsuit would often be prohibitive to any small entity. This is especially so when suing an ISP giant like Google, who is likely willing to throw money at the problem until it goes away.¹⁶³ This not only makes it essentially impossible for a small copyright holder without proper funding to succeed in such a suit, it further creates a discrepancy in the use of embedding between the internet giants and their smaller competitors while essentially providing a remedy only for the rich.

Fair use and DMCA § 512 takedown notices are not enough to address the gap. The availability of the fair use defense typically hinges on the noncommercial use of the copyrightable material.¹⁶⁴ Since a lot of websites posting copyrightable materials receive money through advertising, they are unlikely to fall within the protection of fair use due to their commercial

159. Jonathan Bailey, *5 Ways Copyright is Screwing Smaller Creators*, PLAGIARISMTODAY (Oct. 20, 2010), <https://www.plagiarismtoday.com/2010/10/20/5-ways-copyright-is-screwing-smaller-creators/>.

160. C. Daniel Baker, *Many Small Businesses Don't Seek Legal Help Despite Risks*, BLACKENTERPRISE (Oct. 25, 2013), <https://www.blackenterprise.com/small-businesses-need-legal-help/>.

161. See *Oversight of the U.S. Copyright Office: Hearing on H.R. 113-116 before the Subcomm. On Courts, Intellectual Property, and the Internet*, 113th Congress 2 (2014) (statement of Mary Rasenberger on Behalf of the Authors Guild).

162. Cost of litigating such a case can go well into six figures. See *How much does it cost to pursue a copyright infringement claim*, TRAVERSE LEGAL, <https://www.traverselegal.com/blog/how-much-does-it-cost-to-pursue-a-copyright-infringement-claim/> (last visited Feb. 8, 2020).

163. Some of those cases might settled in favor of the author. However, settlements for many of these companies can create a dangerous precedent, and thus the companies are likely to want to avoid them.

164. 17 U.S.C. § 107.

nature.¹⁶⁵ The DMCA § 512 takedown notice is also a problematic remedy for this problem, because although it would protect copyright holders' exclusive rights, it disregards the access side of the incentive-access balance equation by preventing end users from having any access to the copyrightable content.

If the roles are reversed, however, and it is the "little guy" who infringes a copyright of a large copyright holder, the copyright holder will have the resources under the current regime to enforce their exclusive rights. This once gain favors the larger party against the "little guy," both in terms of small copyright holders and small website providers, as with the DMCA § 512 takedowns discussed above. Strong copyright holders' rights, particularly in a regime which favors large parties, can so heavily distort the incentive-access balance that it can entirely cut end users off from access to content.

d) Creation of Walled Gardens

As shown above, more stringent copyright protections would further skew the delicate incentive-access balance that has already been heavily tilted in favor of the copyright holders. The most dangerous and extreme consequence of stringent copyright protections is the potential for so-called "walled gardens"—a term used broadly in the tech world to signify a system where the service provider has the ability to limit access to content, applications, and other platforms.¹⁶⁶ The slippery slope of prohibiting one form of linking creates a precedent of willingness to compromise our access to information. This, in turn, would disrupt the functioning of the web¹⁶⁷ and potentially infringe on basic human rights such as rights of freedom of expression and opinion.¹⁶⁸

3. *The Effects of the European Directive*

The policies favoring stringent copyright enforcement on the internet, such as those established in Europe through Directive (EU) 2019/790,¹⁶⁹ would largely have the same effects as those established in the United States

165. See 17 U.S.C. § 107 (1).

166. *Escaping the Walled Gardens in the Clouds*, *supra* note 3.

167. Imagine that every time you tried to access any copyrightable material (a picture, a music video, an article) you'd have to go through paywalls—the cost for simple research on internet would become prohibitive to an average user. In scholarly research many of the websites rely on such paywalls (for example, JSTOR). Students and faculty therefore rely on their Universities' subscriptions in order to utilize such services. In the face of too many restrictions on the access to information, the internet would stop being the great equalizer and instead become another elitist institution.

168. See *supra* Section II.A and Part IV.

169. See *supra* Section IV.C.

through recent court rulings.¹⁷⁰ And the Directive is not without its critics—it has already been challenged. Poland has filed an action for annulment with the Court of Justice of the European Union, alleging that such laws would bring “unwanted censorship” on the internet.¹⁷¹

Of special concern to Poland is Article 17, which is likely to have a “chilling effect”¹⁷² on online expression. Article 17 encourages even the smallest web services “to do everything they can to stop users [from] uploading copyrighted content without authorization,” if the ISPs want to avoid direct liability for infringement.¹⁷³ If this form of censorship was not daunting enough, the type of filtering technology necessary to accommodate Article 17 is extremely expensive if not currently technologically infeasible.¹⁷⁴ These costs are likely to be prohibitive to smaller ISPs, thus once again stifling competition and promoting monopolistic practices.

The consequences of Article 15, which grants print publishers a year-long direct copyright over materials they publish, are far less theoretical than those of Article 17. The ideas behind Article 15 have been tested by both Germany and Spain, where referrals to major newspapers plummeted when Google simply withdrew embedding of certain news sources that wanted to enforce the “link tax.”¹⁷⁵ When major newspapers in Germany demanded payment for their copyrighted content, Google simply ceased to use their content in Google News, causing dramatic drops in readership of source websites.¹⁷⁶

One way to address the imbalance in the relationships between the ISPs and news providers would be for the world to uniformly adopt the “link tax” that was passed in the European Union. This would force the ISPs to comply

170. *See supra* Section IV.C.2.

171. *See* Andrew Liptak, *Poland has filed a complaint against the European Union’s copyright directive*, THE VERGE (May 25, 2019), <https://www.theverge.com/2019/5/25/18639963/poland-european-union-copyright-directive-filed-complaint-court-of-justice> (Articles 15 and 17 of the Directive “would be ripe for abuse by copyright trolls and would make millions of mistakes. The technology simply does not exist to scan the internet’s content in this way.”).

172. *See* David Meyer, *Tech Industry and Activists Still Hope to Sink New EU Copyright Rules*, FORTUNE (Feb. 14, 2019), <https://fortune.com/2019/02/14/eu-copyright-directive-trilogue-deal/>.

173. *Id.*

174. *See generally* Felipe Romero-Moreno, ‘Notice and staydown’ and social media: amending Article 13 of the Proposed Directive on Copyright, 33:2 INT’L REV. L. COMPUTS. & TECH. 187, 189 (2019).

175. *See* David Meyer, *EU Lawmakers Are Still Considering This Failed Copyright Idea*, FORTUNE (Mar. 24, 2016), <https://fortune.com/2016/03/24/eu-ancillary-copyright/> (“Referrals to Springer properties such as *Bild* plummeted by as much as 80% and the publishers quickly retreated, granting Google a temporary exemption from having to pay them.”).

176. *Id.*

and work with copyright holders to find a mutually beneficial arrangement. Such a resolution, however, is both unlikely to happen and undesirable because it would further skew the copyright incentive-access balance. It would likely lead to reduced access to copyrightable material on the internet, as it provides copyright holders with all the power, and does not consider the rights of end users to access information at all.

On the other hand, if no one outside of the European Union adopts the link tax, the ISP giants will remain largely unaffected by copyright limitations. Only the companies of countries with stringent protections are likely to suffer economic downturn when their services are underutilized. As the German example shows, newspaper organizations and publishers, which often are at the forefront of advocating for stringent copyright protections, are likely to suffer if they try to enforce their rights while the laws are not uniform across the internet.¹⁷⁷ Furthermore, according to experts, such a link tax is also likely to cause a shutdown of small media companies across Europe and turn the internet into a “tool of surveillance.”¹⁷⁸

B. CONSEQUENCES OF NO COPYRIGHT IN EMBEDDING ON THE INTERNET

Understanding the consequences of not providing copyright protections is also key to appreciating the licensing solution proposed in this Note. The internet was dreamt up as a tool for free access to information.¹⁷⁹ The logical question to ask, then, is what would happen if no or limited copyright protection existed for embedded links on the internet. Would content creation cease entirely? Maybe, but it is very unlikely. Would the potential loss of economic incentives affect the amount and type of content created? The answer to this question is far more likely to be yes.

177. See *EU Lawmakers Are Still Considering This Failed Copyright Idea*, *supra* note 175.

178. Some of the brightest minds of technology and the pioneers of the internet such as Mitch Kapor, Tim Berners-Lee, Vint Cerf, Guido van Rossum, James Cornin, and Desiree Miloshevic (amongst many others) warn that “Article 13 takes an unprecedented step towards the transformation of the Internet, from an open platform for sharing and innovation, into a tool for the automated surveillance and control of its users.” See Danny O’Brien & Jeremy Malcolm, *70+ Internet Luminaries Ring the Alarm on EU Copyright Filtering Proposal*, EFF (June 12, 2018), <https://www.eff.org/deeplinks/2018/06/internet-luminaries-ring-alarm-eu-copyright-filtering-proposal>.

179. See *supra* Section II.A; see also Alex Hern, *Tim Berners-Lee on 30 years of the world wide web: We can get the web we want*, THE GUARDIAN (Mar. 12, 2019), <https://www.theguardian.com/technology/2019/mar/12/tim-berners-lee-on-30-years-of-the-web-if-we-dream-a-little-we-can-get-the-web-we-want> (discussing the initial intent for the internet to be accessible, interoperable and open).

The internet has permeated our society and lives to such degrees it is almost impossible to avoid it. Copyright holders could try to stop the linking and embedding of their works on the internet, however, the type of exposure that an author can get through others reposting their work is unmatched virtually anywhere else.¹⁸⁰ Exposure is desirable, because it has the potential to generate additional revenue for the copyright holders.¹⁸¹

It is this motivation for publicity (for shares and likes) that makes embedding such a powerful tool. Although theoretically the same level of exposure could be achieved through simple or deep linking, the convenience of embedding is more likely to draw the attention of end users.¹⁸² Additionally, modern embedding functions as a deep link to the source website. In the absence of embedding, internet users who are often unaware of copyright restrictions are more likely to download the copyrighted material and upload it back without giving any credit to the creator.

The issue with embedding in the current scheme, however, is that it is can be difficult for copyright holders to monetize all of their exposure. Without the users clicking on embedded content, copyright holders cannot collect the advertisement revenue they often rely on.¹⁸³ Unlike the current scheme, the licensing system proposed in the next Section ensures that copyright holders will get reimbursed for their embedded works.

Finally, limiting copyright protections would help fulfill the basic human right of freedom of opinion and expression, which includes the freedom to information and ideas.¹⁸⁴ Access to the internet is crucial to fulfilling one of the most basic human rights. In 2016, the U.N. Human Rights Council released a nonbinding resolution emphasizing the importance of access to the internet by all members of society and condemning “measures to intentionally prevent or disrupt access to or dissemination of information online.”¹⁸⁵

180. See, e.g., Arjun Kharpal, *Pop stars the Internet made famous*, CNBC (Apr. 23, 2015), <https://www.cnbc.com/2015/04/21/popstars-the-internet-made-famous.html> (showing pop stars that became famous thanks to the internet).

181. *Id.*

182. See Dan Taylor, *The Advantages & Disadvantages of Object Linking & Embedding*, TECHWALLA, <https://www.techwalla.com/articles/the-advantages-disadvantages-of-object-linking-embedding> (last visited Aug 8 2021).

183. See *Kelly v. Arriba Soft Corp.*, *supra* note 39. Although some will likely click on the embedded materials, there is no way for copyright holders to assure that every user that sees their material embedded on another website, chooses to click on the embedded material, and thus gets redirected to the source website.

184. See G.A. Res. 217 (III) A, Universal Declaration of Human Rights (Dec. 10, 1948).

185. U.N. General Assembly, *The promotion protections and enjoyment of human rights on the Internet Statement*, United Nations General Assembly 1, 4 (June 27, 2016), https://www.article19.org/data/files/Internet_Statement_Adopted.pdf.

Although the resolution itself was initiated to address issues of silencing dissidents by cutting off access to the internet, the identified principle can also be applied to copyright law, if such is enforced to the extreme.¹⁸⁶ The removal of copyright protection on the internet is likely to further promote the equal access to information by all members of society, by preventing the creation of “walled gardens.”¹⁸⁷

That same system, however, would to some degree undermine the compensation model that currently exists for content on the internet. And it is this economic aspect of reimbursing copyright holders that is the most problematic in a world of no copyright protections. Although the equal access to information and clearly set expectations have the potential to equalize (at least to the greatest degree possible) the economic playing field both for small and large ISPs, lack of copyright protections could prevent some of those organizations from monetizing their content. Businesses would still be able to monetize on the embedded material through the publicity that embedding provides, and ad revenue.¹⁸⁸ These two sources of revenue are significantly smaller than the ones currently available on the internet.

The removal of copyright protections for linking practices on the internet affects the production of content only to a certain degree, while also promoting basic human rights. Although this Note does not advocate for a complete removal of copyright protections for linking on the internet, understanding the extreme views on copyright protections on the internet allows the reader to fully appreciate the balance struck in the proposed licensing resolution. The goal of that licensing system is to return to the original intent behind incentivizing progress by giving authors and copyright holders a certain level of protection, while keeping in mind the public need for access to their works. The remaining sections, therefore, attempt to reconcile the potential consequences of copyright protections and the goals of the internet while addressing existing law.

C. PROPOSED LICENSING REGIME RECONCILING THE GOALS OF THE INTERNET AND COPYRIGHT LAW AND ADDRESSING THE INCENTIVE-ACCESS BALANCE

As this Note has shown, neither the European nor American systems properly balance the access and incentive goals of copyright. As both the

186. *See id.*

187. *See supra* Section V.A.2.d).

188. For example, YouTube allows a content poster to profit from ads in the video even if the video is not viewed in the platform. For more information, see *How Youtube Ad Revenue Works*, INVESTOPEDIA, <https://www.investopedia.com/articles/personal-finance/032615/how-youtube-ad-revenue-works.asp> (last visited Mar. 31, 2020).

European Union and United States continue on the trajectory of increasing copyright holders' exclusive rights in linking on the internet, we have to make a choice between reconciling the goals of the internet and copyright or continuing to put them at odds with each other. A licensing system akin to the one outlined below could strike the right balance.

The ultimate goal of copyright law in the United States is to drive progress in society through content creation and access to art and information,¹⁸⁹ while the goal of the internet is to provide free and open access to that very content.¹⁹⁰ These goals can be reconciled by removing the assumption that the only way to incentivize content creation and economic reimbursement to copyright holders is through copyright enforcement. The pervasiveness of embedding on the internet and the continual production of content, show that there are other factors that continue to drive authors to produce on the internet. However, it is important to account for the economic needs that will continue to drive the production of content.¹⁹¹

Instead of relying on threat of copyright litigation, the internet could follow the lead of the music industry.¹⁹² Congress could create a licensing scheme that allows businesses to pay a certain range of prices, dependent on their size, to embed copyrighted material. Such a scheme should prohibit the collection of monetary reimbursement from noncommercial website holders that do not generate a profit from the website they are running and create caps for small businesses and startups that would allow them to compete with their wealthier counterparts. This would prevent copyright from stifling small businesses and non-profit online ventures and allow for true, noncommercial hobby websites to continue to thrive.

In order for such a licensing scheme to be successful, the government would have to set up a non-profit to facilitate licenses between online service organizations and copyright holders.¹⁹³ The copyright holders would then be reimbursed based on the licensing fees paid to the non-profit. This would protect small artists, while also driving the price of licensing down for big business, because of their ability to bargain for a price to access all copyrighted material. Large ISPs also would not have to worry about negotiating separate agreements with anyone who puts copyrighted material on the internet. This

189. *See supra* Section II.B.

190. *See supra* Section II.A.

191. *See supra* Section V.A.1.

192. *See* Kevin R. Fish, *Restaurants, Music Licensing, and the United States Copyright Code*, 5:1 J. FOOD SERV. BUS. RES. 119, 121–25 (2002) (discussing the music licensing industry's facilitation of licenses).

193. Such an organization could mirror BMI, SCAP, or SESAC from the music licensing industry. *Id.* at 123.

type of approach would make it easier for big businesses to buy into a licensing scheme.

On its surface, the licensing system may seem similar to the resolution suggested in recently passed European Directives. There is, however, one key distinction between the licensing system proposed in this Note and the European Resolution. The licensing model provides a balance between incentives for copyright holders and the need to access copyrighted information by end users, which has long been ignored in copyright regulation.¹⁹⁴

The emphasis of the proposed licensing system would focus on making sure that materials are widely accessible and that copyright holders get reimbursed for their work. The licensing system would not, however, permit the copyright holders to cherry-pick where their work is available, or to serve takedown notices to remove their works whenever they wish to withhold them from the public.

The proposed licensing system would also work better for the European Union in spite of the fact that it emphasizes authors' exclusive rights.¹⁹⁵ The question of reconciling the goals of copyright and the internet in the European Union is much more complicated because of that emphasis.¹⁹⁶ It will be virtually impossible to find a solution that accommodates both the need for author control, and the free and open access to information on the internet.

The European Union, therefore, finds itself at critical crossroads. If it continues to favor copyright protections over openness of the internet, Europe will likely stifle the enormous economic potential of the web. At the same time, Europe is pushing away large, established ISPs by creating a form of liability that does not exist for those ISP's anywhere else. Europe could, however, build in moral rights (but not absolute rights of control) into the outlined licensing system. This would allow the copyright holders to receive reimbursement and make sure they are able to exercise a certain level of control, while still providing its citizens the same level of access that they have grown accustomed to.

194. See PATRICIA AUFDERHEIDE & PETER JASZI, RECLAIMING FAIR USE 71 (2d ed. 2011) (arguing that access has lost out to exclusive rights of copyright holders and that this should change).

195. See PETER BALDWIN, THE COPYRIGHT WARS: THREE CENTURIES OF TRANS-ATLANTIC BATTLE 15–16 (2014).

196. See *supra* Section II.B.

VI. CONCLUSION

The recent developments in online copyright protections in the United States and Europe are worrisome at best. They go against the fundamental spirit of the web and have the potential to render reference techniques, such as embedding, too expensive to be used. This in turn endangers the ease of access to information as we know it today.

In the United States, a battle is emerging between the server and incorporation tests. As the district courts continue to push for a potential split by embracing one of the two tests and circuit courts remain silent on the issue, they create uncertainty, likely inhibiting the economic and cultural growth of internet. That is why it is crucial for the legislature to take up the issue as soon as possible. By passing the Music Modernization Act,¹⁹⁷ Congress has already recognized the need to adapt current intellectual property protections to the digital world. Congress is in a better position than the courts to resolve this issue in a way that preserves the goals of the internet in light of copyright laws. The legislature could protect linking from copyright liability while still providing economic reimbursement for the efforts of copyright holders, by creating a licensing system specific to copyrighted works on the internet.

In the European Union, Directive (EU) 2019/790 clearly favors stricter copyright protections in linking practices on the internet.¹⁹⁸ Although its consequences are yet to be seen, the European Union is likely to shortly become a hotbed of linking and embedding issues. If the United States and Europe continue to take different approaches to embedding, the world might find itself in a precarious position. This in turn, is likely to continue to spur internet innovation in the United States, while creating a larger gap between technology innovation in Europe and the rest of the world. By adopting the proposed licensing system, the United States could once again become a leader in internet regulation and lead the world in creating a solution to the currently inevitable clash between copyright and the internet.

197. *See supra* Section III A.

198. *See supra* Section IV.C.

**FLAUNTING THE SCARLET LETTER:
CONSUMER REGULATION OF
TRADEMARK MORALITY AFTER *IANCU V.*
*BRUNETTI***

Kyung-Lee Kelly Go[†]

I. INTRODUCTION

No matter how shocking or offensive the word “FUCT” may be to the general public, the Supreme Court in *Iancu v. Brunetti* held that the government violated Erik Brunetti’s First Amendment right to free speech by refusing to register his clothing brand name “FUCT” as a trademark.¹ Under the Lanham Act, the United States Patent and Trademark Office (PTO)—the government agency in charge of registering trademarks—was required to reject “FUCT” and similar trademarks that are “immoral . . . or scandalous.”² Only the trademarks that, in the eyes of the government, met the society’s moral standard passed the trademark registration test and received the government seal of approval. A trademark like “FUCT” that was “the equivalent of [the] past participle form of a well-known word of profanity” and associated with “extreme nihilism” and “anti-social” behavior was not worthy of government approval.³

“So the Lanham Act,” Justice Kagan stated in the opinion of the Court, “allows registration of marks when their messages accord with, but not when their messages defy, society’s sense of decency or propriety.”⁴ The statute, “on its face, distinguishes between two opposed sets of ideas: those aligned with conventional moral standards and those hostile to them; those induc[ing] societal nods of approval and those provoking offense and condemnation.”⁵ The government’s ban on registering “immoral or scandalous” trademarks “discriminates on the basis of viewpoint” and “results in viewpoint-

DOI: <https://doi.org/10.15779/Z38QF8JK5G>

© 2020 Kyung-Lee Kelly Go.

† J.D., 2021, University of California, Berkeley, School of Law.

1. *See Iancu v. Brunetti*, 139 S. Ct. 2294 (2019).

2. 15 U.S.C. § 1052(a) (2012).

3. *See Brunetti*, 139 S. Ct. at 2297–98.

4. *Id.* at 2300.

5. *Id.*

discriminatory application,” thereby “collid[ing] with our First Amendment doctrine.”⁶

The Court’s decision in *Brunetti* to strike down the “immoral or scandalous” ban is a logical extension of its earlier decision in *Matal v. Tam*,⁷ in which the Court invalidated “a neighboring provision”⁸ in the Lanham Act that prohibited registering “disparaging” trademarks.⁹ Indeed, as if picking up where it had left off in *Tam*, the Court in *Brunetti* began with a reference to *Tam* and stated that it had “made clear in *Tam*” already that “a law disfavoring ‘ideas that offend’ discriminates based on viewpoint, in violation of the First Amendment.”¹⁰ Thus, the Court declared: “We hold that this [immoral or scandalous] provision infringes the First Amendment for the same reason [that the disparaging provision did]: It too disfavors certain ideas.”¹¹

The “immoral or scandalous” ban disfavored trademarks that would “shock” or “offend” the public’s “sense of truth, decency, or propriety,”¹² a standard that is difficult if not impossible to grasp and certainly not for the government to define. Accordingly, the PTO was given unfettered discretion in working with this test, as evidenced by its inconsistent and unpredictable results. For instance, based on the PTO’s determination, “HAVE YOU HEARD THAT SATAN IS A REPUBLICAN?” flunked the test,¹³ but “THE DEVIL IS A DEMOCRAT” passed the test and became a registered trademark.¹⁴ Justice Gorsuch’s questions during oral argument best sum up the problem: “How is a person -- a person who wants to get a mark supposed to tell what the PTO is going to do? Is it a flip of the coin?”¹⁵ Similarly, many

6. *Id.* at 2296.

7. See Megan Carpenter, *Opinion Analysis: Long-Time Prohibition on Registration of Scandalous and Immoral Trademarks is Struck Down*, SCOTUSBLOG (June 24, 2019, 6:59 PM), <https://www.scotusblog.com/2019/06/long-time-prohibition-on-registration-of-scandalous-and-immoral-trademarks-is-struck-down/>.

8. *Brunetti*, 139 S. Ct. at 2297.

9. See *Matal v. Tam*, 137 S. Ct. 1744, 1753 (2017); 15 U.S.C. § 1052(a).

10. *Brunetti*, 139 S. Ct. at 2301.

11. *Id.* at 2297.

12. *Id.* at 2298.

13. *Id.*

14. See PETER S. MENELL, MARK A. LEMLEY & ROBERT P. MERGES, *INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE* 972 (2019). For more examples, see also *Brunetti*, 139 S. Ct. at 2300–01, where Justice Kagan provides a lengthy list of examples of inconsistent decisions made by the PTO.

15. Transcript of Oral Argument at 21, *Iancu v. Brunetti*, 139 S. Ct. 2294 (2019) (No. 18–302), available at https://www.supremecourt.gov/oral_arguments/argument_transcripts/2018/18-302_7k47.pdf.

trademark scholars “widely agree” that the current trademark registration system is “broken.”¹⁶

As a trademark speaks to the quality of the product and reputation of the seller, trademark law and regulations should facilitate commerce by protecting consumers from deceitful sellers’ unauthorized misappropriation of trademarks of high-quality products. Regulating trademarks that are “immoral or scandalous” does not achieve this goal, as shown by Erik Brunetti’s trademark case. As the founder and long-time owner of the skater fashion brand FUCT, Brunetti saw many counterfeit sellers on eBay and Amazon selling knockoff FUCT clothes and “costing him real money.”¹⁷ Hence, he became motivated to register his brand name as a trademark so that he could “go after the copycats and shut them down.”¹⁸ Trademark registration is not necessary since “the owner of an unregistered mark may still use it in commerce and enforce it against infringers” to use the trademark in commerce and enforce it against infringers.¹⁹ However, registration gives trademark owners “valuable benefits” such as “prima facie evidence” of the mark’s validity and “constructive notice” of the registrant’s claim of ownership, thereby allowing the registered owner to defeat several defenses in trademark infringement suits.²⁰ Denying Brunetti these benefits and protection because he chose to use a trademark that the government deems “immoral or scandalous” misses the mark. Determining what is “immoral or scandalous” is neither within the government’s capability nor responsibility. Instead, trademark regulations should seek to protect Brunetti and consumers from counterfeit sellers and their low-quality products.

Now that the “disparaging” and “immoral or scandalous” bans have been struck down by the Supreme Court, how might the government regulate these marks? Congress is certainly free to rewrite the statute, since Congress “alone has the institutional competence, democratic legitimacy, and (most importantly) constitutional authority to revise statutes”²¹ The dissents in *Brunetti* raise salient policy concerns that support rewriting rather than completely eliminating the government’s regulation of trademark morality.²²

16. Rebecca Tushnet, *Registering Disagreement: Registration in Modern American Trademark Law*, 130 HARV. L. REV. 867, 868 (2017).

17. Nina Totenberg, *Supreme Court Dances Around the F-Word with Real Potential Financial Consequences*, NPR (Apr. 16, 2019, 5:00 AM), <https://www.npr.org/2019/04/16/713632552/supreme-court-dances-around-the-f-word-with-real-potential-financial-consequence>.

18. *Id.*

19. *See Brunetti*, 139 S. Ct. at 2297.

20. *Id.* at 2297–98 (quoting 15 U.S.C. §§ 1115(a), 1072 (2012)).

21. *See Wis. Cent. Ltd. v. United States*, 138 S. Ct. 2067, 2074 (2018).

22. *See infra* Section III.B.3.

Nevertheless, the “disparaging” and “immoral or scandalous” bans should not be revived or revised. It is true, as Justice Sotomayor pointed out, that the elimination of the “immoral or scandalous” ban would leave the government with “no choice but to begin registering marks containing the most vulgar, profane, or obscene words and images imaginable.”²³ It is also understandable, as Justice Breyer noted, to be concerned with words like “FUCT” being used as the name of “a clothing line that includes apparel for children and infants”²⁴ and even “believe that such words should not be displayed in public spaces” where “children are likely to be present.”²⁵ However, refusing to register “FUCT” as a trademark would not necessarily result in a fewer number of shirts with the FUCT logo; instead, it may increase counterfeit uses of the FUCT logo and make it more difficult for Brunetti to win fights against misappropriators of his brand.

The Court made the right decision to invalidate the “immoral or scandalous” provision, and Congress should not attempt to resurrect it. The Court and the government agree that it is unconstitutional to “deny registration based on the views expressed” by a mark.²⁶ No matter how the “immoral or scandalous” provision is rewritten, it will be inherently viewpoint-discriminatory and thus unconstitutional. Trademark law and regulations should be designed to ensure consumer protection and fair competition: incentivize sellers to sell high-quality products and build good reputation by protecting their trademarks from unfair use.²⁷ Whether and to what extent sellers may use offensive or scandalous words as their trademarks are questions left to consumers, not the government.

This Note argues consumers are the real players in the market with the purchasing power to favor and disfavor certain trademarked products, and they are in the best position to gauge the current level of morality in society. The government should stay out of regulating trademark morality, and let consumers choose what kinds of trademarks they want to see displayed in the marketplace. Let them decide whether they want to see “immoral or scandalous” trademarks and which trademarks should be deemed immoral or scandalous. After all, what was thought to be scandalous a couple of decades ago may “now be thought merely humorous (or even quaint)” due to changes

23. *Brunetti*, 139 S. Ct. at 2308 (Sotomayor, J., concurring in part and dissenting in part) (parenthesis omitted).

24. *Id.* at 2304 (Breyer, J., concurring in part and dissenting in part).

25. *Id.* at 2307.

26. *Id.* at 2299.

27. *See infra* Section II.A.1.

in societal attitudes.²⁸ As Judge Clevenger of the Federal Circuit pointed out, “Today’s scandal can be tomorrow’s vogue.”²⁹

The following Part II provides a brief overview of trademark law and examines the PTO’s interpretation and application of the “immoral or scandalous” provision. Part III discusses the *Iancu v. Brunetti* decision and the four separate concurring and dissenting opinions that suggest the possibility of a congressional rewrite of the now-invalidated provision. Part IV argues against a rewrite and in support of entrusting consumers with regulating trademark morality in the market. The Note also highlights some of the broader policy implications of the *Brunetti* decision that demonstrate why this kind of trademark regulation should be left to consumers rather than to the government.

II. TRADEMARK LAW AND THE PTO’S APPLICATION OF THE PROVISION AT ISSUE

A. TRADEMARK LAW

1. *History and Purpose*

Trademarks have co-existed with humans for the last 4,000 years.³⁰ Early merchants trading in China, Persia, Egypt, and Greece marked their clothing and pottery goods with words or symbols to identify themselves.³¹ These marks served several purposes that resemble those of today’s trademarks. First, the early marks acted as advertisements, allowing merchants to promote their names in front of potential customers.³² Second, the marks helped resolve ownership disputes by proving that the goods with a particular mark were being sold by a specific merchant.³³ Third, these marks provided a guarantee of quality, as merchants were putting their reputations on the line by identifying themselves with their goods.³⁴ In serving these purposes,

28. See *In re Old Glory Condom Corp.*, 26 U.S.P.Q.2d (BNA) 1216, 1218 (I.T.A.B. 1993).

29. *In re Mavety Media Group Ltd.*, 33 F.3d 1367, 1371 (1994) (discussing, amongst other cases, QUEEN MARY, in reference to women’s underwear, which was held as scandalous in *Ex parte Martha Maid Mfg. Co.*, 37 U.S.P.Q. (BNA) 156 (Comm’r Pats. 1938)).

30. MENELL ET AL., *supra* note 14, at 873 (quoting WILLIAM H. BROWNE, A TREATISE ON THE LAW OF TRADEMARKS 1–14 (1885)).

31. *Id.*

32. *Id.*

33. *Id.*

34. *Id.*

trademarks have helped reduce transaction costs by making it easier for consumers to identify desired goods quickly and accurately.³⁵

The purpose of trademarks, then, is to help consumers distinguish between the goods that they “wish to purchase” and those that they “want to avoid.”³⁶ Hence, it is not surprising to find the roots of trademark law in the soil of anti-fraud law³⁷ as trademark-related disputes began arising among merchants and counterfeiters. Judge Joseph Story granted the first trademark injunction in 1844 against the counterfeiters of “Taylor’s Persian Thread.”³⁸ Then, in 1870 Congress enacted a federal trademark law that grounded trademark protection in the Intellectual Property Clause of the Constitution.³⁹ However, the Supreme Court struck it down and held that trademark should not be grounded in the Intellectual Property Clause, as it was not an “invention[] and discover[y] in the arts and sciences,” but rather a “useful and valuable aid or instrument of commerce” that belonged in the Commerce Clause.⁴⁰ Congress followed the Court’s recommendation and enacted the Trademark Law of 1905 under the Commerce Clause.⁴¹ Consequently, use of a mark in commerce has long been a requirement for trademark protection.⁴²

Although trademark registration is not required for bringing a civil action against alleged counterfeiters,⁴³ owners who register their marks enjoy additional “valuable benefits,”⁴⁴ including a presumptive right of validity, constructive nationwide rights, and customs rights to restrict imports of goods labeled with infringing marks.⁴⁵ The Lanham Act, which lays out the foundation of the modern trademark law,⁴⁶ specifies various requirements and restrictions regarding trademark registration—such as the prohibition of

35. *Id.*

36. *Iancu v. Brunetti*, 139 S. Ct. 2294, 2306 (2019) (Breyer, J., concurring in part and dissenting in part) (quoting *Matal v. Tam*, 137 S. Ct. 1744, 1751 (2017)).

37. See Kenneth J. Vandeveld, *The New Property of the Nineteenth Century: The Development of the Modern Concept of Property*, 29 BUFF. L. REV. 325, 340 (1980).

38. See *Taylor v. Carpenter*, 23 F. Cas. 742 (1844).

39. MENELL ET AL., *supra* note 14, at 874.

40. *The Trade-Mark Cases*, 100 U.S. 82, 93–95 (1879).

41. MENELL ET AL., *supra* note 14, at 874.

42. *Id.*

43. See 15 U.S.C. § 1125(a) (2012).

44. *Iancu v. Brunetti*, 139 S. Ct. 2294, 2297 (2019).

45. See 15 U.S.C. §§ 1115(a) (2016), 1072 (2011), 1125(b) (2012).

46. Mark Lemley, *The Modern Lanham Act and the Death of Common Sense*, 108 YALE L.J. 1687, 1687 (1999). The Lanham Act is also known as the Trademark Act of 1946.

registering any mark that “so resembles” another mark⁴⁷ or that is “merely descriptive” of the good on which it is used.⁴⁸

2. *At Issue: Section 2(a) of the Lanham Act*

Besides the provisions prohibiting registration of marks that are likely to deceive or confuse consumers, the Lanham Act also includes a provision prohibiting registration of marks that may shock public morality. According to section 2(a) of the Lanham Act, marks that consist of “immoral . . . or scandalous” matter “shall be refused registration.”⁴⁹ The legislative history of the Lanham Act, which spans across four congressional terms, does not provide much helpful guidance as to the meaning of “immoral” or “scandalous.”⁵⁰ Instead, this prohibition has been part of the Lanham Act for so long that it was taken as a given, without any showing of how it relates to and furthers the purpose of broader trademark law.⁵¹ Hence, the PTO was left on its own to interpret and apply the statutory prohibition, which resulted in “a long line of arbitrary and contradictory decisions.”⁵²

B. THE PTO’S INTERPRETATION AND APPLICATION

In its Trademark Manual of Examining Procedure (TMEP), the PTO explicitly states that “[t]here is little legislative history concerning the intent of Congress with regard to the [immoral or scandalous] provision.”⁵³ Accordingly, the PTO has interpreted this provision by looking into “its ordinary and common meaning.”⁵⁴ Based on that interpretation, it created a test that asks whether a “substantial composite of the general public” would find the mark “shocking to the sense of propriety, offensive to the conscience

47. 15 U.S.C. § 1052(d) (2012).

48. 15 U.S.C. § 1052(e) (2012).

49. 15 U.S.C. § 1052(a) (2012). Such prohibition on “immoral or scandalous” trademarks was enacted initially as section 5(a) of the Trademark Act in 1905 and was reenacted as part of the Lanham Act in 1946. THE PATENT AND TRADEMARK OFFICE, TRADEMARK MANUAL OF EXAMINING PROCEDURE § 1203.01 (Oct. 2018), <https://tmepp.uspto.gov/RDMS/TMEP/current#/current/TMEP-1200d1e3042.html> [hereinafter TMEP].

50. TMEP, *supra* note 49, at § 1203.01.

51. Alan Behr, *The Implications are Potentially Broad, in Iancu v. Brunetti Ruling – Trademark Community Has Its Say on Implications of Momentous US Supreme Court Decision* (Tim Lince, ed.), WORLD TRADEMARK REV. (June 25, 2019), <https://www.worldtrademarkreview.com/enforcement-and-litigation/iancu-v-brunetti-ruling-trademark-community-has-its-say-implications>.

52. Brief for the American Civil Liberties Union and the ACLU of the District of Columbia at *25, as Amici Curiae Supporting Respondent, *Iancu v. Brunetti*, 139 S. Ct. 2294 (2019) (No. 18-302).

53. TMEP, *supra* note 49, at § 1203.01.

54. *Id.* (quoting *In re Riverbank Canning Co.*, 95 F.2d 327, 328 (C.C.P.A. 1938).

or moral feelings or calling out for condemnation,” or “‘vulgar,’ defined as ‘lacking in taste, indelicate, [or] morally crude.’”⁵⁵

Unfortunately, this test did not work as intended when carried out in practice. As Justice Gorsuch pointed out in *Brunetti*, “there are shocking numbers of [marks] granted [registration] and ones refused that . . . do look remarkably similar.”⁵⁶ For example, the mark BITCH was approved and registered over 150 times, while KICKABITCH was rejected.⁵⁷ Other examples include: GRAMMAR NAZI (approved) versus COFFEE NAZI (rejected); F·A·G FABULOUS AND GAY (approved) versus MARRIAGE IS FOR FAGS (rejected); WHORES FROM HELL (approved) versus THE CHRISTIAN PROSTITUTE (rejected).⁵⁸ Justice Gorsuch asked, “How is a reasonable citizen supposed to know . . . how the government’s going to treat their mark?”⁵⁹

Erik Brunetti was one such citizen, wondering why FCUK was able to pass the test eight times,⁶⁰ while his mark, FUCT, could not pass even once. By the end of his nearly thirty-year-long journey to successful trademark registration, he and his mark ended up all the way in the Supreme Court.

III. IANCU V. BRUNETTI

A. FUCT: THE ORIGIN OF ITS GOODS AND SERVICES

Despite its showcase in the nation’s highest Court, the FUCT brand rose from a rather humble beginning in the 1980s in Venice, California, where street skateboarding culture was vibrant and thriving.⁶¹ Erik Brunetti and Natas Kaupas, an entrepreneurial graphic designer and a famous professional skateboarder in their twenties, were trying to come up with a name for a new graphic design company that they were about to start together.⁶² They thought it would be clever to call their company “FUCT” and present it as “very

55. *Id.* (quoting *In re Boulevard Entm’t, Inc.*, 334 F.3d 1336, 1340 (Fed. Cir. 2003) (quoting *In re McGinley*, 660 F.2d 481, 486 (C.C.P.A. 1981) (quoting *In re Runsdorf*, 171 USPQ 443, 444 (T.T.A.B. 1971))).

56. Transcript of Oral Argument, *supra* note 15, at 19.

57. See Joint Appendix for Noel J. Francisco & John R. Sommer at 74, *Iancu v. Brunetti*, 139 S. Ct. 2294 (2019) (No. 18-302), 2019 WL 914147.

58. *Id.*

59. Transcript of Oral Argument, *supra* note 15, at 19.

60. FCUK AT HOME also was among those approved for registration. See Francisco & Sommer, *supra* note 57, at 71.

61. See Ian Michna, *Discussing the History of Fuct & the Current Streetwear Market*, JENKEM (May 23, 2019), <http://www.jenkemmag.com/home/2019/05/23/discussing-history-fuct-current-streetwear-market/>.

62. *Id.*

corporate,” so as to have people question the pronunciation based on the way the logo looked.⁶³ Soon, one of their corporate clients suggested that they make and sell skater pants and shorts with the FUCT logo⁶⁴—and that is how the FUCT brand, which caters to “wealthy rebellious skaters”⁶⁵ today, was born.⁶⁶ The skater-fashion brand continued to grow, both at home and abroad, even after Brunetti and Kaupas parted ways.⁶⁷ Brunetti, who remained with the brand, has since established great pride in his skater customers who have “a sense of community . . . [and are] very communal and very unified.”⁶⁸

With popularity, however, came imitators and knockoffs. As Brunetti described it, big corporations and fashion houses “vulturized” street skateboarding fashion’s “real and legit” personality to compete in their “very dishonest” and cut-throat industry.⁶⁹ More importantly, Brunetti found numerous “copycats” on eBay and Amazon, who were selling “lots of counterfeits” that hurt the FUCT brand’s sales.⁷⁰ Triggered by the ongoing financial and reputational harm, Brunetti applied for trademark registration at the PTO, in hopes of “shut[ting] down the tremendous amount of bootlegging that[] [had] been happening for years.”⁷¹

Brunetti’s journey to successful registration of his mark, however, did not turn out to be so straightforward. Both the PTO and the Trademark Trial and Appeal Board (TTAB) determined that FUCT “flunked” the Lanham Act’s “immoral or scandalous” test.⁷² Defining the mark as the slang-equivalent of “fucked,” i.e., the past participle form of the quintessential word of profanity, the TTAB concluded that FUCT was unregistrable for being just as offensive, profane, or vulgar as “fuck.”⁷³

Nevertheless, the Federal Circuit disagreed with the TTAB’s decision and sided with Brunetti, when he brought a facial challenge to the statutory

63. *Id.*

64. *Id.*

65. Dennis Crouch, *Oops, I Did It Again: Time to Register Those Scandalous Marks*, PATENTLYO (June 24, 2019), <https://patentlyo.com/patent/2019/06/again-register-scandalous.html>.

66. *Id.*; see Thomas Key, *Freedom of Expression Transcends Morality in US Trademark Registration*, THE IPKAT (July 14, 2019), <http://www.ipkitten.blogspot.com/2019/07/freedom-of-expression-transcends.html>.

67. Michna, *supra* note 61.

68. *Id.*

69. *Id.*

70. Totenberg, *supra* note 17.

71. Michna, *supra* note 61.

72. *Iancu v. Brunetti*, 139 S. Ct. 2294, 2298 (2019).

73. *See In re Brunetti*, 877 F.3d 1330, 1337 (Fed. Cir. 2017).

prohibition.⁷⁴ The court concluded that the “century-old provision” of the Lanham Act was in violation of the First Amendment.⁷⁵ While acknowledging that there was a question of whether the “immoral or scandalous” provision was viewpoint-neutral, the court explained that it did not need to resolve that issue: “Independent of whether the immoral or scandalous provision is viewpoint discriminatory, . . . the provision impermissibly discriminate[d] based on content in violation of the First Amendment.”⁷⁶ Furthermore, the court asserted that strict scrutiny should be applied in this case since the statutory prohibition targeted the expressive—and not commercial—content of speech.⁷⁷ When strict scrutiny is applied, the court held, the prohibition would not withstand the review.⁷⁸ Even if the court were to assume *arguendo* that trademarks are “purely commercial speech,” the prohibition still would not have survived the applicable intermediate scrutiny analysis.⁷⁹

In sum, the court concluded that there was no reasonable definition of “scandalous” or “immoral” that would preserve the statutory prohibition’s constitutionality.⁸⁰ In response to the invalidation of a federal statute by a lower court, the Supreme Court then granted certiorari.⁸¹

B. IANCU V. BRUNETTI

1. *The Precedent: Matal v. Tam*

As if picking up where it had left off in *Matal v. Tam*, the Court in *Iancu v. Brunetti* began the opinion by reciting its holding in the former case.⁸² “Two terms ago, in *Matal v. Tam*,” Justice Kagan wrote, “this Court invalidated the Lanham Act’s bar on the registration of ‘disparag[ing]’ trademarks.”⁸³ In accordance with that statutory prohibition, the PTO had refused to register THE SLANTS, the name of an Asian-American band that is also a derogatory

74. *See id.* at 1357.

75. *See id.* (Dyk, J., concurring).

76. *Id.* at 1341. The court defined content-based government restriction on speech as “a law [that] applies to particular speech because of the topic discussed or the idea or message expressed.” *See id.* at 1341–42 (quoting *Reed v. Town of Gilbert*, 135 S. Ct. 2218, 2227 (2015)).

77. *Id.* at 1340.

78. *Id.* at 1342, 1349. To survive strict scrutiny review, the government must “prove that the restriction furthers a compelling interest and is narrowly tailored to achieve that interest.” *Id.* at 1342 (quoting *Reed v. Town of Gilbert*, 135 S. Ct. 2218, 2231 (2015)).

79. *See id.* at 1350. Intermediate scrutiny requires the government to “show at least that the statute directly advances a substantial governmental interest and that the measure is drawn to achieve that interest.” *Id.* (quoting *Sorrell v. IMS Health Inc.*, 564 U.S. 552, 572 (2011)).

80. *Id.* at 1355.

81. *See Iancu v. Brunetti*, 139 S. Ct. 2294, 2298 (2019).

82. *Id.*

83. *Id.* at 2297.

term used in reference to persons of Asian descent.⁸⁴ The Court there rejected the government’s argument that trademarks constitute government speech, and explained that the content of a registered trademark could not be government speech because, if that were true, the government would have been “babbling prodigiously and incoherently,” saying many “unseemly” things.⁸⁵ The Court warned of the danger of passing off private speech as government speech by “simply affixing a government seal of approval,” which would allow the government to “silence or muffle the expression of disfavored viewpoints.”⁸⁶

The Court also refused to resolve the issue of whether trademarks constitute commercial speech, since the disparagement clause “cannot withstand even *Central Hudson* review” that applies “relaxed scrutiny” by requiring only that the restriction of speech serve a “substantial interest” and be “narrowly drawn.”⁸⁷ The Court concluded that such prohibition on disparaging trademarks “violates the Free Speech Clause” because it “offends a bedrock First Amendment principle: Speech may not be banned on the ground that it expresses ideas that offend.”⁸⁸

2. *The Majority’s Opinion*

Accordingly, the Court in *Iancu v. Brunetti* held that the “immoral or scandalous” provision “infringes the First Amendment for the same reason: It too disfavors certain ideas.”⁸⁹ Justice Kagan explained that the problem was not that the meanings of “immoral” and “scandalous” are mysterious. Rather,

84. *Matal v. Tam*, 137 S. Ct. 1744, 1751 (2017). When Simon Tam, the band’s leader, questioned the PTO why it “[chose] to apply the racial connotations to this application” when it had “never done that before in the entire history of this country,” the PTO explained that “it was because [Tam] was Asian-American.” Kat Chow, *Asian-American Band Fights to Trademark Name ‘The Slants’*, NPR (Oct. 20, 2013 6:14 AM), <https://www.npr.org/sections/codeswitch/2013/10/20/236235813/asian-american-band-fights-to-trademark-name-the-slants>.

85. *See Matal*, 137 S. Ct. at 1758. The Court has defined commercial speech as speech that does “no more than propose a commercial transaction . . .” *Va. State Bd. of Pharmacy v. Va. Citizens Consumer Council Inc.*, 425 U.S. 748, 762 (1976) (quoting *Pittsburgh Press Co. v. Human Relations Comm’n*, 413 U.S., 376, 385 (1973)).

86. *Matal*, 137 S. Ct. at 1758.

87. *See id.* at 1763–64 (quoting *Cent. Hudson Gas & Electric Corp. v. Public Service Comm’n*, 447 U.S. 557, 564–65 (1980)). A restriction on commercial speech is subject to a four-part analysis that tests whether (1) the speech concerns “lawful activity and [is] not . . . misleading”; (2) the “asserted governmental interest” to be served by the restriction is “substantial”; (3) the restriction “directly advances the governmental interest asserted”; and (4) “whether [the restriction] is not more extensive than is necessary to serve that interest.” *Cent. Hudson*, 447 U.S. at 566.

88. *Matal*, 137 S. Ct. at 1751.

89. *Iancu v. Brunetti*, 139 S. Ct. 2294, 2297–98 (2019).

the problem was that the provision that discriminated against marks meeting such criteria was viewpoint-based.⁹⁰ The provision at issue, she argued, allowed registering marks “when their messages accord with, but not when their messages defy, society’s sense of decency or propriety.”⁹¹ In other words, it distinguished between “two opposed sets of ideas”: those “aligned with conventional moral standards” versus those “hostile to them”; those “inducing societal nods of approval” versus those “provoking offense and condemnation.”; hence, “the statute favors the former, and disfavors the latter.”⁹²

The Court also rejected the government’s proposal to apply a narrowing construction to the statute to remove viewpoint bias.⁹³ Justice Kagan questioned, “How . . . can the Government claim that the ‘immoral or scandalous’ bar is viewpoint-neutral? The Government basically asks us to treat [the PTO’s inconsistent] decisions . . . as PTO examiners’ mistakes.”⁹⁴ She contended that the government was instructing the Court to “ignore” that the text of the statute, “on its face, disfavor[ed] some ideas.”⁹⁵ The government’s assertion was that the Court should try to construe the statute in a way that renders it constitutional, i.e., to narrow the prohibition to “marks that are offensive [or] shocking to a substantial segment of the public because of their *mode* of expression,” regardless of any viewpoints that they might express.⁹⁶ Justice Kagan rejected “that canon of construction,” which is applicable “only when ambiguity exists.”⁹⁷ Here, no such ambiguity existed, and the statute as written “[did] not draw the line at lewd, sexually explicit, or profane marks,” let alone at marks whose mode of expression was offensive.⁹⁸ She concluded that “once the ‘immoral or scandalous’ bar is interpreted fairly, it must be invalidated.”⁹⁹

3. *The Four Separate Opinions*

Much debate over the majority’s decision ensued: Chief Justice Roberts and Justices Alito, Breyer, and Sotomayor each filed his or her own concurring or concurring-in-part-and-dissenting-in-part opinions.¹⁰⁰ Justice Alito, for

90. *Id.* at 2299.

91. *Id.* at 2300.

92. *Id.*

93. *Id.* at 2301.

94. *Id.*

95. *Id.*

96. *Id.* (quoting Transcript of Oral Argument, *supra* note 15, at 11).

97. *Id.*

98. *Id.*

99. *Id.*

100. *See id.*

instance, joined the majority in full yet wrote a separate concurring opinion to emphasize how “important [it was] for this Court to remain firm on the principle that the First Amendment does not tolerate viewpoint discrimination,” especially “[a]t a time when free speech is under attack.”¹⁰¹ Moreover, he made clear that the members of the Court “are not legislators and cannot substitute a new statute for the one now in force,” noting that the majority’s decision “does not prevent Congress from adopting a more carefully focused” provision.¹⁰²

On the other hand, Chief Justice Roberts and Justice Breyer joined the majority only in part and each filed an opinion concurring in part and dissenting in part.¹⁰³ They both agreed with Justice Sotomayor’s argument that the majority erred in treating the “immoral” and “scandalous” prongs as one provision, when the latter is “susceptible of . . . a narrowing construction.”¹⁰⁴

Specifically, Justice Breyer expressed in detail his disagreement with the Court’s usual practice of taking a categorical approach to First Amendment questions.¹⁰⁵ Instead of focusing on categorizing the statute at issue “as an example of ‘viewpoint discrimination,’ ‘content discrimination,’ ‘commercial speech,’ ‘government speech,’ or the like,” he asserted, “I would appeal more often and more directly to the values the First Amendment seeks to protect,”¹⁰⁶ and focus on whether the statute “works speech-related harm that is out of proportion to its justifications.”¹⁰⁷

Justice Breyer also addressed the practical effect of the majority’s decision. “Just think about how you might react if you saw someone wearing a t-shirt or using a product emblazoned with an odious racial epithet.”¹⁰⁸ He contended that the government may have an interest in protecting “the sensibilities of children” by prohibiting trademark registration of words like FUCT that “should not be displayed . . . where children are likely to be present.”¹⁰⁹

101. *Id.* at 2302–03 (Alito, J., concurring).

102. *Id.* at 2303.

103. *See id.* at 2302–06 (Roberts, C. J., concurring in part and dissenting in part).

104. *Id.* at 2303.

105. *See id.* at 2304–08 (Breyer, J., concurring in part and dissenting in part).

106. *Id.* at 2304–05.

107. *Id.* at 2305 (quoting *United States v. Alvarez*, 567 U.S. 709, 730 (2012)).

108. *Id.* at 2307.

109. *Id.* Regarding the question of “whether the First Amendment permits the Government to rely on [the narrowly constructed] statute . . . to deny the benefits of federal trademark registration to marks like the one at issue here, which involves the use of the term ‘FUCT,’” Justice Breyer stated, “I believe the answer is ‘yes.’” *Id.* at 2304.

Justice Sotomayor, with whom Justice Breyer joined, voiced the strongest dissent against the majority's decision.¹¹⁰ At the outset, Justice Sotomayor conceded and agreed with the majority that there was "no tenable way to read" the word "immoral," as it "clearly connotes a preference for 'rectitude and morality' over its opposite."¹¹¹ Nonetheless, she argued that the word "scandalous," although "on its own . . . is ambiguous," could mean "simply indecent, shocking, or generally offensive" and thus could be saved from invalidation.¹¹² The majority, therefore, "ill-advisedly collapse[d]" the two words, when it should have treated "immoral" as covering marks that are "offensive because they transgress social norms" and "scandalous" as covering marks that are "offensive because of the *mode* in which they are expressed."¹¹³ She concluded that she would apply a narrowing construction to the "scandalous" prong to address "only obscenity, vulgarity, and profanity" and reject Brunetti's facial challenge.¹¹⁴

The majority responded to Justice Sotomayor's argument in a footnote, explicitly "reject[ing] the dissent's statutory surgery."¹¹⁵ "[T]he dissent thinks that the term 'scandalous' can be read [narrowly] as the Government proposes," the majority noted, "[b]ut that term is not 'ambiguous,' as the dissent argues . . . it is just broad."¹¹⁶

As Justice Alito suggested, Congress can decide whether to enact a new statute to replace section 2(a). With the majority's decision in hand, Congress now faces several options: for instance, it may adopt "a more carefully focused" provision.¹¹⁷ It can take Justice's Sotomayor's approach of a narrowing construction by reviving only the "scandalous" prong of the provision.¹¹⁸ Alternatively, Congress can decide to do nothing with the invalidated provision altogether, keeping the PTO's trademark registration system free of any test for immoral or scandalous marks. For the reasons below, this Note argues that Congress should resist any temptation to rewrite the "immoral or scandalous" statute. Instead, Congress should leave the task of regulating trademark morality to consumers, who are positioned to be the

110. "The Court's decision today will beget unfortunate results." *Id.* at 2308 (Sotomayor, J., concurring in part and dissenting in part).

111. *Id.* at 2309.

112. *Id.*

113. *Id.* at 2311 (emphasis added).

114. *Id.* at 2313.

115. *Id.* at 2302 n*.

116. *Id.*

117. *Id.* at 2303 (Alito, J., concurring).

118. *Id.* at 2308.

most effective and efficient regulators of potentially immoral or scandalous trademarks in the market.

IV. CONSUMERS AS THE REGULATORS OF TRADEMARK MORALITY

Entrusting consumers with the task of regulating trademark morality resolves several legal and practical problems posed by the test in the PTO's trademark registration process. Consumers have what the PTO does not: purchasing and boycotting power that influences businesses seeking to promote their brands to consumers. Additionally, it is consumers, not the PTO, who can best measure the current temperature of public sentiment and sense of morality. Lastly, policy implications of the Court's decision in *Brunetti*—such as upholding First Amendment values, promoting predictability and the equal treatment of trademark applicants, and adhering to trademark law's intended purpose—also demonstrate why evaluation of trademark morality is best left in the hands of consumers.

A. CONSUMERS' PURCHASING AND BOYCOTTING POWER: THE REAL DISINCENTIVE TO USING IMMORAL OR SCANDALOUS TRADEMARKS

In practice, it is consumers' taste and their own sense of morality, rather than a hypothetical test set up by the government, that can best gauge which trademarks are offensive to the public and which ones are not. Since the fate of a trademark ultimately depends on whether and how much it is used in commerce, trademark registration and its accompanying benefits become meaningless if the mark is not even in use. Therefore, consumers' purchasing decisions, which drive the usage and popularity of trademarks in the market, inevitably exert significant influence on business owners' branding and marketing decisions—as illustrated by the actions of Gap and the Redskins.

1. *Gap*

The actions of the clothing retailer company Gap serve as a good example of a brand owner's response to consumer influence. In 2016, Gap publicized an ad featuring a taller Caucasian girl posing next to and resting her arm on top of the head of a shorter African-American girl (as shown in Figure 1A).¹¹⁹ As soon as the ad went live, it ignited a huge public outrage and criticism against the casting of the black child as “an armrest.”¹²⁰ One Twitter comment, which received more than a thousand Likes, read: “@GapKids proving girls

119. John Kell, *Gap Apologizes For 'Racist' Ad*, ENTREPRENEUR (Apr. 5, 2016), <https://www.entrepreneur.com/article/273593>.

120. *Id.*

can do anything . . . unless she's Black. Then all she can do is bear the weight of White girls."¹²¹ Without any hesitation or delay, Gap immediately issued an apology and replaced the image with a different one (as shown in Figure 1B) that looked similar but less offensive.

Figure 1A: The Original Version of the Ad by Gap



121. Fatima La’Juan (@TheTherapyDiva), TWITTER (Apr. 2, 2016, 2:21 PM), https://twitter.com/TheTherapyDiva/status/716375045421944832?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E716375045421944832&ref_url=https%3A%2F%2Ffortune.com%2F2016%2F04%2F05%2Fgap-pulls-controversial-ad-image%2F.

Figure 1B: The Edited Version After Consumer Complaint



This controversy demonstrates how essential it has become for big brands like Gap to “navigate the turbulent nature of social media,” where anything from ad campaigns to new products can be “harshly criticized in a matter of hours.”¹²² As brand owners are well aware that such criticism and backlash can slash profits and even lead to a complete boycott, they understandably react quickly to consumers’ feedback by “pulling the offensive material and issuing a swift apology, as Gap did in this instance.”¹²³

Although this incident is an example of consumers reacting to a controversial ad, rather than a controversial trademark, it nevertheless shows that business owners would be even more careful when it comes to choosing and promoting their trademarks. An advertisement may be ephemeral and separable from the product, but a trademark is often permanently imprinted onto and thus inseparable from the product that consumers decide to buy and own. Therefore, it would be much more difficult to rectify the kind of mistake that Gap made, had that mistake involved a trademark instead of a mere seasonal ad, the latter of which Gap was able to quickly take down and isolate from its brand.

122. Kell, *supra* note 119.

123. *Id.*

Therefore, the natural process—in which consumers react disapprovingly to what they find as offensive or shocking messages, and businesses respond promptly by making necessary changes to conform to contemporaneous public sentiment—renders superfluous any government regulation of trademark morality on behalf of consumers. Consumers would be unlikely to take home a product that bears a trademark they find offensive—let alone wish to be perceived by others as being associated with that trademark—especially if, like the FUCT apparel, the product is one that consumers wear or carry in public. As the following example shows, the success and lifespan of a particular trademark depend on its usage in the market, which is largely driven by consumers themselves—regardless of the trademark’s registration status.

2. REDSKINS

REDSKINS is both the name of a national football team and a word viewed by many Native Americans as “a deeply offensive slur.”¹²⁴ For nearly half a century, Native American leaders have been urging the football team to change its name, and also have been challenging the validity of the REDSKINS trademark registration based on the Lanham Act’s “disparagement” provision.¹²⁵ The team has refused to change its name but did agree to replace “Scalp ’em” in the team’s fight song with “Beat ’em” and to get rid of the cheerleaders’ black braided wigs.¹²⁶ Daniel Snyder, the current owner, has sworn that he will never change the name of his football team: “It’s that simple. NEVER — you can use caps,” he said in an interview.¹²⁷ Even former President Barack Obama, who expressed his personal view that he would change the name if he were the owner, could not convince Snyder.¹²⁸

Then, the decision in *Matal v. Tam* came down, in which the Court invalidated the “disparagement” provision in support of THE SLANTS.¹²⁹ This meant that the Native Americans who had been fighting long legal battles against the Washington Redskins team had to “call[] it quits” and concede that

124. Ian Shapira & Ann E. Marimow, *Washington Redskins Win Trademark Fight Over the Team’s Name*, WASH. POST (June 29, 2017), https://www.washingtonpost.com/local/public-safety/2017/06/29/a26f52f0-5cf6-11e7-9fc6-c7ef4bc58d13_story.html?utm_term=.b5cb74ceed07.

125. *Id.*

126. *Id.*

127. *Id.*

128. *Id.*

129. *Matal v. Tam*, 137 S. Ct. 1744 (2017); 15 U.S.C. § 1052(a).

“[t]here’s no more challenge to make.”¹³⁰ Snyder, on the other hand, did not hesitate to express his excitement: “I am THRILLED! Hail to the Redskins.”¹³¹

The practical effect, however, would not have been much different for REDSKINS, had the decision in *Matal v. Tam* come out the other way. Even if the Court had not decided to invalidate the “disparagement” provision—and even if the Native Americans somehow had won the legal battle against “the most entrepreneurial of haters”¹³² by successfully invalidating the trademark registration of REDSKINS—that mark nevertheless would survive and continue to thrive in the market so long as its fans continue to consume it and associate with it as part of their identities as they do so today. In other words, even as an *unregistered* mark, REDSKINS would still be emblazoned on team jerseys and flags, being “displayed . . . where children are likely to be present,” just as it had been as a registered trademark, only without the ® symbol.¹³³ As unfortunate as it is that the Redskins team and its owner are not as receptive to external feedback as Gap was, trademark law has proven itself to be unhelpful in this case. What may be helpful, instead, is the public awareness of the trademark’s offensiveness and the resulting decrease in the public’s consumption of and association with the REDSKINS brand. Therefore, it is consumers’ purchasing and boycotting power that can drive this trademark out of the marketplace, rather than the cancellation of its trademark registration with the PTO. So long as consumers continue to support a brand and choose not to change their consumption behavior, that brand will live on in the marketplace regardless of whether it is granted or denied trademark registration.

B. FLUIDITY IN WORDS’ DEFINITIONS AND PUBLIC SENTIMENT

Meanings that society attaches to words and symbols evolve over time, making it difficult for the PTO to decide which ones are meant to offend or shock “society’s sense of decency or propriety” today.¹³⁴ Public sentiment changes over time as well, and such fluid nature of public sentiment and words’

130. Shapira & Marimow, *supra* note 124.

131. JP Finlay, *Dan Snyder ‘THRILLED’ with Supreme Court Decision that Should Protect Redskins Name*, NBC SPORTS (June 19, 2017), <https://www.nbcsports.com/washington/washington-redskins/dan-snyder-thrilled-supreme-court-decision-should-protect-redskins-name>.

132. Sonia K. Katyal, *Commentary: Brands Behaving Badly*, 109 L.J. INT’L TRADEMARK ASS’N 819, 824 (2019); *see also* *Harjo v. Pro-Football Inc.*, 50 U.S.P.Q.2d 1705, 1747 (BNA) (noting that Redskins fans “engag[e] in antics that clearly poke fun at Native American culture and portray Native Americans as savages and buffoons”).

133. *See Iancu v. Brunetti*, 139 S. Ct. 2294, 2307 (2019) (Breyer, J., concurring in part and dissenting in part).

134. *Id.* at 2300.

meanings undermine any trademark registration system that requires the PTO to fix a particular meaning onto a word so as to determine whether that word is “immoral or scandalous.” Such decision on trademark registrability of a word—which locks the word’s meaning at the time of review—becomes susceptible to being outdated or challenged if the public subsequently changes its sentiment towards the word, or, if the word itself changes in meaning. Therefore, the determination as to whether a word is offensive to the general public should be made by the contemporaneous public, rather than the examiners at the PTO who rely on dictionary definitions.

Take the following OK-sign as an example. Zina Bash—a former White House staffer and clerk to then-nominee Brett Kavanaugh—was sitting at the Supreme Court nomination hearing in 2018, where she was shown on camera as if making an OK sign with her hand resting on her arm.¹³⁵ Unbeknownst to her, however, the OK sign had been gathering much attention in the media after several segments of far-right groups and white supremacists launched the so-called “Operation O-KKK” in 2017 in an effort to claim the OK sign as a symbol of white supremacy.¹³⁶ Bash’s hand gesture immediately ignited an “online furor,” despite attempts by her husband (and U.S. Attorney) John Bash to assure “[e]veryone tweeting this vicious conspiracy theory” that Zina—who is Jewish, part Mexican, and a descendant of Holocaust survivors—would never associate herself with such meaning.¹³⁷ As this short-lived but nevertheless highly controversial incident demonstrates, both a symbol’s meaning and social attitudes towards that symbol are susceptible to rapid change that can take place in a matter of a year or two.

If the OK sign in this scenario had been reviewed under the PTO’s “immoral or scandalous” trademark ban, would it have passed the test? If the white supremacists were to succeed in their Operation O-KKK to claim the OK sign as their own, the PTO may have had to consider denying trademark registration on the basis that the sign would shock or offend a substantial composite of the public. But if such offensive meaning did not permanently stick to the OK sign and instead faded away after a while, the PTO then would have had to reassess the meaning and perhaps start approving the sign for trademark registration again.

135. Mahita Gajanan, *A Kavanaugh Supporter Was Accused of Making a White Power Symbol. She’s a Descendant of Holocaust Survivors*, TIME (Sept. 5, 2018), <https://time.com/5386860/zina-gelman-bash-white-power-symbol/>.

136. *How the “OK” Symbol Became a Popular Trolling Gesture*, ADL (May 1, 2017), <https://www.adl.org/blog/how-the-ok-symbol-became-a-popular-trolling-gesture>; *4CHAN-OPERATIONOKKK.JPG*, SOUTHERN POVERTY LAW CENTER, <https://www.splcenter.org/files/4chan-operationokkkjpg> [<https://perma.cc/F4PQ-QXLP>].

137. Gajanan, *supra* note 135.

Thus, the government should not attempt to fix a particular meaning onto a symbol and guess whether the general public would find it offensive. Instead, the general public should decide for itself as to whether, based on contemporaneous public sentiment, it finds the meaning of the symbol offensive enough to ban it from the market. Put simply, the question of whether “a substantial composite of the general public” would find a mark shocking or offensive¹³⁸ is best answered by the general public itself, not the government. When a substantial composite of the general public indeed finds a mark—registered or unregistered—offensive or scandalous, it may choose to avoid purchasing products associated with that mark and eventually drive it out of the market; alternatively, it may induce the business owner directly to abandon the mark for a less offensive one. This organic process is adaptive to changes in words’ meanings and social attitudes over time, thereby enabling regulation of trademark morality to be both effective and efficient.

C. POLICY IMPLICATIONS OF ENTRUSTING CONSUMERS WITH
TRADEMARK REGULATION

1. *First Amendment Values in Potentially Scandalous Words*

By invalidating the “immoral or scandalous” provision, the Court in *Iancu v. Brunetti* also has implicitly acknowledged the value in using offensive words. Whether or not to use swear words is a choice—and a meaningful one—that a person makes in not only the *mode* of their expressions but also the *attitude* and *meaning* of their ideas.¹³⁹ Labeling swear words as mere modes of expression trivializes the irreplaceability and uniqueness of swear words as stand-alone expressions of ideas. As Professor Franklyn Haiman pointed out, the message “Fuck the Draft” at issue in *Cohen v. California* cannot be replaced by phrases like “Repeal the Draft” or “Resist the Draft” that just do not carry the same meaning.¹⁴⁰ How people use swear words may be part of their “deep emotional investment in a personal identity” that they use to represent themselves out in the world, to differentiate themselves from others, and to express their feelings and attitudes about others.¹⁴¹

138. See TMEP, *supra* note 49, at § 1203.01.

139. Brief for Cato Institute et al. as Amici Curiae Supporting Respondent at 13, *Iancu v. Brunetti*, 139 S. Ct. 2294 (2019).

140. See Franklyn S. Haiman, *Speech v. Privacy: Is There a Right Not to Be Spoken To?*, 67 NW. U. L. REV. 153, 189 (1972) (referring to *Cohen v. California*, 403 U.S. 15 (1971)).

141. Cato Institute, *supra* note 139, at 13. (quoting TIMOTHY JAY, *WHY WE CURSE: A NEURO-PSYCHO-SOCIAL THEORY OF SPEECH* 82 (1999)).

The Court's decision thus implicitly recognizes that, since people use language as a mode of their thoughts,¹⁴² a ban on certain types of language—including “even the most viscerally offensive words and images that one can imagine”¹⁴³—constitutes a ban on people's thoughts. The Court previously has established that “forbid[ding] particular words . . . run[s] a substantial risk of suppressing ideas in the process.”¹⁴⁴ Such a ban by the government then erodes people's freedom to choose what ideas to express and how to exchange those ideas with others. But the government has no right to “cleanse public debate to the point where it is grammatically palatable to the most squeamish among us.”¹⁴⁵ Without the government's say in determining which ideas are acceptable, people should be able to “bid[] for the minds of men in the market place [sic] of ideas”¹⁴⁶ and to let that marketplace sort out which ideas are going to be accepted as moral or valuable.

It is worth noting that there are many scholars who argue that obscene speech should not be excluded from First Amendment protection.¹⁴⁷ Excluding obscenity from the area of protected speech on the basis that it is “utterly without redeeming social importance”¹⁴⁸ impermissibly allows the government to make value judgments on people's thoughts.¹⁴⁹ Additionally, just as the PTO has struggled to determine what is “immoral or scandalous,” the Court, too, has “struggled to define what is ‘obscene.’”¹⁵⁰ Justice Potter Stewart, for example, had stated that the Court was “faced with the task of trying to define what may be indefinable.”¹⁵¹ He then declared, “I shall not today attempt further to define [it] . . . and perhaps I could never succeed in intelligibly doing so. But I know it when I see it, and [this] motion picture . . . is not that.”¹⁵² Due to such difficulty in coming up with a workable definition, some scholars argue that obscenity should not be a category of unprotected speech.¹⁵³ Nonetheless, as long as obscene speech remains unprotected under

142. Eran Asoulin, *Language as an Instrument of Thought*, GLOSSA (Nov. 16, 2016), [https://www.glossa-journal.org/articles/10.5334/gjgl.34/\[https://perma.cc/NZ7T-PV6J\]](https://www.glossa-journal.org/articles/10.5334/gjgl.34/[https://perma.cc/NZ7T-PV6J]).

143. See *Brunetti*, 139 S. Ct. at 2318 (Sotomayor, J., concurring in part and dissenting in part).

144. *Cohen*, 403 U.S. at 26.

145. *Id.* at 25.

146. *United States v. Rumely*, 345 U.S. 41, 56 (1953) (Douglas, J., concurring).

147. See ERWIN CHEMERINSKY, *CONSTITUTIONAL LAW: PRINCIPLES AND POLICIES* 1110 (6th ed. 2019).

148. *Roth v. United States*, 354 U.S. 476, 484 (1957).

149. See CHEMERINSKY, *supra* note 147, at 1308.

150. *Id.* at 1067.

151. *Jacobellis v. Ohio*, 378 U.S. 184, 197 (1964) (Stewart J., concurring).

152. *Id.*

153. See CHEMERINSKY, *supra* note 147, at 1112.

the First Amendment, the Court must continue to attempt to draw the line between what is obscene and what is not.

As the categories of today's speeches often overlap and thus are not mutually exclusive of one another, this Note agrees with Justice Breyer that the Court should take a more flexible proportionality approach—instead of the strict categorical approach—when reviewing the constitutionality of speech-restricting statutes. He suggested treating the Court's "speech-related categories not as outcome-determinative rules, but instead as rules of thumb" in asking whether the statutory restriction "works speech-related harm that is out of proportion to its justifications."¹⁵⁴ Despite any difficulty in weighing proportions between the speech-related harm and the government's interest, adopting the proportionality analysis may be desirable for future First Amendment cases to allow the Court to conduct review without having to try to squeeze the speech at issue into a particular category. This is especially the case when that speech, considered in its given context, may not fit nicely into one category. For example, "I REALLY DON'T CARE, DO U?" written across the back of a jacket arguably could be viewed as its maker's commercial speech that promotes the brand's theme or campaign. At the same time, it also could be viewed as political speech and even government speech when First Lady, Melania Trump, wears the jacket on a trip to visit detained immigrant children in Texas.¹⁵⁵

Still, to the extent that this Note agrees with Justice Breyer's proportionality approach, it does not agree with his conclusion. In comparing the weight of the speech-related harm posed by the "immoral or scandalous" prohibition against that of its justifications, Justice Breyer concluded that the latter outweighs the former and therefore the statute does not violate the First Amendment.¹⁵⁶ He argued that the speech-related harm here was "[n]ot much," since the provision does not completely shut down the speech but merely denies certain benefits; businesses are free to continue using their marks, even when denied trademark registration.¹⁵⁷ By contrast, the government may have interests in preventing "creation of public spaces that many will find repellant" and "protecting the sensibilities of children."¹⁵⁸ However, those interests do not seem justified when considering both the

154. *Iancu v. Brunetti*, 139 S. Ct. 2294, 2304–05 (2019) (Breyer, J., concurring in part and dissenting in part) (quoting *United States v. Alvarez*, 567 U.S. 709, 730 (2012)).

155. See Katie Rogers, *Melania Trump Wore a Jacket Saying 'I Really Don't Care' on Her Way to Texas Shelters*, NY TIMES (Jun. 21, 2018), <https://www.nytimes.com/2018/06/21/us/politics/melania-trump-jacket.html>.

156. *Brunetti*, 139 S. Ct. at 2305–06.

157. *Id.* at 2306.

158. *Id.* at 2307.

government's record of inconsistency in reviewing similarly offensive marks and the lack of such interests in other government registration systems, e.g., for copyright and patent. By contrast, there are substantial speech-related harms caused by the trademark prohibition, e.g., restriction on freedom of speech and unequal treatment of trademark applicants. Therefore, even under Justice Breyer's proportionality approach, the "immoral or scandalous" prohibition is still unconstitutional, and thus invalid.

Hence, leaving the trademark regulation to consumers prevents the government from discriminating against those who use offensive words to express their thoughts and identities. As the opinions of both the majority in *Matal v. Tam* and Justice Breyer in *Iancu v. Brunetti* indicate, trademarks have "an expressive component in addition to a commercial one . . ." ¹⁵⁹ Eliminating such discrimination places offensive words on the same legal ground as inoffensive words, allowing consumers to associate more freely and easily with certain marks and expressions of ideas.

The decision to cease the "immoral or scandalous" ban also recognizes the reality that people may use offensive marks to pursue more than mere financial values. Justice Breyer argued that swear words are "attention-grabbing words . . . financially valuable to some businesses that seek to attract interest in their products."¹⁶⁰ What he may have overlooked, however, is that those who use potentially offensive words, especially in commerce, may be doing so with the full awareness that such usage comes with risks—that the public attention they are about to receive by cursing may not be so financially beneficial. When a clothing brand chooses FCUK as its name, for instance, it has embraced "the risqué element, the apparent avoidance through mis-ordering, the in-your-face-ness of *fuck*, which they appear to sidestep but in fact highlight."¹⁶¹ The government should stand clear of mark owners' path to the market and their right to express their ideas and identities in the form of trademarks. The true value in using an offensive mark, then, will be determined by consumers who will see and react to it.

2. *Predictability and Equal Treatment in the Trademark Registration Process*

The Court's invalidation of the "immoral or scandalous" prohibition also promotes both predictability of rules and equal treatment of applicants in the government's trademark registration system. In support of denying FUCT's trademark registration, the dissents argued that the government has an interest

159. *Id.* at 2305.

160. *Id.* at 2307.

161. Cato Institute, *supra* note 139, at 13 (quoting RUTH WAJNRYB, EXPLETIVE DELETED: A GOOD LOOK AT BAD LANGUAGE 187 (2005)).

in not being “associated with” highly vulgar and offensive words and in seeking to “disincentivize the use” of such words.¹⁶² Justice Breyer pondered, “Just think about how you might react if you saw someone wearing a t-shirt or using a product emblazoned with an odious racial epithet.”¹⁶³ Nevertheless, as the above REDSKINS example illustrates, the government has undercut such claimed interest by approving trademark registration for words like REDSKINS. The Washington football fans wearing jerseys and waving flags emblazoned with the REDSKINS trademark are precisely what Native Americans and their children have been seeing for the last half a century. From the public’s perspective, it may be difficult to discern the justification for the government’s interest when that interest has been clouded by the government’s inconsistent treatment of potentially offensive marks.

It may be true, as Chief Justice Roberts asserted, that the government is not required “to give aid and comfort” to those using profane or vulgar words.¹⁶⁴ However, a problem arises when the government gives aid and comfort to some but not others who use words that “do look remarkably similar.”¹⁶⁵ For example, when the government approves BITCH but rejects KICKABITCH, the government cannot be said to be pursuing its interest in disincentivizing the use of vulgar words; rather, it is arbitrarily discriminating between two similarly vulgar words.

Indeed, in the eyes of trademark applicants who seek statutory benefits and rightfully expect equal review of their marks in the registration process, the government’s determination at the end of that process has been as unpredictable as a “flip of a coin.”¹⁶⁶ Even if what trademark registration provides is “merely . . . certain additional benefits,” the government should give those benefits equally to mark owners and let them compete fairly—against one another and also against counterfeiters.¹⁶⁷ In furtherance of the purpose of trademark law, the government’s job is to armor trademark owners—including owners of FCUK and FUCT alike—with those statutory benefits and enable them to fight off confusing and deceptive knockoffs more effectively. By treating all words equally, regardless of their offensiveness or vulgarity, the government would strengthen its stance that its act of “registration does not constitute approval of a mark.”¹⁶⁸ In sum, the government’s equal treatment of marks in its trademark registration process

162. *See Brunetti*, 139 S. Ct. at 2307.

163. *Id.*

164. *Id.* at 2303–04 (Roberts, C.J., concurring in part and dissenting in part).

165. Transcript of Oral Argument, *supra* note 15, at 19.

166. *Id.* at 21.

167. *Brunetti*, 139 S. Ct. at 2303.

168. *Matal v. Tam*, 137 S. Ct. 1744, 1759 (2017).

would put trademark applicants on clearer notice of what is registrable and allow trademark owners to compete fairly with one another and against counterfeiters.

3. *Adherence to Trademark Law's Intended Purpose*

Lastly, leaving the task of regulating trademark morality to consumers allows trademark law to better adhere to its original purpose. The purpose of trademark law is to help consumers identify desired goods quickly and accurately by reducing any confusion or deception in their commercial transactions. Considering offensiveness or scandalousness of certain marks, and denying registration on such grounds, interfere with the goal of reducing trademark confusion and deception. By refusing to register FUCT, the government undermines the strength of trademark law in helping to sort out counterfeit sellers, whose products are designed to confuse and deceive consumers.

In similar registration systems such as those for copyright and patent, the government does not regulate morality as it is not relevant to the purpose of the underlying laws, i.e., to “promote the Progress of Science and useful Arts.”¹⁶⁹ Consequently, the general public is left to evaluate the morality of copyrighted or patented works and to make subsequent determinations on its own. Although trademark law is inherently different from copyright and patent laws in that the former is rooted in the Commerce Clause, regulating morality is still not part of the purpose of trademark law. Unless and until the government can provide a legitimate justification for its interest in regulating morality in trademarked words but not in copyrighted works or patented inventions, it remains questionable as to why only trademark law should diverge from its intended purpose to regulate morality.

In particular, unlike in trademark law, the government does not impose any ban on registering offensive or shocking materials under copyright or patent laws. For example, Erik Brunetti was able to obtain copyright registration for his book, titled *FUCT*, which features on its front cover a hat with the FUCT logo (see Figure 2).¹⁷⁰ Moreover, it is not difficult for the general public to find books with even more blunt titles: *On Bullshit* (2005), *Shit My Dad Says* (2009), and *Go the Fuck to Sleep* (2011), are some of the copyright-protected books that are on the *New York Times* best-seller list.¹⁷¹ Similarly, three top songs on the *Billboard* chart in March 2011 were “Fuck You!”

169. U.S. Const., Art. I, § 8, cl. 8.

170. U.S. Copyright No. TX0008351677 (issued July 6, 2016).

171. Cato Institute, *supra* note 139, at 18 (quoting MELISSA MOHR, HOLY SH*T: A BRIEF HISTORY OF SWEARING 247 (2013)).

“Tonight (I’m Fucking You),” and “Fuckin’ Perfect.”¹⁷² Then there is “Real N**** Roll Call,” which is not only registered with the U.S. Copyright Office¹⁷³ but also a winner of the Guinness World Record for having the most swear words in one song—295 to be exact.¹⁷⁴

Figure 2: Front Cover of the Book *FUCT* by Erik Brunetti



172. *Id.* (quoting MOHR, *supra* note 171, at 246).

173. U.S. Copyright No. V3529D821 (issued Oct. 28, 2005) (censorship added).

174. Kate Hutchinson, *The List: A Brief History of Swearing in Music*, REDBULL (July 22, 2013), <https://www.redbull.com/int-en/a-brief-history-of-swearing-in-music>.

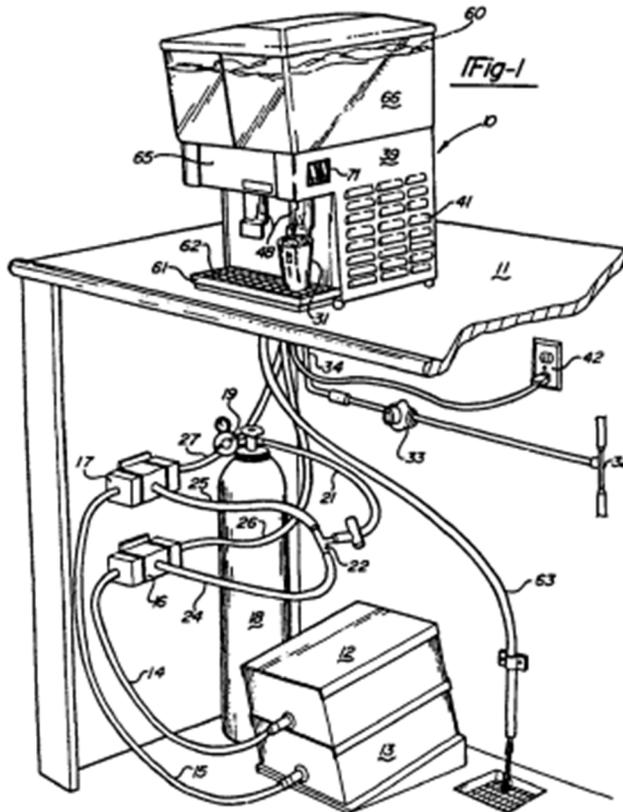
The government would not deny copyright registration for a song entitled “BITCH” while approving one entitled “KICKABITCH,” and thus any interest it asserts to vest in making such a distinction between vulgar words in trademark law does not hold much water. Moreover, that copyright and trademark laws exist to serve different purposes, e.g., that prohibiting curse words in copyrightable works may work against promoting creativity, still does not save the government’s justification. In practice, the general public—including children whom the government claims to seek to “protect” from exposure to offensive trademarks—is not likely to appreciate the different purposes that copyright law and trademark law serves when it sees a shirt and a book emblazoned with the same FUCT logo. Indeed, the public’s inability to see such distinction is underscored by the fact that copyrighted works and trademarked words overlap today.

Likewise, the government does not regulate fraud in patent registration, a criterion that carries little relevance to the underlying purpose of patent law. Consider the invention shown below in Figure 3: a juice dispensing system that *appears* to blend fresh juice in a mixer on top but actually dispenses its juice from a tank hidden underneath.¹⁷⁵ The Federal Circuit affirmed the validity of this invention’s patent registration on the grounds that this invention, although imitating freshness to defraud the public, still had *utility* within the meaning of the patent statute.¹⁷⁶ As the PTO does not regulate fraud in patent registrations, the public is left to evaluate that factor on its own. Similarly, leaving it up to the public to evaluate scandalousness or offensiveness in trademarks would allow trademark law to better adhere to its intended purpose and not diverge away from that purpose.

175. See *Juicy Whip, Inc. v. Orange Bang, Inc.*, 185 F.3d 1364 (Fed. Cir. 1999).

176. *Id.*; see also 35 U.S.C. § 101 (2012) (“Whoever 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”) (emphasis added).

Figure 3: U.S. Patent No. 5,575,405



These examples also weaken the *Brunetti* dissents' argument in support of refusing registration of marks like FUCT on the ground that the government has "a reasonable interest" in ensuring that it is not involved in "promoting highly vulgar or obscene speech, and that it [is not] associated with such speech."¹⁷⁷ Although the dissents conceded that trademarks are private speech and not government speech," they contended that, when the government registers a mark, it is "necessarily 'involv[ed] in promoting' that mark";¹⁷⁸ upon registering a mark, the government is required to publish the mark "as well as to take steps to combat international infringement."¹⁷⁹ However, that interest is not reasonable as it clashes with the fact that the government regularly registers copyrights in creative works that are obscene, vulgar, or profane.

177. *Iancu v. Brunetti*, 139 S. Ct. 2294, 2307 (2019) (Roberts, C.J., concurring in part and dissenting in part).

178. *Id.* at 2307 (quoting Sotomayor, J., *Brunetti*, 139 S. Ct. at 2317–18 (concurring in part and dissenting in part)).

179. *Id.* at 2317 (Sotomayor, J., concurring in part and dissenting in part).

Furthermore, the government should not worry about being perceived as supporting trademarks it registers, since the Court in *Matal v. Tam* has already assured that there is no evidence that the public “associates the contents of trademarks with the Federal Government.”¹⁸⁰ If there were any evidence of such association, there seems to be no reason the same public should think differently about copyrighted works and patented inventions that the government also registers. Under that scenario, the government’s concern then should not be uniquely focused on trademark registration. Thus, denying trademark registration—on the basis that the government has an interest in not promoting certain kinds of speech—not only undermines the immediate trademark registration system, but also constitutes “a huge and dangerous extension” of such restriction as “other systems of government registration could easily be characterized in the same way.”¹⁸¹

No government-created test can substitute, let alone outperform, consumers’ own moral standards. Just as it is up to the general public to evaluate the morality of uncensored explicit songs and deceptive juice dispensers, it should be up to the same general public to evaluate the morality of offensive brand names. That way, trademark morality can be more effectively and efficiently tested, and the government can avoid regulating beyond the realm originally intended for trademark law.

V. CONCLUSION

The majority in *Iancu v. Brunetti* made the right decision in invalidating the “immoral or scandalous” provision of the Lanham Act, and in deciding not to apply a narrowing construction to the “scandalous” prong. Accordingly, Congress should not reintroduce questions of “scandalousness” by rewriting the statute. Instead, it should leave the task of sorting out such matters to consumers who have been the de facto regulators of trademark morality thus far.

The government does not need to worry when entrusting consumers with this task, since consumers are well-equipped with their power to purchase products bearing trademarks they favor as well as the power to boycott products bearing trademarks they find highly offensive or scandalous. A government restriction on “immoral or scandalous” marks has not determined the ultimate fate and success of trademarks in the marketplace, and no new replacement statute will.

180. *Matal v. Tam*, 137 S. Ct. 1744, 1759–60 (2017).

181. *Id.* at 1760; *see Brunetti*, 139 S. Ct. at 2317.

Congress nevertheless may be triggered to replace the statute if, and as Justice Sotomayor predicts, people seek to take advantage of the *Iancu v. Brunetti* decision and flood the PTO with trademark applications for the “most viscerally offensive words” possible.¹⁸² Nonetheless, empirical evidence has proved otherwise. A study that focused on the change in the volume of trademark registration applications for such words following the *In re Brunetti* decision (that the Supreme Court subsequently affirmed in *Iancu v. Brunetti*) indicates “a future not much altered, despite early concerns to the contrary.”¹⁸³ The absence of “a flood behind the anticipated floodgates” may suggest that the actual commercial value of highly offensive words is not as substantial as may have been expected.¹⁸⁴ Provided that this trend in volume follows the *Iancu v. Brunetti* decision, it may alleviate the concern for commercial abuse of curse words as mere attention-grabbing modes of expression.

Regardless of any effects felt at the PTO, the *Iancu v. Brunetti* decision was an important move for upholding First Amendment values in free speech, for promoting equal treatment of trademark applicants, and for bringing trademark law closer to its intended purpose. More importantly, the decision has allowed people to flaunt the scarlet letters of their choice as registered trademarks, and it is up to consumers to decide whether those letters are unworthy of acceptance in the marketplace. With the PTO out of the picture, consumers now stand as the sole regulators of trademark morality—and should remain so.

182. *See Brunetti*, 139 S. Ct. at 2318 (Sotomayor, J., concurring in part and dissenting in part).

183. Vicenc Feliu, *The F Word – An Early Empirical Study of Trademark Registration of Scandalous and Immoral Marks in the Aftermath of the In re Brunetti Decision*, 18 JOHN MARSHALL REV. INTELL. PROP. L. 404, 404 (2019).

184. *Id.* at 418.

THE GDPR: A RETROSPECTIVE AND PROSPECTIVE LOOK AT THE FIRST TWO YEARS

Erin Hilliard

I. INTRODUCTION

“The Clock Is Ticking: Is Your Company Ready for GDPR?”¹

“Marriott Faces Massive \$123 Million GDPR Fine For 2018 Security Breach”²

“General Data Protection Regulation (GDPR): What you need to know to stay compliant”³

For three years, these four letters, G-D-P-R, have been making headlines around the world. In May 2018, Google searches for the GDPR were more popular than searches for Beyoncé.⁴ Consumers have received countless notifications from companies about their data collection practices. The new E.U. General Data Protection Regulation (GDPR or “Regulation”) has made a significant impact on companies of all sizes and industries—but why?

The right to privacy, the protection of one’s private and family life, home, and correspondence, has a longstanding history in Europe. Its origins can be

DOI: <https://doi.org/10.15779/Z384J09Z3K>

© 2020 Erin Hilliard.

† J.D., 2021, University of California, Berkeley, School of Law. I would like to recognize the people who have been instrumental in the successful completion of this publication. I would like to express my sincere gratitude to Professor Chris Jay Hoofnagle for his valuable feedback and suggestions throughout the publication process. I would also like to acknowledge Professor Kenneth A. Bamberger and BCLT Executive Director Jim Dempsey for their critical assessments and encouragement during the drafting process. A huge thank you also goes out to Dr. Su Li for her statistical expertise.

1. *The Clock Is Ticking: Is Your Company Ready For GDPR?*, THE ONE BRIEF, <https://theonebrief.com/the-clock-is-ticking-is-your-company-ready-for-gdpr/> (last visited May 18, 2020).

2. Nicole Lindsey, *Marriott Faces Massive \$123 Million GDPR Fine For 2018 Security Breach*, CPO MAGAZINE (July 23, 2019), <https://www.cpomagazine.com/data-protection/marriott-faces-massive-123-million-gdpr-fine-for-2018-security-breach/>.

3. Michael Nadeau, *General Data Protection Regulation (GDPR): What You Need to Know to Stay Compliant*, CSO ONLINE (May 29, 2019), <https://www.csoonline.com/article/3202771/general-data-protection-regulation-gdpr-requirements-deadlines-and-facts.html>.

4. Rachel Thompson, *GDPR is Currently Searched for More Than Beyoncé on Google*, MASHABLE (May 23, 2018), <https://mashable.com/2018/05/23/google-trends-gdpr-beyonce/>.

traced as far back as the European Convention on Human Rights,⁵ enacted in 1953.⁶ The right to data protection is recognized in Article 8 of the E.U. Charter of Fundamental Rights.⁷ The E.U. Court of Justice often conflates the two rights, treating data protection as a subset of the right to privacy,⁸ while some academics argue a distinction between the two is necessary.⁹ Regardless of the specific terminology used, Europe's privacy protections are derived from these two rights and the GDPR is the result of a tremendous amount of iterative legislation on the part of the European Union and its member states to ensure the fundamental rights and freedoms of its citizens remain protected.¹⁰

Prior to proposing the GDPR, lawmakers recognized that technology had advanced exponentially since the Data Protection Directive—the data protection legislation preceding the GDPR—was written in 1995, and they were concerned the Directive no longer provided E.U. citizens with adequate protection.¹¹ Lawmakers also wanted to make data protection practices across the European Union more harmonious and consistent.¹² Now, with the GDPR in effect, hefty penalties of millions of euros,¹³ a widened territorial scope of E.U. data protection law,¹⁴ and compliance demands from vendors and partners have successfully motivated many companies into compliance.

5. Peter Hustinx, *EU Data Protection Law: The Review of Directive 95/46/EC and the Proposed General Data Protection Regulation*, EUROPEAN DATA PROTECTION SUPERVISOR 3 (Sept. 15, 2014), https://edps.europa.eu/data-protection/our-work/publications/speeches-articles/eu-data-protection-law-review-directive_en.

6. *Details of Treaty No.005*, COUNCIL OF EUROPE, <https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/005> (last visited May 12, 2020).

7. Charter of Fundamental Rights of the European Union, art. 8, 2000 O.J. (C 364) and art. 8, 2010 O.J. (C 83).

8. Orla Lynskey, *Deconstructing Data Protection: The 'Added-value' of a Right to Data Protection in the EU Legal Order*, 36 INT'L & COMPARATIVE L.Q. 569, 597 (2014).

9. See, e.g., *id.* (arguing a judicial recognition of the distinction is necessary because data protection provides individuals with more rights over more types of data than the right to privacy).

10. See Convention 108 and Protocols, COUNCIL OF EUROPE, <https://www.coe.int/en/web/data-protection/convention108-and-protocol> (last visited May 12, 2020); Hustinx, *supra* note 5, at 2, 4, 9, 29.

11. *Id.* at 26–27.

12. *Id.*

13. *Id.* at 33.

14. E.U. data protection law has expanded under the Regulation to encompass not only companies based in the European Union, but also companies who target or monitor the behavior of E.U. citizens. *Guidelines 3/2018 on the Territorial Scope of the GDPR (Article 3)—Version for Public Consultation*, EUROPEAN DATA PROTECTION BOARD (Nov. 16, 2018), https://edpb.europa.eu/sites/edpb/files/consultation/edpb_guidelines_3_2018_territorial_scope_en.pdf.

Privacy professionals are enjoying phenomenal job security. But has the GDPR achieved its goals? What did enforcement look like for the first two years?

The GDPR is an enormous and comprehensive regulation that affects thirty countries, each with its own unique culture, objectives, and needs. It took four years of negotiations for all parties subject to the Regulation to finally agree on a draft.¹⁵ From the very beginning, the Article 29 Working Party (now the European Data Protection Board), which was charged with providing guidance on GDPR implementation, acknowledged that consistent implementation of the GDPR would require national supervisory authorities to work together on “sub-national, national and cross-border levels.”¹⁶

In the first year of implementation, over 144,000 individual complaints and more than 89,000 data breach notifications were reported.¹⁷ Countries have worked diligently to restructure and operationalize their methods for responding to and investigating E.U. citizens’ complaints and company data breach notifications. Companies continue to navigate GDPR requirements, which were completely foreign to many organizations based outside of Europe. They seek guidance from European data protection authorities and E.U. institutions and advisory bodies regarding the interpretation and implementation of specific articles. While complete harmonization of data protection practices across thirty countries (twenty-seven E.U. member states and three European Economic Area (EEA) countries) may not be realistic, some of the early fragmentation in enforcement will become inconsequential with time.

Many practitioners and scholars have been looking to the number of imposed administrative fines across the European Union as a measure of overall “success” of the GDPR. Counting fines is an empirical way to track enforcement across countries. One way to assess the consistency of GDPR enforcement, a key goal of the new regulation, is to ask when, where, and why administrative fines were imposed during the first two years of GDPR enforcement. At this point in time, there are not any known academic

15. See *EU General Data Protection Regulation - Background*, DLA PIPER, <https://www.dlapiper.com/en/portugal/focus/eu-data-protection-regulation/background/> (last visited May 12, 2020).

16. *Guidelines on the Application and Setting of Administrative Fines for the Purposes of the Regulation 2016/679*, Article 29 Data Protection Working Party, 17 (Oct. 3, 2017) [hereinafter Article 29 Data Protection Working Party].

17. *GDPR One Year Anniversary—Infographic*, INTERNATIONAL ASSOCIATION OF PRIVACY PROFESSIONALS, <https://iapp.org/resources/article/gdpr-one-year-anniversary-infographic/> (last visited May 12, 2020).

resources that have comprehensively tracked all fines across each E.U. member state and provided an empirical analysis of the data.

This Note addresses that gap in the literature. We have done an exhaustive search to identify all GDPR fines and nonmonetary sanctions—imposed or pending enforcement—across the E.U. member states, three EEA countries, and the United Kingdom for the time period May 25, 2018, the date of enactment, through March 31, 2020. (The United Kingdom withdrew from the European Union in January 2020,¹⁸ but remained subject to the GDPR until December 2020,¹⁹ and is therefore included in all analyses.) Our search yielded a total of 311 enforced fines and forty-nine pending actions across the thirty-one countries²⁰ where the GDPR was enforced during this time period. In this Note, we use descriptive statistics (mean, median, mode, range, and linear regression) to analyze the data and present our results, including noteworthy findings, patterns, and trends of enforcement since implementation.

The number of administrative fines imposed is increasing rapidly, and the median value of these fines is also increasing. In the first six months after the GDPR was enacted, fines were imposed in smaller numbers than expected. This may be because many underestimated how each country's micro-decisions regarding national privacy legislation and GDPR implementation would initially impact the harmonization of GDPR enforcement practices across the European Union. Two behaviors played a role: first, it was difficult for a country to impose fines for GDPR violations without having first passed national privacy legislation that includes the nation's selected GDPR derogations (discretionary articles), and second, countries with education or warning-first implementation approaches fined much less than countries who took active enforcement stances. However, the impact of these micro-decisions on fine consistency will decrease with time.

Looking at the 311 fines imposed in the first two years of enforcement, the total number and median fine value increased from 2018 to 2019. The number of fines issued increased nearly eightfold from 2018 (twenty-five fines) to 2019 (193 fines). In the first quarter of 2020 alone, ninety-three fines were issued. The median fine in 2018 was €3,200, while the median fine in 2019 was

18. The data set includes enforcement actions brought by the United Kingdom until it withdrew from the European Union in January 2020. *See generally* *Brexit: All You Need to Know About the UK Leaving the EU*, BBC (Feb. 17, 2020), <https://www.bbc.com/news/uk-politics-32810887>.

19. *Information Rights and Brexit Frequently Asked Questions*, INFORMATION COMMISSIONER'S OFFICE (May 20, 2020), <https://ico.org.uk/for-organisations/data-protection-and-brexit/information-rights-and-brexit-frequently-asked-questions/>.

20. The data set includes enforcement actions brought by the United Kingdom. *Id.*

€11,380, demonstrating that fine values are increasing but small fines are still more prevalent than massive fines. The median fine in the first quarter of 2020 (accounting for the first three months of the year) was €6,670, more than double the median fine in the seven months of 2018 following enactment.

This Note proceeds as follows. Part II explains the motivations behind the GDPR and provides an overview of key aspects of the GDPR that are crucial to assessing the enforcement actions that have occurred. Part III analyzes each country's enforcement behavior and GDPR implementation approach. Part IV presents the data, including a breakdown of fines by country and violation type, a comparison of total fines by country to gross domestic product, and a discussion of the industries most affected by enforcement actions. Part V discusses anticipated enforcement trends based on the data and findings presented in Parts III and IV.

II. BACKGROUND

A. WHAT IS THE GDPR?

The GDPR became enforceable on May 25, 2018.²¹ Thirty countries—all twenty-seven E.U. member states and three EEA countries—are subject to the GDPR.²² The protection of personal data is a fundamental right of all E.U. citizens, and the GDPR serves to enhance and unify data protection processes across Europe.²³ The GDPR defines personal data as “any information relating to an identified or identifiable natural person.”²⁴ These identifiable, natural persons are referred to as “data subjects” throughout the Regulation.²⁵

The GDPR has a laser focus on “[t]he protection of natural persons in relation to the processing of personal data.”²⁶ Data “processing” is viewed as any operation(s) “performed on personal data [. . .] whether or not by automated means.”²⁷ Unlike the GDPR's predecessor, the Data Protection

21. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Apr. 27, 2016), art. 99(2), 2016 O.J. (L 119), <https://eur-lex.europa.eu/eli/reg/2016/679/oj> [hereinafter GDPR].

22. *GDPR National Implementation Legislation Tracker*, THOMSON REUTERS PRACTICAL LAW UK (Sept. 20, 2019), <https://uk.practicallaw.thomsonreuters.com/w-013-1949> [hereinafter THOMSON REUTERS PRACTICAL LAW UK].

23. Charter of Fundamental Rights of the European Union (Oct. 26, 2012), art. 8(1), 2012 O.J. (C 326).

24. GDPR, *supra* note 21, art. 4(1).

25. *Id.*

26. *Id.* Recital 1; *see also id.* art. 94.

27. *Id.* art. 4(2).

Directive, the GDPR is a regulation that is wholly enforceable as law in each E.U. member state.²⁸ The Data Protection Directive placed obligations on individual member states to implement data protection laws locally.²⁹ In contrast, the GDPR provides one set of data protection rules that must be followed in each member state, with the exception of a number of derogations where articles can be tailored to, and supplemented by, national law.³⁰

B. INTENT OF THE GDPR

The European Commission was particularly concerned that the lack of consistency regarding data protection standards across E.U. member states would hamper economic development.³¹ In 1990, the Commission submitted a proposal for a directive that would create an E.U.-wide standard for personal data protection, making it easier for businesses to operate.³² After four years of negotiations, the 1995 Data Protection Directive was adopted.³³ The overarching goals of the Directive were twofold: (1) to protect E.U. citizens' rights to privacy concerning the processing of personal information, and (2) to encourage development of the E.U. market by enabling the free flow of data.³⁴ The Directive led to more consistent data protection standards across countries than had previously existed.³⁵

Over time, however, it became clear that the Directive needed to be updated. First and foremost, technology had evolved tremendously since the Directive became enforceable in 1998.³⁶ When the Directive was implemented, the internet was just taking off. Today, advanced technologies are a part of everyday life, and online monitoring and data collection are commonplace. The technological age presents new challenges that require more effective protections.³⁷ Second, implementation of the Directive varied greatly across

28. European Union, *Regulations, Directives and Other Acts*, OFFICIAL WEBSITE OF THE EUROPEAN UNION, https://europa.eu/european-union/eu-law/legal-acts_en [hereinafter European Union] (last visited May 12, 2020).

29. *Id.*

30. GDPR, *supra* note 21, Recital 10.

31. Hustinx, *supra* note 5, at 9.

32. *Id.*

33. *Id.*

34. *Id.*

35. *Id.*

36. *First Report on the Implementation of the Data Protection Directive (95/46/EC)*, COMMISSION OF THE EUROPEAN COMMUNITIES (May 15, 2003), <https://op.europa.eu/en/publication-detail/-/publication/ff783aa5-5770-42e8-bac3-917fe0a361d7/language-en> [hereinafter COMMISSION OF THE EUROPEAN COMMUNITIES].

37. *Id.* at 26.

the then twenty-eight³⁸ E.U. member states.³⁹ Because the Directive had to be transposed into the national law of each country, there were discrepancies in interpretation, application, and implementation across the member states.⁴⁰ After six years of planning and negotiations, the GDPR was adopted in 2016, replacing the Directive that had been law for over two decades.⁴¹

While the GDPR's steep administrative fines have garnered a lot of public attention, the success of the GDPR largely depends on the Regulation's ability to adapt to new technologies. To secure compliance, lawmakers drafted laws they hoped would be flexible enough to stay relevant as technological innovation continues at a rampant pace.⁴² The introduction of larger fines was just one of many key changes that occurred in data protection law with the implementation of the GDPR. Table 1 provides an overview of the key attributes of the previous law, the Data Protection Directive, and the current law, the GDPR.

38. The United Kingdom withdrew from the EU in January 2020. *See generally* *Brexit: All You Need to Know About the UK Leaving the EU*, BBC (Feb. 17, 2020), <https://www.bbc.com/news/uk-politics-32810887>.

39. *See* Katie McMullan, *Legislative Framework*, in *European Data Protection: Law and Practice*, 49, 52 (Eduardo Ustaran ed., 2018).

40. Hustinx, *supra* note 5, at 9, 24–25.

41. *The History of the General Data Protection Regulation*, EUROPEAN DATA PROTECTION SUPERVISOR, https://edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en (last visited May 12, 2020).

42. GDPR, *supra* note 21, Recitals 6–7.

Table 1: Key Attributes of the Data Protection Directive and the GDPR

Data Protection Directive	General Data Protection Regulation
Date Adopted: October 24, 1995 ⁴³	Date Adopted: April 27, 2016 ⁴⁴
Year Enforceable: 1998 ⁴⁵	Date Enforceable: May 25, 2018 ⁴⁶
Geographical Scope: Data processors established in the territory, or using equipment established in the territory, of a country subject to the Directive ⁴⁷	Geographical Scope: ⁴⁸ <ul style="list-style-type: none"> • Companies established in the EU • Companies established outside the EU, who offer goods or services to the EU, or monitor individuals in the EU
Implementation Process: Transposed into the national law of each country ⁴⁹	Implementation Process: Enforceable as law in all EU member states ⁵⁰
Goals: ⁵¹ <ul style="list-style-type: none"> • Protection of fundamental privacy rights • Development of the EU market • Harmonization of data protection practices across countries 	Goals: ⁵² <ul style="list-style-type: none"> • Update data protection law to account for technological innovation • Harmonization of data protection practices across countries

43. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995 O.J. (L 281), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31995L0046> [hereinafter Directive 95/46/EC] (last visited May 12, 2020).

44. GDPR, *supra* note 21.

45. COMMISSION OF THE EUROPEAN COMMUNITIES, *supra* note 36, at 19.

46. GDPR, *supra* note 21, art. 99.

47. Directive 95/46/EC, art. 4.

48. GDPR, *supra* note 21, arts. 3(1)–(3).

49. Directive 95/46/EC, art. 4.

50. GDPR, *supra* note 21, Recital 10.

51. Hustinx, *supra* note 5, at 9.

52. *Id.* at 26–27.

Table 1 (continued): Key Attributes of the Data Protection Directive and the GDPR

New Requirements Under the GDPR ⁵³
<ul style="list-style-type: none"> • Now applies to all companies engaging with the EU market • Data subject right to Data Portability • Consent and explicit consent require affirmative action • 72 hours to notify regulator of a data breach • Appointment of Data Protection Officers in companies • Much larger, standardized administrative fines

C. DATA PROTECTION AUTHORITIES

Data protection authorities (DPAs), also referred to as supervisory authorities, are key supervisors and enforcers of the GPDR. DPAs are independent regulators⁵⁴ who are designated by member states to monitor the implementation of the GDPR and enforce penalties. Article 58(1) grants DPAs three types of power: investigatory, corrective, and advisory. Their investigatory powers include the ability to access all evidence necessary to fulfill their responsibilities and start investigations.⁵⁵ DPA corrective powers are vast and range from issuing reprimands to imposing administrative fines; they can even ban an organization's data processing activities.⁵⁶ DPA advisory powers concern advising companies when consulted, and the ability to issue, adopt, and approve codes of conduct and certifications.⁵⁷ DPAs use their powers to fulfill their tasks (enumerated in Article 57), which include handling complaints, carrying out investigations, promoting awareness of data protection, and cooperating with other nations to ensure consistent application of the Regulation.⁵⁸ Some countries have one DPA while other countries have many. When a country has more than one DPA, as is the case in Germany, that member state must ensure consistency across all of its national DPAs.⁵⁹ Although member states had already appointed DPAs under the Data

53. *EU General Data Protection Regulation—Key Changes*, DLA PIPER, <https://www.dlapiper.com/en/portugal/focus/eu-data-protection-regulation/key-changes/#wider%20territorial%20scope> (last visited May 17, 2020).

54. GDPR, *supra* note 21, art. 52(1).

55. *Id.* art. 58(1).

56. *Id.* art. 58(2).

57. *Id.* art. 58(3).

58. *Id.* art. 57.

59. *Id.* art. 51(3).

Protection Directive,⁶⁰ DPAs now have significantly larger caseloads as a result of the GDPR's broader territorial scope and more serious penalties.

Although this Note focuses heavily on the role DPAs play in enforcement, specifically their power to issue administrative fines, it is important to note that DPAs are not the only enforcement mechanism available under the Regulation. Individuals can pursue legal action in accordance with their country's national laws.⁶¹ Industry self-regulation is another enforcement tool the GDPR advances through the data processing accountability requirement (Table 2), the mandated appointment of Data Protection Officers within organizations,⁶² and the creation of codes of conduct and data protection certification processes.⁶³

D. GDPR DATA PROCESSING PRINCIPLES

Article 5(1) of the GDPR sets out seven data processing principles that inform the purpose and intent of the legislation (Table 2). These principles are similar to the principles of the Data Protection Directive, with the exception of the Accountability principle, which is new under the GDPR.⁶⁴ These seven data processing principles “embody the spirit” of the GDPR.⁶⁵ The highest tier of administrative fines can be imposed for violating these principles.⁶⁶ Fines are administered by member state DPAs⁶⁷ (or the European Data Protection Supervisor, if the suspected GDPR violator is an E.U. institution)⁶⁸ using both the text of the Regulation and interpretation guidance from E.U. advisory bodies.⁶⁹ Below is a brief explanation of each guiding principle. Recitals, while not binding, are used in legal documents to explain the reasoning behind certain terms and decisions. Recital 39 provides important context for how to interpret and follow the seven Article 5(1) data processing principles.

60. Detlev Gabel and Tim Hickman, *Chapter 14: Data Protection Authorities—Unlocking the EU General Data Protection Regulation*, WHITE & CASE (Apr. 5, 2019), <https://www.whitecase.com/publications/article/chapter-14-data-protection-authorities-unlocking-eu-general-data-protection>.

61. GDPR, *supra* note 21, art. 79.

62. *Id.* art. 37.

63. *Id.* arts. 40–43.

64. *The Principles*, INFORMATION COMMISSIONER'S OFFICE, <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/> (last visited May 12, 2020).

65. *Id.*

66. GDPR, *supra* note 21, art. 83(5)(a).

67. *Id.* art. 83(1).

68. *Complaints*, EUROPEAN DATA PROTECTION SUPERVISOR, https://edps.europa.eu/node/75_en (last visited May 1, 2020).

69. *See* Article 29 Data Protection Working Party, *supra* note 16.

Table 2: Data Processing Principles

Article 5(1) ⁷⁰	Recital 39 Explanation ⁷¹
(a) “[L]awfulness, [F]airness and [T]ransparency”	<p>“Any processing of personal data should be lawful and fair.”</p> <p>“It should be transparent to natural persons that personal data concerning them are collected, used, consulted or otherwise processed and to what extent the personal data are or will be processed.”</p> <p>“Natural persons should be made aware of risks, rules, safeguards and rights in relation to the processing of personal data and how to exercise their rights in relation to such processing.”</p>
(b) Purpose Limitation	<p>“[T]he specific purposes for which personal data are processed should be explicit and legitimate and determined at the time of the collection of the personal data.”</p>
(c) Data Minimization	<p>“The personal data should be adequate, relevant and limited to what is necessary for the purposes for which they are processed.”</p>
(d) Accuracy	<p>“Every reasonable step should be taken to ensure that personal data which are inaccurate are rectified or deleted.”</p>

70. GDPR, *supra* note 21, art. 5.

71. *Id.* Recital 39.

Table 2 (continued): Data Processing Principles

Article 5(1)	Recital 39 Explanation
(e) Storage Limitation	<p>“[It must be ensured] that the period for which the personal data are stored is limited to a strict minimum.”</p> <p>“In order to ensure that the personal data are not kept longer than necessary, time limits should be established by the controller for erasure or for a periodic review.”</p>
(f) Integrity and Confidentiality	<p>“Personal data should be processed in a manner that ensures appropriate security and confidentiality of the personal data, including for preventing unauthorised access to or use of personal data and the equipment used for the processing.”</p>
Article 5(2) ⁷²	Article Text
Accountability	<p>“The controller⁷³ shall be responsible for, and be able to demonstrate compliance with, paragraph 1.”</p>

E. DATA SUBJECT RIGHTS

Chapter three of the GDPR enumerates seven rights that data subjects can exercise under the Regulation. These rights are listed in Table 3. The largest possible fines can also be imposed for infringement of these rights.⁷⁴

72. GDPR, *supra* note 21, art. 5.

73. “[C]ontroller’ means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data.” GDPR, *supra* note 21, art. 4(7).

74. *Id.* art. 83.

Table 3: Data Subject Rights

Article 15 ⁷⁵	Recital 63 Explanation ⁷⁶
Right of Access	“A data subject should have the right of access to personal data which have been collected concerning him or her.”
Article 16 ⁷⁷	Article Text
Right to Rectification	The right to “obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her[.]” including the right to complete personal data that is “incomplete.”
Article 17 ⁷⁸	Recital 65 Explanation ⁷⁹
Right to Erasure	“In particular, a data subject should have the right to have his or her personal data erased and no longer processed where the personal data are no longer necessary in relation to the purposes for which they are collected or otherwise processed, where a data subject has withdrawn his or her consent or objects to the processing of personal data concerning him or her, or where the processing of his or her personal data does not otherwise comply with this Regulation.”

75. *Id.* art. 15.76. *Id.* Recital 63.77. *Id.* art. 16.78. GDPR, *supra* note 21, art. 17.79. *Id.* Recital 65.

Table 3 (continued): Data Subject Rights

Article 18 ⁸⁰	Recital 67 Explanation ⁸¹
Right to Restriction of Processing	“Methods by which to restrict the processing of personal data could include, inter alia, temporarily moving the selected data to another processing system, making the selected personal data unavailable to users, or temporarily removing published data from a website.”
Article 20 ⁸²	Article Text
Right to Data Portability	“[T]he right to receive [] personal data” that a data subject “has provided to a controller, in a structured, commonly used and machine-readable format and [] the right to transmit those data to another controller without hindrance.” Also, “the right to have personal data transmitted from one controller to another [controller], where technically feasible.”
Article 21 ⁸³	Article Text
Right to Object	A data subject’s right to object, “at any time to processing of personal data concerning him or her.”
Article 22 ⁸⁴	Recital 71 Explanation ⁸⁵
“[R]ight [N]ot to [B]e [S]ubject to a [D]ecision [B]ased [S]olely on [A]utomated [P]rocessing”	“The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her.”

80. *Id.* art. 18.81. *Id.* Recital 67.82. *Id.* art. 20.83. GDPR, *supra* note 21, art. 21.84. *Id.* art. 22.85. *Id.* Recital 71.

Companies should have mechanisms in place for data subjects to exercise these rights and request, access, obtain, rectify, or erase their data.⁸⁶ These company mechanisms and expectations are further explained in Articles 12, 13, 14, and 19. Companies are expected to respond “without undue delay” to data subject requests, within one month at the latest.⁸⁷ This can be an operationally difficult requirement for many companies to meet, especially very small and very large companies where the cost of compliance is significant, for the former in terms of budget and the latter in terms of sheer scale.

F. WHY THE GDPR AFFECTED COMPANY BEHAVIOR ACROSS THE WORLD

Almost all modern businesses are affected by the GDPR because the law defines relevant data processing so broadly. A business can be subject to GDPR compliance even when the business is not physically present in the European Union. Two provisions in particular might ensnare a U.S. business: if it monitors Europeans or if it offers goods or services to Europeans. Failure to comply with the GDPR can trigger heavy fines.

The GDPR applies to data processing activities in three contexts:

- (1) A company is established in the European Union (regardless of where the data processing actually takes place);⁸⁸ or
- (2) A company is not established in the European Union, but its data processing activities include “the offering of goods or services” or the monitoring of E.U. citizens within the European Union;⁸⁹ or
- (3) A company is established in a place where E.U. member state law applies.⁹⁰

Recital 23 provides some guidance regarding what constitutes “offering goods and services” to E.U. citizens. While mere accessibility of a website or use of a member state language on a website is insufficient, actions such as using the language and currency of a member state on a website with the possibility of ordering goods and services in that language could be viewed as targeting individuals within the European Union and weigh in favor of required GDPR compliance.⁹¹ Because of the broad territorial scope of the Regulation, the threat of massive fines for noncompliance, and the fear of

86. *Id.* Recital 59.

87. *Id.*

88. GDPR, *supra* note 21, art. 3(1).

89. *Id.* art. 3(2)(a).

90. *Id.* art. 3(3).

91. *Id.* Recital 23.

losing access to the entire European market, many companies not established in the European Union have felt pressure to comply with the GDPR.⁹²

Furthermore, the Regulation applies to a variety of business processes that are commonplace in our technological age. Any company that processes personal data “wholly or partly by automated means” or has information that forms (or is intended to form) a nonautomated filing system must comply with GDPR data protection standards, provided that the company’s activities are within the territorial scope of the Regulation (explained above).⁹³ Purely personal or household activities are exempt, and there are limited exceptions for micro, small, and medium-sized enterprises.⁹⁴ However, the material scope of the Regulation remains broad, with the goal of achieving a consistent level of protection for European consumers across business sectors and industries.⁹⁵

G. GDPR ADMINISTRATIVE FINES

The GDPR changed the way data protection law is implemented across the European Union and also significantly increased sanctions for noncompliance. Companies are now facing heightened compliance requirements that, if violated, can result in massive monetary penalties. The GDPR has generated a huge shift in the relative importance of data protection law when compared to the trivial fines that were administered under the Data Protection Directive.⁹⁶ The risk of such severe penalties has forced companies to prioritize GDPR compliance.

There are two categories of fines: a higher category and a lower category.⁹⁷ The highest fine category, outlined in GDPR Article 83(5), sets a maximum fine of €20 million or four percent of a company’s total worldwide turnover and can be imposed for violating:

- “Basic principles [of data] processing”;
- “[D]ata subjects’ rights”;

92. See, e.g., Ivana Kottasova, *These Companies Are Getting Killed by GDPR*, CNN BUSINESS (May 11, 2018), <https://money.cnn.com/2018/05/11/technology/gdpr-tech-companies-losers/index.html>; Daniel Mikkelsen, Henning Soller, Malin Strandell-Jansson & Marie Wahlers, *GDPR Compliance Since May 2018: A Continuing Challenge*, MCKINSEY & COMPANY (July 2019), <https://www.mckinsey.com/business-functions/risk/our-insights/gdpr-compliance-after-may-2018-a-continuing-challenge>.

93. GDPR, *supra* note 21, art. 2(1).

94. *Id.* art. 2(2).

95. *Id.* Recital 13.

96. See Chris Jay Hoofnagle, Bart van der Sloot & Frederik Zuiderveen Borgesius, *The European Union General Data Protection Regulation: What It Is and What It Means*, INFO. & COMM. TECH. L., 28:1, 94 (Feb. 10, 2019), <https://www.tandfonline.com/doi/full/10.1080/13600834.2019.1573501>.

97. GDPR, *supra* note 21, arts. 83(4)–(5).

- International data transfer processes;
- Member state law regarding the processing of specific types of data; and
- Orders from a Supervisory Authority.⁹⁸

The lesser category of fines, found in Article 83(4), sets a maximum fine of €10 million or two percent of a company's total worldwide turnover, and applies to violations of:

- “[O]bligations of [] controller[s] and [] processor[s]”;
- “[O]bligations of [] certification bod[ies]”; and
- “[O]bligations of [a] monitoring body.”⁹⁹

Data Protection Authorities are not required to impose fines for infringement of the GDPR.¹⁰⁰ They have the power to choose the most appropriate corrective measures in each instance.¹⁰¹ Corrective measures may include, but are not limited to, warnings, reprimands, fines, banning data processing, and ordering that data breach notifications be sent to data subjects.¹⁰²

When a DPA does elect to impose an administrative fine for GDPR infringement, there are four guiding principles that should be considered.¹⁰³ First, administrative fines should be equivalent across member states.¹⁰⁴ Similar fines should be imposed for similar violations.¹⁰⁵ Article 83(2) provides eleven criteria that shall be given “due regard” when deciding whether to impose a fine, and determining the amount of the fine.¹⁰⁶ Second, corrective measures (including fines) should be “effective, proportionate and dissuasive.”¹⁰⁷ DPAs must assess all the facts of a case in a consistent and objective manner, and respond adequately to the severity of the infringement.¹⁰⁸ The Article 29 Working Party guidelines suggest that this “effective, proportionate and dissuasive” standard will be more precisely defined through practice and future case-law.¹⁰⁹ Third, each case must be assessed individually, starting with the

98. *Id.* art. 83(5).

99. *Id.* art. 83(4).

100. Article 29 Data Protection Working Party, *supra* note 16, at 6–8.

101. GDPR, *supra* note 21, Recital 148.

102. *Id.* art. 58(2).

103. Article 29 Data Protection Working Party, *supra* note 16, at 5–8.

104. GDPR, *supra* note 21, Recital 11.

105. Article 29 Data Protection Working Party, *supra* note 16, at 5.

106. GDPR, *supra* note 21, art. 83(2).

107. *Id.* art. 83(1).

108. Article 29 Data Protection Working Party, *supra* note 16, at 6.

109. *Id.* (emphasis removed).

Article 83(2) fine considerations.¹¹⁰ Finally, DPAs are expected to cooperate with one another and the European Commission, through formal and informal means, to achieve a harmonized approach to administrative fines.¹¹¹

Fines should be used as an effective tool, neither overused nor viewed as a last resort.¹¹² Individual countries can choose whether or not to apply the Article 83 administrative fine structure to public authorities and bodies established in that country.¹¹³ When a country's legal system does not allow for administrative fines, as is the case in Denmark and Estonia, Article 83 can be applied to initiate and impose a fine.¹¹⁴ Some countries have specified when certain corrective measures will be used. In Austria, for example, first-time infringers are only issued a warning.¹¹⁵

With an understanding of the goals and compliance requirements of the GDPR, Parts III through V will now present an empirical analysis of the administrative fines that have been imposed thus far. Part III analyzes each country's enforcement behavior and GDPR implementation approach. Part IV presents the data, including a breakdown of fines by country and violation type, a comparison of total fines by country to gross domestic product, and a discussion of the industries most affected by enforcement actions. Part V discusses anticipated enforcement trends based on the data and findings presented in Parts III and IV.

III. A COMPARISON OF THE APPROACHES TAKEN ACROSS THE EU TO IMPLEMENT AND SUPPLEMENT THE GDPR

As a regulation, the GDPR became immediately enforceable as law, in what was then twenty-eight E.U. member states,¹¹⁶ on May 25, 2018,¹¹⁷ but member states implemented the Regulation through their own national privacy legislation in order to incorporate permitted country-specific alterations to the core text of the GDPR. The GDPR was incorporated into the EEA Agreement and enacted in Iceland, Norway, and Lichtenstein, the three EEA

110. *Id.* at 6–7.

111. *Id.* at 8.

112. *Id.* at 7.

113. GDPR, *supra* note 21, art. 83(1).

114. GDPR *Member State Permitted Variations and Requirements Chart: Overview*, THOMSON REUTERS PRACTICAL LAW UK, <https://uk.practicallaw.thomsonreuters.com/w-012-6272> (last visited May 19, 2020).

115. *The GDPR: One Year On*, IUS LABORIS (May 24, 2019), <https://theword.iuslaboris.com/hrlaw/insights/the-gdpr-one-year-on>.

116. European Union, *supra* note 28.

117. GDPR, *supra* note 21, art. 99.

countries not part of the European Union, on July 6, 2018.¹¹⁸ While enforceable as law, the GDPR allows for a number of derogations where member states can exercise discretion over how specific articles are applied.¹¹⁹ These derogations include the option to lower the age of consent from sixteen years to no lower than thirteen years of age,¹²⁰ introduce further limitations on the processing of health data,¹²¹ and retain certain data processing laws that member states may already have in place.¹²² Member states were given the ability to restrict and supplement specific GDPR articles through derogations in order to ensure national and public security, as well as to safeguard other member state interests.¹²³ As a result, countries adopted their own national data privacy legislation which included nation-specific alterations to the core text of the GDPR.¹²⁴ Some member states released updated privacy laws quickly, while other member states did not implement the GDPR into their national privacy legislation until well into 2019.¹²⁵ For example, the United Kingdom's national Data Protection Act 2018 became enforceable the same day as the GDPR: May 25, 2018.¹²⁶ The United Kingdom's Data Protection Act is closely related to the GDPR, using the GDPR as a baseline and extending GDPR principles to additional data processing scenarios through derogations.¹²⁷ As illustrated in Table 4, twelve member states (approximately thirty-nine percent) implemented their own national privacy legislation by the GDPR enforcement date of May 25, 2018. The remaining countries implemented their legislation

118. THOMSON REUTERS PRACTICAL LAW UK, *supra* note 22.

119. Andrew Clearwater and Brian Philbrook, *GDPR Derogations and How to Prepare for Member State Variation*, CPO MAGAZINE (Sept. 29, 2017), <https://www.cpomagazine.com/data-protection/gdpr-derogations-prepare-member-state-variation/>.

120. GDPR, *supra* note 21, art. 8(1).

121. *Id.* art. 9(4).

122. Andrew Clearwater and Brian Philbrook, *GDPR Derogations and How to Prepare for Member State Variation*, CPO MAGAZINE (Sept. 29, 2017), <https://www.cpomagazine.com/data-protection/gdpr-derogations-prepare-member-state-variation/>.

123. GDPR, *supra* note 21, Recital 73.

124. *See GDPR Genius—Chapter 1 - General Provisions: Nation-Specific Notes*, INTERNATIONAL ASSOCIATION OF PRIVACY PROFESSIONALS (2019), <https://iapp.org/resources/tools/gdpr-genius-chapter-1/>.

125. *See Data Protection Laws of the World*, DLA PIPER (Jan. 2020), <https://www.dlapiperdataprotection.com/>.

126. *Data Protection Laws of the World—United Kingdom*, DLA PIPER (Jan. 14, 2020), <https://www.dlapiperdataprotection.com/?t=law&c=GB>.

127. Dan Swinhoe, *GDPR vs UK Data Protection Act 2018: What's the difference?*, CSO ONLINE (Aug. 5, 2019), <https://www.csoonline.com/article/3410039/gdpr-vs-uk-data-protection-act-2018-whats-the-difference.html>.

later in 2018, with nine countries not implementing national privacy legislation until 2019.¹²⁸ Slovenia has yet to pass local privacy legislation post-GDPR.¹²⁹

Countries also had discretion over their GDPR implementation approach regarding how quickly and heavily they fined violators. Some countries were slow to fine, prioritizing education regarding proper GDPR implementation and delaying penalization for infringements. For example, Hungary's Data Protection Act requires its DPA to only issue warnings to companies who violate the Regulation for the first time.¹³⁰ In contrast, the Lithuanian DPA took a more active approach, announcing in January 2019 a list of seventy-five organizations it planned to inspect for GDPR compliance that year.¹³¹

To understand the different implementation approaches used, we assessed each country's approach to GDPR enforcement and categorized it as vigorous, progressing, or cautious based on the country's implementation approach, the total number of fines issued as of March 31, 2020, the cumulative total of all fines imposed, and the date when national privacy legislation was enacted (Table 4). Vigorous countries actively fined companies for violations, imposed a high number of fines or a few extremely large fines, and implemented their national privacy legislation before or near the date the GDPR became enforceable. In contrast, cautious countries were warning or education-first focused, imposed relatively few fines, and frequently enacted national privacy legislation later than the majority of E.U. member states. Progressing countries were countries with a combination of vigorous and cautious attributes. Based on these metrics, the five countries considered vigorous as of March 2020 were France, Germany, Italy, Spain, and the United Kingdom. Of the remaining twenty-six countries, thirteen (forty-two percent) were evaluated to be progressing, and thirteen (forty-two percent) were deemed cautious in their GDPR implementation approach.

The analysis presented in Table 4 was done twice in the process of writing this Note, once in May 2019 and again in March 2020. Both analyses used the same assessment factors. In May 2019, only four countries were categorized as vigorous. Italy moved from progressing to vigorous after it imposed a few large fines from May 2019 to March 2020, increasing its cumulative fine total from €50,000 to €39,454,946 (third only to the United Kingdom and France).

128. See THOMSON REUTERS PRACTICAL LAW UK, *supra* note 22.

129. *Data Protection Laws of the World—Slovenia*, DLA PIPER (Jan. 15, 2021), <https://www.dlapiperdataprotection.com/index.html?t=law&c=SI>.

130. *Data Protected—Hungary*, LINKLATERS (Nov. 2018), <https://www.linklaters.com/en-us/insights/data-protected/data-protected---hungary>.

131. Neil Hodge, *GDPR Enforcement Varies Widely by Country*, COMPLIANCE WEEK (July 19, 2019), <https://www.complianceweek.com/gdpr/gdpr-enforcement-varies-widely-by-country/27436.article>.

Five countries transitioned from cautious in May 2019 to progressing in March 2020: Austria, Hungary, the Netherlands, Romania, and Sweden. The Austrian privacy advocacy organization, None of Your Business, has been incredibly active in filing actions against companies for GDPR violations, moving Austria from the cautious to progressing category. Hungary's cumulative fine total more than doubled from May 2019 to March 2020, and Romania's total fine count increased fivefold in that period of time. Sweden imposed a few massive fines, increasing its cumulative fine total by more than €7,000,000. Similarly, the Netherlands both doubled its fine count and its cumulative fine total, warranting a progressing categorization. Lithuania, on the other hand, stated an intent to begin over seventy-five investigations in 2019,¹³² but no evidence of additional sanctions, monetary or nonmonetary, were found. Therefore, Lithuania was recategorized from progressing to cautious.

132. Neil Hodge, *GDPR Enforcement Varies Widely by Country*, COMPLIANCE WEEK (July 19, 2019), <https://www.complianceweek.com/gdpr/gdpr-enforcement-varies-widely-by-country/27436.article>.

IV. A CLOSER LOOK AT GDPR FINES: GROWTH RATES, VIOLATION TYPES, QUANTITY BY COUNTRY AND ECONOMIC SECTOR

A. METHODS

To understand the GDPR enforcement landscape, we performed an exhaustive search to identify all fines and nonmonetary sanctions imposed and pending in the European Union as a result of the Regulation. This included an internet review of GDPR enforcement actions from May 25, 2018, the date of enactment, through March 31, 2020. We assembled a dataset of publicly available fines and nonmonetary sanctions for all twenty-seven E.U. member states, the United Kingdom, and three EEA—Lichtenstein, Iceland, and Norway—which are also subject to the GDPR.¹³³ Key sources included the CMS GDPR Enforcement Tracker,¹³⁴ Linklaters' Data Protected tracker,¹³⁵ DLA Piper's Data Protection Laws of the World Handbook,¹³⁶ PwC's Global Privacy and Security Enforcement Tracker,¹³⁷ Nathan Trust's GDPR Fines and Penalties News Feed,¹³⁸ Bird & Bird's GDPR Tracker,¹³⁹ None of Your Business's GDPRhub,¹⁴⁰ and country-specific DPA Annual Reports. News articles, Bloomberg Law resources, Thomson Reuters Practical Law charts, and European Data Protection Board (EDPB) reports were also examined. While every attempt was made to be comprehensive, the data sources are public information and impacted by variations in reporting, timeliness, and consistency.

Our search strategy yielded a total of 311 enforced fines and forty-nine pending actions across thirty-one countries for the time period of May 25, 2018, through March 31, 2020. We collected numerous data elements about each enforcement action, including imposed or pending status, fine amount,

133. THOMSON REUTERS PRACTICAL LAW UK, *supra* note 22.

134. GDPR Enforcement Tracker, CMS, <https://enforcementtracker.com/> (last visited Mar. 31, 2020).

135. *Data Protected Home: Your Global Guide to Data Protection*, LINKLATERS, <https://www.linklaters.com/en/insights/data-protected/home> (last visited March 31, 2020).

136. *Data Protection Laws of the World*, DLA PIPER (Jan. 2020), <https://www.dlapiperdataprotection.com/>.

137. *Global Privacy and Security Enforcement Tracker*, PWC (2018), <https://www.pwc.com/gx/en/issues/regulation/general-data-protection-regulation/hot-topics/enforcement-tracker.html>.

138. *GDPR Fines and Penalties*, NATHAN TRUST, <https://www.nathantrust.com/gdpr-fines-penalties> (last visited Mar. 31, 2020).

139. *GDPR Tracker*, BIRD & BIRD, <https://www.twobirds.com/en/in-focus/general-data-protection-regulation/gdpr-tracker> (last visited Mar. 31, 2020).

140. *GDPRhub*, NONE OF YOUR BUSINESS, https://gdprhub.eu/index.php?title=Welcome_to_GDPRhub (last visited Mar. 31, 2020).

company, date of decision, and type of violation. Country-specific information such as gross domestic product (GDP), local privacy legislation implementation date, and GDPR enforcement approach were also included in the data set. We conducted quantitative analyses using descriptive statistics (mean, median, mode, range, linear regression) and used data visualization tools to create graphs that summarized our findings.

Not all countries' records are publicly available, and many DPAs only release annual reports. We have attempted to use the most up-to-date records and sources available, but it is possible that some of the reported information used to create the data set is outdated or incomplete due to collection limitations.

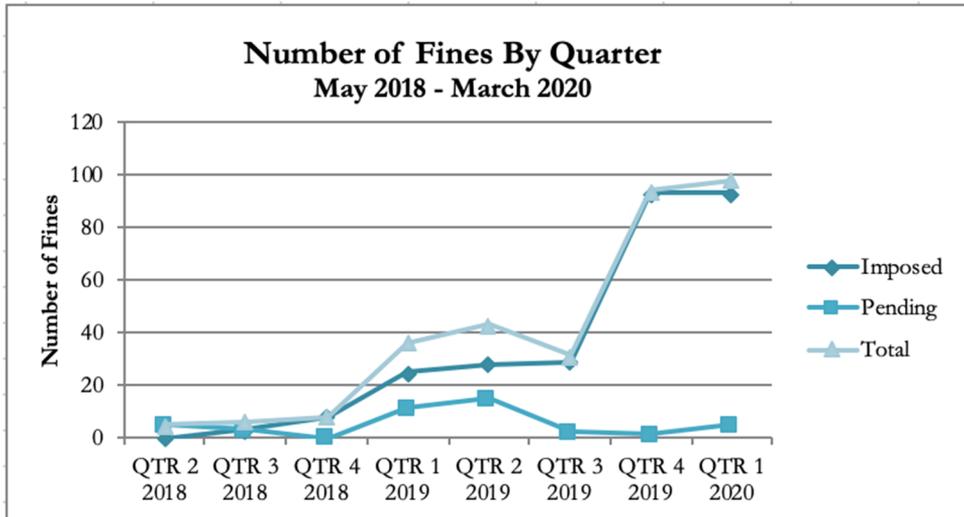
B. FINDINGS

1. *Violations: Quantity and Type*

The number of fines has increased steadily quarter-over-quarter and year-over-year since May 2018 with the greatest number of new fines imposed in Q4 2019. Many companies worked diligently for months to be compliant with the GDPR, in order to avoid millions of euros in penalties. In the first six months after the GDPR was enacted, fines were imposed, but in perhaps smaller numbers than expected. Imposed fines first started to increase notably in Quarter 1 of 2019. It is highly probable that DPAs started identifying targets of enforcement before May 2018, while looking for enforceable violations that took place after GDPR's enforcement date. Assuming that is the case, the initial enforcement trends suggest that it takes at least nine to ten months for a DPA to investigate and assess a fine. This timeline stands in contrast to the U.S. Federal Trade Commission, which, in non-fraud privacy cases, typically takes more than one year to investigate and assess penalties.¹⁴¹ Figure 1 illustrates that the number of imposed fines increased significantly in 2019, with sixty-three new enforcement actions initiated in Quarter 4 of 2019. The leveling off of fines from the end of 2019 though the start of 2020 was likely due to lag time in the reporting of fine decisions from the end of our collection period.

141. Commenting on the FTC's relatively slow approach in 2016, Chris Hoofnagle observed, "In a single year, the FCC [Federal Communications Commission] levied \$42 million in privacy fines. The FCC will soon eclipse the FTC's records of fines, which is approximately \$60 million in the FTC's eighteen-year history of online privacy cases." *In* Chris Jay Hoofnagle, *Federal Trade Commission Privacy Law and Policy* (Cambridge University Press 2016).

Figure 1: Number of Fines Per Quarter



Pending cases are not visibly increasing at a rate comparable to imposed fines, but it is difficult to track pending cases. Some DPA records are much more accessible and detailed than others. It is also important to note that a number of pending cases are cross-border. In these instances, multiple countries are working together to pursue enforcement actions. This requires significant cooperation and coordination, as well as record sharing. There are also some situations where one country has been designated to lead an investigation based on where the violator is established within the European Union, one example being Ireland.¹⁴² The majority of Ireland's pending cases are directed at large U.S. technology companies whose E.U. headquarters are located in Ireland.

Fines imposed due to a violation of one or more of the seven data processing principles (Table 2) were two and a half times more frequent than fines imposed due to the violation of data subject rights (Table 3), and nearly three times more frequent than fines imposed due to a violation of controller and processor duties (Table 5). The type of violation dictates the maximum administrative fine that can be imposed. Data protection principle violations and data subject right violations (Chapters II and III in Table 5) are viewed as the most serious of all infringements, and give rise to the highest maximum

142. See, e.g., Dara Murphy, *Dara Murphy: Ireland Is Up To the Data Protection Task Speech*, INTERNATIONAL ASSOCIATION OF PRIVACY PROFESSIONALS DATA PROTECTION INTENSIVE (Apr. 21, 2015), <https://www.youtube.com/watch?v=CMVuk0Cgg5o> [hereinafter IAPP Videos].

fine allowed under the GDPR: €20 million, or four percent of a company's total worldwide annual turnover.¹⁴³

It is interesting that data processing principle violations were a greater source of fines than data subject rights violations. Data processing principles include transparency, data minimization, and storage limitation, to name a few. Data subject rights, such as the right to data access and erasure, appear simpler to enforce than the more obscure issues of transparency and data minimization. However, DPAs can easily download companies' privacy policies and begin monitoring their compliance with the GDPR data processing principles. Individual complaints regarding data subject rights, on the other hand, probably get stuck in the review pipeline as DPAs struggle to keep pace administratively.

Table 5: GDPR Articles Cited as Reason for Violation(s)
May 25, 2018-March 31, 2020

GDPR Article Cited in Fine	Number of Citations
Chapter I – General Provisions	6
Chapter II – Principles	339
Chapter III – Rights of the Data Subject	135
Chapter IV – Controller and Processor	117
Chapter V – Transfers of Personal Data to Third Countries/Int'l Organizations	0
Chapter VI – Independent Supervisory Authorities	18
Chapter VII – Cooperation and Consistency	0
Chapter VIII – Remedies, Liability and Penalties	8
Chapter IX – Provisions Relating to Specific Processing Situations	0
Chapter X – Delegated Acts and Implementing Acts	0
Chapter XI – Final Provisions	0
Not Reported	52

A more granular breakdown of the Article 5 data processing principle violations shows that actions were most frequently brought for the infringement of the 5(a) lawfulness, fairness, and transparency principle (Figure 2). Of the 221 fines that cited a specific Article 5 violation, eighty-nine of those fines (forty percent) cited a violation of 5(a). Integrity and confidentiality and data minimization were the second and third most common principle violations, accounting for twenty-one percent and nineteen percent of Article 5 violations, respectively. Of the 135 cases that cited specific data

143. GDPR, *supra* note 21, art. 83.

subject right violations, the right to access and companies failing to provide information about the processing of information received from data subjects were the two most commonly cited infringements (Figure 3). These two violations were cited in fifty-three percent of data subject right actions where administrative fines or other nonmonetary sanction(s) were imposed or a final decision was pending.

Figure 2: Breakdown of Article 5 Data Processing Principle Violations
May 25, 2018-March 31, 2020

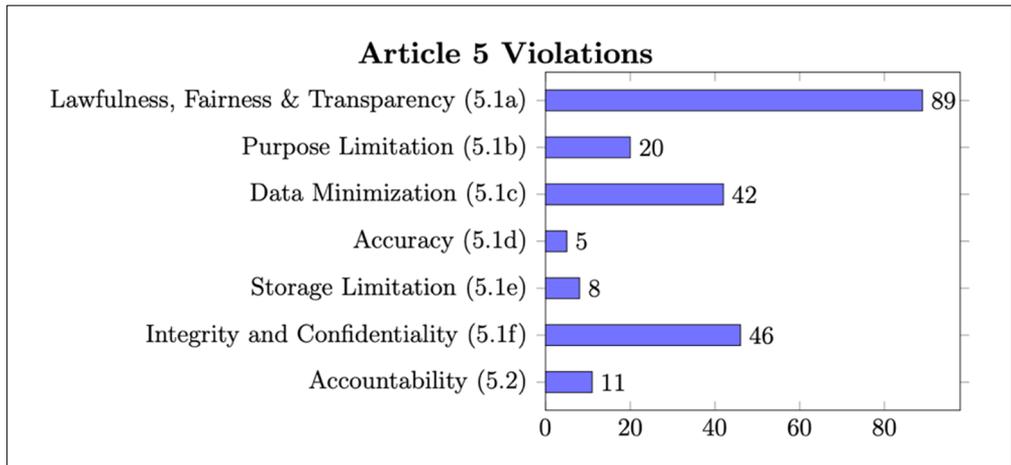
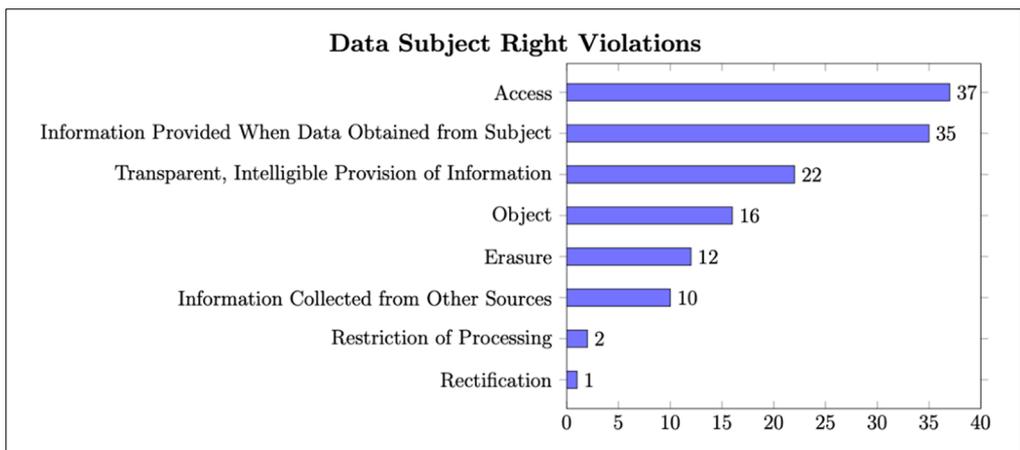


Figure 3: Breakdown of Data Subject Right Violations
May 25, 2018-March 31, 2020

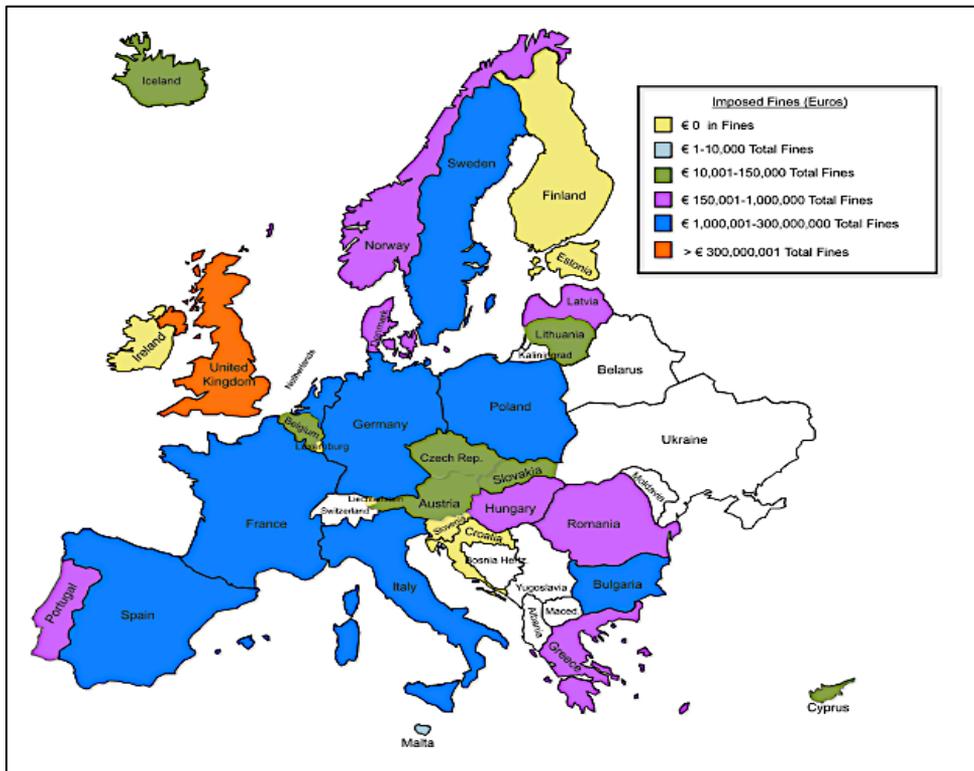


2. Total Fines Imposed by Country

Nine countries emerged as leaders in total fines imposed. The United Kingdom imposed €315,869,695 in GDPR fines, the highest cumulative fine total of any European country. Eight countries (Spain, France, Italy, Germany, Poland, Bulgaria, Sweden, and the Netherlands) imposed more than €1,000,001 each in total fines during the period of May 25, 2018 through March 31, 2020 (Figure 4). It could be argued that €1,000,001 in total fines over two years is actually a low total given that individual administrative fines for GDPR violations can reach €20 million, or four percent of a company's total worldwide annual turnover.¹⁴⁴

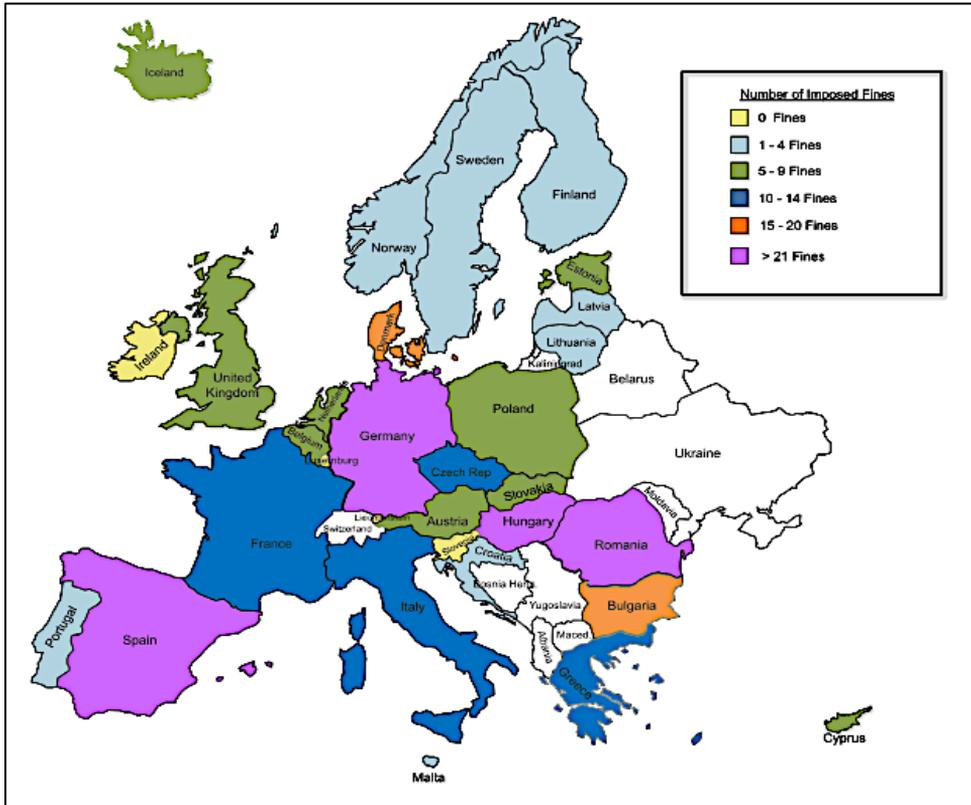
A high cumulative fine total can occur from a few large fines or a number of small fines. In the case of the United Kingdom, Sweden, and the Netherlands, fines have been few but mighty. Spain, in contrast, imposed one hundred fines, more than double the total fine count of any other country.

Figure 4: Total Fines Imposed by Country
May 25, 2018-March 31, 2020



144. GDPR, *supra* note 21, art. 83.

Figure 5: Number of Imposed Fines by Country
May 25, 2018-March 31, 2020



Ireland had yet to impose any significant fines as of March 31, 2020. Ireland's Data Protection Commission (DPC) was said to be conducting as many as twenty investigations into multinational companies with highly anticipated final decisions on the horizon, but the Irish DPC did not impose its first GDPR fine until mid-May 2020, and surprisingly, it wasn't against a U.S.-based technology company.¹⁴⁵ This first fine was not confirmed in the Dublin Circuit Court until November 4, 2020, almost two and a half years after

145. Elizabeth Schulze, *Big Tech Fears US Regulation, But It May Be Ireland That Should Scare Them*, CNBC (June 20, 2019), <https://www.cnbc.com/2019/06/20/technology-regulation-irelands-helen-dixon-has-attention-of-big-tech.html>. Colm Keena, *Tulsa Becomes First Organization Fined for GDPR Rule Breach*, IRISH TIMES (May 17, 2020), <https://www.irishtimes.com/news/crime-and-law/tulsa-becomes-first-organisation-fined-for-gdpr-rule-breach-1.4255692>.

the GDPR went into effect.¹⁴⁶ Ireland's actions are of great interest to many American companies who have established their E.U. headquarters there. As shown below in Table 7, as of March 31, 2020, Ireland had ten pending fines that were public, second only to Austria, the home of an active data protection advocacy organization called None of Your Business.¹⁴⁷ In many cross-border cases involving American technology companies, Ireland has been designated the lead supervisory authority in charge of penalty assessment because the companies' E.U. headquarters are in Ireland.¹⁴⁸ Because of its many pending actions, Ireland will continue to be a key country of interest. But at the same time, businesses are likely to invest heavily in defending these cases to retain Ireland's aversion to stricter continental enforcement approaches. The Irish DPC not only has many companies to oversee, it also has respondents known for scorched-earth litigation tactics. For example, Facebook (a frequent GDPR litigant), sought dismissal of one Dutch case because the court failed to adhere to the country's strict language requirements by using the words "browser" and "cookie" rather than "internetsnuffelaar" and "koekje zijn."¹⁴⁹

From May 25, 2018, through March 31, 2020, there were six notably large fines, the two largest coming from the United Kingdom and followed, in order of magnitude, by France, Italy, and Germany. The majority of these fines were imposed in the second half of 2019. British Airways received the largest fine for GDPR noncompliance, resulting in a fine of 1.5% of its total revenue for the year 2018.¹⁵⁰ However, the United Kingdom's Information Commissioner's Office deferred the payment of the enormous fines levied against both British Airways and Marriott International twice, pending further investigations later in 2020.¹⁵¹ The British Airways fine was ultimately reduced

146. *Data Protection Commission Fine on Tusla Child and Family Agency Confirmed in Court*, DATA PROTECTION COMMISSION (Nov. 4, 2020), <https://www.dataprotection.ie/en/news-media/press-releases/data-protection-commission-fine-tusla-child-and-family-agency-confirmed-court>.

147. *See Our Detailed Concept*, NONE OF YOUR BUSINESS, <https://noyb.eu/en/our-detailed-concept> (last visited Mar. 31, 2020).

148. *See, e.g.*, IAPP Videos, *supra* note 142.

149. Michaël Temmerman, *With This (Strange) Argument, Facebook Strikes Back to Our Country*, HET NIEUWSBLAD (Jan. 27, 2016), https://m.nieuwsblad.be/cnt/dmf20160127_02093367.

150. Ingrid Lunden, *UK's ICO Fines British Airways a Record £183M Over GDPR Breach That Leaked Data From 500,000 Users*, TECHCRUNCH (July 8, 2019), <https://techcrunch.com/2019/07/08/uks-ico-fines-british-airways-a-record-183m-over-gdpr-breach-that-leaked-data-from-500000-users/>.

151. Melanie Mingas, *ICO Confirms Second Deferral For BA and Marriott's GDPR Fines*, DATA ECONOMY NEWSROOM (Apr. 17, 2020), <https://data-economy.com/ico-confirms-second-deferral-for-ba-and-marriotts-gdpr-fines/>.

to €20 million, a ninety-percent reduction from the initial July 2019 fine.¹⁵² Marriott's fine was also drastically reduced to €18.4 million.¹⁵³

Table 6: Six Largest Fines by Country¹⁵⁴
May 25, 2018-March 31, 2020

Country	Date Imposed	Company	Fine Imposed (Euros)
United Kingdom	July 2019	British Airways	204,600,000
United Kingdom	July 2019	Marriot International	110,390,200
France	January 2019	Google	50,000,000
Italy	January 2020	TIM S.p.A.	27,802,946
Germany	October 2019	Deutsche Wohnen SE	14,500,000
Germany	December 2019	1&1 Telecom GmbH	9,550,000

Four of the six largest fines were imposed for insufficient security measures. The fines against British Airways, Marriott, and 1&1 Telecom GmbH were imposed for a violation of Article 32, which requires companies to implement “appropriate technical and organisational measures to ensure a level of security appropriate to the risk.”¹⁵⁵ The fine in Italy against TIM S.p.A. was also imposed, in part, due to an Article 32 violation.¹⁵⁶ Security vulnerabilities are generally recognized as privacy disasters because those flaws can be exploited by “hackers” and other malicious actors. Once data are stolen, all use-based and policy-based controls on it are impossible to enforce. In some cases, these data end up on publicly-available websites for anyone to download. This represents a total failure of market promises of security. Insufficient security measures will thus likely continue to be a catalyst for high fines because the harm is evident and company expectations under the GDPR

152. *ICO Fines British Airways £20m for Data Breach Affecting More Than 400,000 Customers*, INFORMATION COMMISSIONER'S OFFICE (Oct. 16, 2020), <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2020/10/ico-fines-british-airways-20m-for-data-breach-affecting-more-than-400-000-customers/>.

153. *ICO fines Marriott International Inc £18.4million For Failing to Keep Customers' Personal Data Secure*, INFORMATION COMMISSIONER'S OFFICE (Oct. 30, 2020), <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2020/10/ico-fines-marriott-international-inc-184million-for-failing-to-keep-customers-personal-data-secure/>.

154. Both U.K. fines were significantly reduced by the ICO in October 2020.

155. GDPR, *supra* note 21, art. 32(1).

156. *Marketing: From the Privacy Guarantor a Fine of 27 Million and 800 Thousand Euros to Tim*, GARANTE PER LA PROTEZIONE DEI DATI PERSONALI (Feb. 1, 2020), <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9256409>.

are clear—companies are obligated to implement procedures that appropriately protect consumers’ personal data. In contrast, other privacy violations are more challenging to assess given that many of the GDPR articles are written at principle-level abstraction.

3. *Comparing Total Fines by Country to Gross Domestic Product*

When comparing each country’s gross domestic product (GDP) against the total number of fines each country imposed, countries with high GDPs imposed more fines and higher fines than countries with lower GDPs. GDP is a monetary measure of the market value of total goods and services produced by a country, and is considered a powerful indicator of economic development and progress. With the exception of Ireland, the three countries that had yet to impose a fine under the GDPR as of March 31, 2020, had a thirty-third percentile GDP when compared to the GDP of the thirty-one countries subject to the GDPR. The top nineteen countries by GDP (again excluding Ireland) had all imposed some number of fines. There were three countries in the fifty-fifth percentile by GDP that imposed a number of fines: Romania, Hungary, and Bulgaria (Table 7). In Table 4 above, we categorized each of these countries in our March 2020 assessment as having a “progressing” approach towards GDPR enforcement. Nearly all of the fines, in each of these three countries, were imposed in 2019 and 2020. This suggests that countries with less economic resources took slightly longer to mobilize, but fines are now on the rise in these countries as well.

In Table 4 above, we categorized five countries in our March 2020 analysis as having a “vigorous” approach to GDPR enforcement: France, Germany, Italy, Spain, and the United Kingdom. In 2019, these countries were the top five GDP earners across the European Union. It is not surprising that countries with strong economies would have the resources to create a robust GDPR enforcement infrastructure. Specific countries, such as Germany and the United Kingdom, arguably took the lead on enforcement actions, while other countries may be waiting for more precedent and cooperation standards to be set by these leaders. The ombudsman of Finland, one of the countries categorized as “cautious,” has publicly spoken about the importance of E.U. harmonization, and it is expected that the European Data Protection Board guidelines will strongly influence Finland’s future GDPR enforcement.¹⁵⁷ This is one example of how some countries, even those with a sizeable GDP, are taking a more cautious approach to enforcement, waiting to see how guidelines and cooperation mechanisms develop across the EU.

157. IUS LABORIS, *supra* note 115.

Additionally, there is a positive correlation between fine amount and the GDP of the country where the fine is issued. GDPR violators pay higher fines in countries with higher GDPs. As GDP increases by €1 million, the fine amount increases by €2,770.¹⁵⁸

158. This is based on a regression of 268 enforcement cases at a ninety-nine-percent confidence level with an R-squared value of 0.032.

Table 7: 2019 Gross Domestic Product¹⁵⁹ and Fines by Country¹⁶⁰
(Count & Monetary Value)

Country	2019 Annual GDP (Euro-Millions)	Imposed Fines (Euros)	Implemented Fines /Sanctions (Count)	Pending Fines /Sanctions (Count)	Total Imposed and Pending Fines/Sanctions (Count)
Germany	3,435,990	25,060,925	21	4	25
United Kingdom	2,523,314	315,869,695	6	1	7
France	2,418,997	52,050,000	12	2	14
Italy	1,787,664	39,454,946	11	1	12
Spain	1,244,757	2,501,270	100	0	100
Netherlands	812,051	2,535,000	6	3	9
Poland	527,033	1,047,248	8	1	9
Sweden	474,683	7,053,630	3	1	4
Belgium	473,639	39,000	6	3	9
Austria	398,522	70,200	7	11	18
Norway	359,109	400,400	4	4	8
Ireland	347,215	0	0	10	10
Denmark	310,576	381,850	17	0	17
Finland	240,924	0	2	1	3
Romania	222,090	495,500	25	0	25
Czech Republic	219,896	19,035	11	2	13
Portugal	212,303	422,000	3	0	3
Greece	187,457	750,000	13	1	14
Hungary	143,826	218,183	24	0	24
Slovakia	94,177	90,000	6	0	6
Luxembourg	63,516	0	0	0	0
Bulgaria	60,675	2,744,820	17	1	18
Croatia	53,937	0	1	0	1
Lithuania	48,339	61,500	1	0	1
Slovenia	48,007	0	0	2	2
Latvia	30,476	157,000	2	0	2
Estonia	28,037	0	8	1	9
Cyprus	21,944	121,000	8	0	8
Iceland	21,603	29,000	5	0	5
Malta	13,209	5,000	1	0	1
Liechtenstein*	5,823	0	0	0	0
Totals		451,577,202	328	49	377

* 2018 GDP in Euro-Millions

159. GDP—Gross Domestic Product, COUNTRYECONOMY.COM, <https://countryeconomy.com/gdp> (last visited Aug. 6, 2021).

160. This table includes forty-one implemented nonmonetary sanctions and one pending nonmonetary sanction.

4. *Fines by Economic Sector*

One goal of the GDPR was to update data protection law so E.U. citizens would be adequately protected in an era of constant technological innovation; the size and quantity of fines imposed against technology companies in the first two years of implementation are consistent with that objective. Economists view the economy as divided into four main sectors: primary, secondary, tertiary, and quaternary. A “sector” is a subset of businesses that share similar product or service offerings, as described in Table 8. Interestingly, the tertiary economic sector, which encompasses companies that provide enterprise and consumer services, received the most fines. The quaternary sector, the sector that includes technology companies, received significantly fewer fines. However, the quaternary sector did receive some of the largest fines imposed under the GDPR and had the most fines pending of any sector.

Perhaps surprisingly, multiple small fines totaling just a few hundred euros were imposed, and primarily against companies in the tertiary sector (enterprise and consumer services). A €300 fine was charged to a private car owner in Austria for inappropriate use of Dashcam, a camera that records a vehicle’s travel through the front and sometimes rear windshields. The Czech Republic imposed a €388 fine against an employer that did not properly delete data on its Facebook page about a former employee.

A variety of industries have received fines, such as schools, taxi companies, restaurants, banks, political campaigns, and even private persons. For example, a school in Sweden was fined €18,630 for inappropriately using facial recognition to monitor student attendance. Simply put, no industry or individual is immune to GDPR enforcement.

Table 8: Imposed and Pending Fines and Nonmonetary Sanctions by Economic Sector¹⁶¹

May 25, 2018-March 31, 2020

Economic Sectors	Description	Imposed Fines	Pending Fines
Primary	Extraction and harvesting of natural resources such as agriculture and mining.	0	0
Secondary	Comprises construction, manufacturing, and processing. Basically, this sector comprises industries that relate to the production of finished goods from raw materials.	2	0
Tertiary	This type of industry provides services and includes companies such as retailers, entertainment, financial, insurance, social and personal services. These companies provide services to consumers and business.	245	5
Quaternary	This sector deals with knowledge or intellectual pursuits including research and development, business, consulting services, and education.	30	43

U.S. companies were not disproportionately fined overall, but U.S. technology companies did receive the majority of quaternary sector fines. Of the 325 fines (imposed and pending) where an economic sector could be identified, only forty-eight actions (eight imposed, forty pending) were against U.S. companies. However, the majority of those forty-eight U.S. actions were against U.S. “big tech” companies. A total of thirty-eight of the forty-eight U.S. actions (seventy-nine percent) involved “big tech” players: Apple, Amazon, Facebook, Google, Microsoft, Twitter, and Uber. European DPAs are fining many more tertiary sector companies than quaternary sector companies, but of the seventy-three actions taken against companies in the quaternary sector (thirty imposed fines, forty-three pending), forty-four of them (sixty percent) were against U.S. technology companies. Additionally, one of the six largest fines imposed during the first two years of GDPR enforcement was against a large U.S. tech company (Google). The five other largest fines were imposed against service industry companies, including the U.S. hotel chain Marriott. Within the quaternary sector, European DPAs are disproportionately fining U.S. technology companies, and the fines they impose are significant. The

161. For fifty-two fines, a company name and/or industry could not be identified.

lowest fine received by a U.S. technology company was a 2019 fine of €51,000, more than four times the 2019 fine median, €11,380.

AdTech, short for advertising technology, is a booming field that is a specific target of the GDPR.¹⁶² AdTech is a term used to describe the software and tools that companies use to send targeted online advertisements to potential customers. In January 2019, France fined Google €50 million for failing to fully explain its personalized advertising data collection process during the Android phone set-up process.¹⁶³ France's DPA, the Commission Nationale de l'Informatique des Libertés (CNIL), charged Google with not being transparent enough in its explanations regarding data collection and use for advertising purposes.¹⁶⁴ During its investigation, the CNIL observed that relevant information about advertisement personalization was disseminated across several documents during the Google account creation process, and the "Ads Personalization" section did not adequately explain to users the multitude of company services and applications involved in the process (e.g., Google search, YouTube, Google Maps, etc.).¹⁶⁵ Advertising technology is complicated and data is used to serve ads in ways that are not always straightforward or easy to explain to a layperson.¹⁶⁶ Moreover, Google is a huge company with vast legal resources. If its legal team can't satisfy France's transparency and consent requirements, smaller companies are surely in trouble if France continues to uphold these interpretations.

France's fine against Google is a case where the findings matter, not the fine. If the CNIL's interpretation of transparency is adopted by other countries, Google might be prohibited from doing the targeted advertising it wants to do because it will be impossible to keep the process transparent. As sophisticated neural networks are increasingly being used to target ads through machine learning, it might not be possible to explain in a human readable format why one profiled data subject receives a certain advertisement while

162. GDPR, *supra* note 21, art. 3(2). Under Article 3(2), "monitoring" Europeans' behavior—even without maintaining a European presence—is a regulated activity under the GDPR.

163. *The CNIL's Restricted Committee Imposes a Financial Penalty of 50 Million Euros Against GOOGLE LLC*, CNIL (Jan. 21, 2019), <https://www.cnil.fr/en/cnils-restricted-committee-imposes-financial-penalty-50-million-euros-against-google-llc>.

164. *Id.*

165. *Id.*

166. Taylor Wessing, *Adtech—What Do the EU Regulators Think?*, LEXOLOGY (Sept. 8, 2019), <https://www.lexology.com/library/detail.aspx?g=6048f2da-ffce-4c44-8225-d0e4027cb0a6>.

another data subject does not.¹⁶⁷ The CNIL's rationale threatens the viability of using complex machine learning systems to improve advertising.

V. ANTICIPATED GDPR ENFORCEMENT TRENDS

The upward trend in fines will likely continue. The number of fines imposed (Figure 6) increased nearly eightfold from 2018 (twenty-five fines) to 2019 (193 fines); ninety-three fines were imposed in the first quarter of 2020 alone. Factors that may have greatly impacted fine variability across the European Union—lack of resources, country implementation approach, and national privacy legislation timeline—will have a lesser impact on individual member state enforcement actions with time.

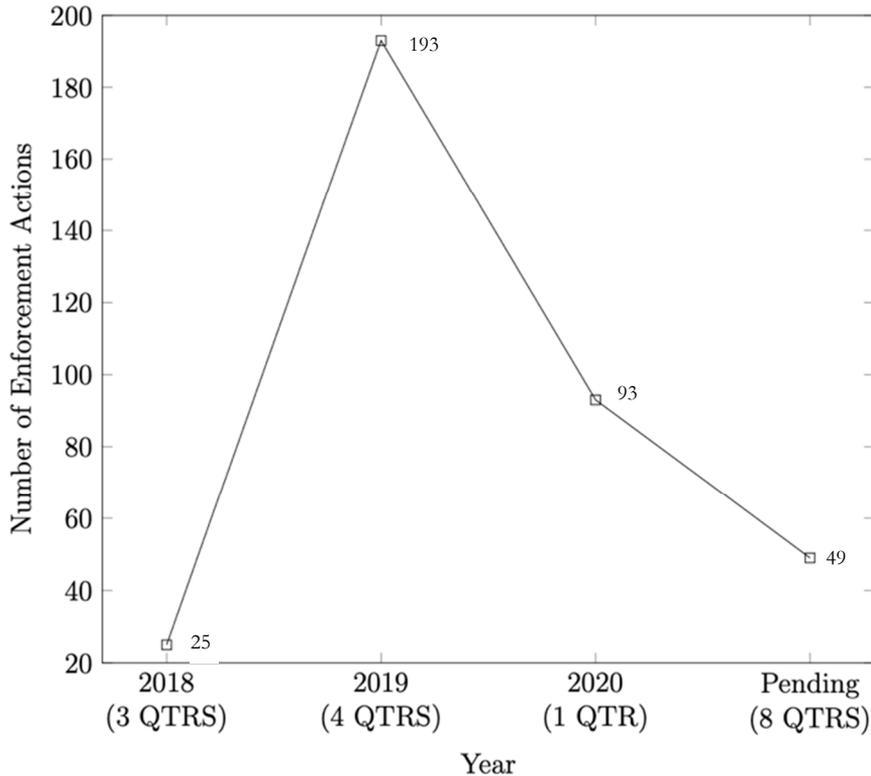
Even though, with time, E.U. member states have become better equipped to enforce the GDPR, there are still a number of factors which affect fine assessment that will continue to evolve and change country-by-country such as political climate, cultural and economic priorities, and industry presence. Even with more resources and improved collaboration methods, complete harmonization of administrative fine practices across all E.U. member states seems unlikely. Country DPAs, like any regulatory authority, set priorities based on the opinions of its internal decision makers and the feedback provided by its constituents. Even with knowledge sharing mechanisms and the EDPB advisory board dedicated to issuing GDPR implementation guidance, individual countries will surely have differing visions regarding on-the-ground enforcement. The question, then, is which metrics, if any, can appropriately evaluate the success of the GDPR? What does success look like in the context of this complex, wide-reaching regulation?

167. See generally Walter A. Mostowy, Note, *Explaining Opaque AI Decisions: How to Satisfy the GDPR's Right to an Ex Post Explanation*, 35 BERKELEY TECH. L.J. 1291 (2020).

Figure 6: Total Fines and Nonmonetary Sanctions Imposed and Pending¹⁶⁸

May 25, 2018-March 31, 2020

Fines and Penalties Imposed by EU Data Authorities



A lot of progress has been made regarding the preparatory work needed to diligently oversee companies' data processing practices, including passing legislation, restructuring agencies, and allocating funding and resources, something many countries initially lacked. Countries like Germany, with a longstanding history of prioritizing privacy and DPAs already equipped with resources and established operational processes, were better prepared to support the changes being made under the GDPR. Other countries were required to restructure their agencies and have struggled to keep up with the increase in data breach notices and complaint submissions. Belgium's Data Protection Authority, for example, was not fully operational until April 25,

168. The graph only includes fines and nonmonetary sanctions with publicly available dates.

2019, almost a full year after the GDPR was enacted.¹⁶⁹ Many countries that needed time to restructure their regulatory bodies, adopt national privacy legislation, and adapt to the new data protection landscape under the GDPR are now, three years later, better equipped to enforce penalties for noncompliance. The work is flooding in: DPAs received over 281,088 cases during the first year of GDPR enforcement.¹⁷⁰ However, enforcement is still challenging for many national DPAs. Some member states voiced concerns about the administrative burden being placed on DPAs and companies under the Regulation.¹⁷¹

Harmonization, one of the key tenants of the GDPR, is in some ways taking root. A number of countries have found ways to be innovative in their approach to GDPR implementation. Austria and Hungary both took the approach of issuing warnings for first infringements. Many countries prioritized educating companies and thus spent time drafting GDPR guidelines and best practices, only issuing fines when companies refused to cooperate. Numerous countries, in adopting new national privacy laws, implemented country-specific derogations unique to their national security, economic, and financial interests. All of these micro-decisions by different countries made harmonization and cross-border cooperation more challenging. But with the GDPR in effect for three years now, the implementation grace period is running out. Some countries are starting to take the lead on harmonization efforts. Germany developed a five-step fining structure to ensure that administrative fines are issued in an accurate and consistent manner.¹⁷² While the German fine model is complex and not without critics, the European Data Protection Board (an independent body in charge of GDPR implementation) could conceivably create a similar standardized fine model that would be binding across the European Union. Furthermore, nineteen countries released their perspectives on GDPR in

169. Laura Brodahl, Laura De Boel, Jan Dhont & Cédric Burton, *Belgian Data Protection Authority Is Up and Running*, WILSON SONSINI (Apr. 26, 2019), <https://www.wsgrdataadvisor.com/2019/04/belgian-dpa/>.

170. Case total based on information provided by twenty-seven countries. *1 Year GDPR—Taking Stock*, EUROPEAN DATA PROTECTION BOARD (May 22, 2019), https://edpb.europa.eu/news/news/2019/1-year-gdpr-taking-stock_en.

171. Preparation of the Council Position on the Evaluation and Review of the General Data Protection Regulation (GDPR) - Comments from Member States, COUNCIL OF THE EUROPEAN UNION (Oct. 9, 2019), <https://data.consilium.europa.eu/doc/document/ST-12756-2019-REV-1/en/pdf>.

172. Tim Wybitul, *German DPAs Push Model for Higher GDPR Fines*, INTERNATIONAL ASSOCIATION OF PRIVACY PROFESSIONALS (Oct. 1, 2019), <https://iapp.org/news/a/german-dpas-push-model-for-higher-gdpr-fines/>.

practice prior to the E.U. Council's formal evaluation of the GDPR,¹⁷³ which was released on January 15, 2020.¹⁷⁴ The European Commission also released an evaluation and review of the first two years of GDPR enforcement.¹⁷⁵ Time will provide opportunities to review the successes and challenges of the GDPR across the European Union, informing and improving harmonization efforts going forward.

Yet certain immeasurable factors that data cannot track—specifically a country's political climate—may be the most indicative of the future GDPR challenges the European Union could face. Fines are not imposed in vacuums; they are imposed by real people with existing loyalties and motivations. Although France sent a strong message to the AdTech industry when it fined Google heavily for not clearly explaining its online advertising practices, it also closed three cases against small French AdTech firms with only a request that the firms update their customer consent collection methods.¹⁷⁶ Ireland has been the hotspot for pending litigation against large U.S. technology companies, but the country abstained from issuing a single fine for two years.¹⁷⁷ As a final example of politics in action, the United Kingdom was engulfed in Brexit turmoil even before the GDPR became enforceable. In January 2020, the United Kingdom formally withdrew from the European Union.¹⁷⁸ Uncertain as to whether it was going to remain an E.U. member state, it is possible that the United Kingdom issued large fines partially to send a message that its data protection standards are high and deserving of an adequacy decision as a sovereign nation. Interestingly, two of the largest fines imposed by the United Kingdom Information Commissioner's Office were ultimately reduced significantly at the end of 2020 despite the fact that the European

173. Müge Fazlioglu, *GDPR in the Eyes of the Member States*, INTERNATIONAL ASSOCIATION OF PRIVACY PROFESSIONALS (Oct. 16, 2019), <https://iapp.org/news/a/gdpr-in-the-eyes-of-the-member-states>.

174. *Council Position and Findings on the Application of the General Data Protection Regulation (GDPR)—Adoption*, COUNCIL OF THE EUROPEAN UNION (Jan. 15, 2020), <https://data.consilium.europa.eu/doc/document/ST-14994-2019-REV-2/en/pdf>.

175. *Data Protection as a Pillar of Citizens' Empowerment and the EU's Approach to the Digital Transition—Two Years of Application of the General Data Protection Regulation*, EUROPEAN COMMISSION (June 24, 2020), https://ec.europa.eu/info/sites/info/files/1_en_act_part1_v6_1.pdf.

176. What the CNIL's Recent Decisions Involving Vectuary, Fidzup, Teemo and Singlespot Reveal About What a Consent UI Should Look Like, ONETRUST (Dec. 14, 2018), <https://www.onetrust.com/cnil-vectuary-fidzup-teemo-singlespot-what-it-means/#>.

177. Colm Keena, *Tulsa Becomes First Organization Fined for GDPR Rule Breach*, THE IRISH TIMES (May 17, 2020), <https://www.irishtimes.com/news/crime-and-law/tulsa-becomes-first-organisation-fined-for-gdpr-rule-breach-1.4255692>.

178. *See generally Brexit: All You Need to Know About the UK Leaving the EU*, BBC (Feb. 17, 2020), <https://www.bbc.com/news/uk-politics-32810887>.

Commission is still finalizing the United Kingdom's adequacy request.¹⁷⁹ In the interim, data flows between the United Kingdom and European Union can continue until June 30, 2021, pursuant to the E.U.-U.K. Trade and Cooperation Agreement.¹⁸⁰ The above examples expose some of the intangible truths behind the data, and their importance should not be overlooked. No matter how standardized the European Union tries to make data protection processes, E.U. member states will continue to have their own values, motivations and challenges that will undoubtedly affect their individual approaches to GDPR enforcement.

VI. CONCLUSION

The number of administrative fines imposed increased almost eightfold from 2018 to 2019, and ninety-three fines were imposed in the first quarter of 2020 alone. Some types of infringements resulted in more fines than others. It may be because those violations are broadly defined, as is the case with the “lawful, fair, and transparent” data processing principle, or the violations are particularly obvious to the user, as is true of the data subject's right to “access” his or her data. Certain countries continue to lead the charge with enforcement, like Germany and France. Their high GDPs and strong history of privacy protection resulted in a vigorous implementation approach that was not hindered by a lack of resources or infrastructure. The technology sector was issued a number of fines, and they were sizeable, but there were also a substantial number of “little guys,” like schools, political campaigns, restaurants, and private individuals fined for noncompliance.

The GDPR is a massive regulation that requires time to implement. The European Union is striving for the GDPR in its entirety, and specifically the administrative fines, to be implemented in a consistent manner across all twenty-seven E.U. member states and three EEA countries. Each iteration of E.U. privacy legislation has pushed for improved harmonization in data protection practices across the European Union. The Article 29 Data Protection Working Party suggested that DPAs have regular exchanges including “case-handling workshops or other events” to allow countries to compare their treatment of cases and resulting fines and corrective measures.¹⁸¹

179. See European Union, *Adequacy Decisions*, OFFICIAL WEBSITE OF THE EUROPEAN UNION, https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en (last visited Apr. 23, 2021).

180. European Union, *Data protection: European Commission launches process on personal data flows to UK*, OFFICIAL WEBSITE OF THE EUROPEAN UNION, https://ec.europa.eu/commission/prescorner/detail/en/ip_21_661 (last visited Apr. 23, 2021).

181. Article 29 Data Protection Working Party, *supra* note 16, at 17.

It will take time for E.U. member states to establish these collaborative methods and channels of communication and build out the resources necessary for member states to work together efficiently.

Early obstacles to harmonization—lack of resources, country-specific implementation approaches, and national privacy legislation timelines—will have less impact on enforcement actions over time. For example, it is difficult for a country to actively enforce the GDPR when it is still deciding which derogations to implement. Once national legislation is passed, DPAs will have the ability to fully enforce the GDPR and work through issues that arise. Cross-border cooperation will also improve and case law will develop, both becoming tools for more consistent decision making. Despite anticipated short-term improvements in harmonization, variability in approaches to GDPR enforcement will remain a long-term challenge due to member states' differing cultural, philosophical, and political opinions on implementation.

The GDPR has had a global impact. Countries around the world (most recently South Korea and the United Kingdom) are seeking adequacy decisions from the European Commission.¹⁸² There have also been discussions about the recently passed California Privacy Rights Act (CPRA) and whether the individual state of California could be eligible for adequacy. With an adequacy decision, personal data can flow between the “adequate” country and all E.U. member states (as well as the three EEA countries subject to the GDPR), without the use of additional data transfer safeguards.¹⁸³ One of the most watched conversations in 2021 will be the renegotiation of the E.U.-U.S. Privacy Shield, a framework for allowing data exchanges between the United States and the European Union, after the Court of Justice of the European Union invalidated the existing agreement in July 2020.¹⁸⁴ Additionally, the COVID-19 pandemic has brought to light many international privacy practices as countries attempt to monitor sick patients and enforce quarantine practices to slow the viral spread. Sensitive health information has never been more top of mind, and the balance between personal privacy and public safety is

182. OFFICIAL WEBSITE OF THE EUROPEAN UNION, *supra* note 179; European Union, Slides—Internal EU27 Preparatory Discussions on the Future Relationship: “Personal Data Protection (Adequacy Decisions), Cooperation and Equivalence in Financial Services”, OFFICIAL WEBSITE OF THE EUROPEAN UNION, https://ec.europa.eu/commission/sites/beta-political/files/seminar_20200110_-_data_protection_adequacy_-_financial_services_en.pdf (last visited Apr. 12, 2020).

183. OFFICIAL WEBSITE OF THE EUROPEAN UNION, *supra* note 179.

184. See C-311/18, *Data Protection Commissioner v. Facebook Ireland Ltd. and Maximilian Schrems*, COURT OF JUSTICE OF THE EUROPEAN UNION (July 16, 2020), <http://curia.europa.eu/juris/document/document.jsf?text=&docid=228677&pageIndex=0&doclang=en>.

precarious. As privacy laws continue to develop and evolve around the globe, there will be many more voices offering input regarding the interpretation and operationalization of E.U. data protections standards.

Fines have been a huge focus since the GDPR became law, but the success of the GDPR is not just defined by fines—the most important question is how the GDPR shapes the future of data privacy. Are administrative fines substantively enforced, or are they merely punitive penalties with little impact on companies' actual behavior? Certain countries, particularly Germany and France, have aggressively imposed fines and taken actions to solidify what it means to “comply” with different provisions of the Regulation. Ireland, comparatively, is the European hub for many large technology companies whose practices are hotly contested, but it has done little to direct the behavior of those companies. If Ireland were to impede the actions of those companies, its economy could be negatively impacted, thus at least partially explaining its inaction. Many of these large technology companies also have deep pockets and are willing to endure expensive litigation. But if some countries are much easier for businesses to operate in than others, businesses will flock to the countries where their operations are least impacted. That could have a ratcheting effect on harmonization, possibly causing countries who are substantively strong on enforcement to suffer economically. Most businesses would prefer to pay a fine and retain the ability to use personal data how they wish over an order to implement restrictive, costly practices. The E.U. member state divide between punitive and substantive enforcement has already begun, and we predict it will continue to grow.

This Note has framed the processing of personal data as a focus of regulation, but the GDPR exists in a larger, complex, growing regulatory environment. All indications point to a desire to impose more rules on companies, especially high technology companies. The European Union's Digital Services Act, aimed at improving safety on digital platforms, is in progress.¹⁸⁵ The goal of the European Union's new ePrivacy Regulation is to safeguard the privacy of electronic communications.¹⁸⁶ The European Commission recently released proposals for an Artificial Intelligence Regulation and a complementary Machinery Regulation, outlining health and

185. European Union, *The Digital Services Act: Ensuring a Safe and Accountable Online Environment*, OFFICIAL WEBSITE OF THE EUROPEAN UNION, https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en#what-are-the-next-steps (last visited Apr. 23, 2021).

186. Müge Fazlioglu, *Next-gen privacy: Examining the EU's ePrivacy Regulation*, INTERNATIONAL ASSOCIATION OF PRIVACY PROFESSIONALS (Feb. 17, 2021), <https://iapp.org/news/a/nextgen-privacy-the-eus-eprivacy-regulation/>.

safety requirements for a wide range of machinery products.¹⁸⁷ These are just three examples of GDPR adjacent legislation that work to limit the freedom of technology companies. In many ways, the GDPR accelerated these other regulatory regimes because it forced companies to be more transparent about how they use customers' personal information, exposing the widespread existence of objectionable business practices. We could move from a world where technological innovation drives progress and policy to one where each technology product or service has specified rules and requirements, with no way to circumvent those specifications. Small companies are already disproportionately burdened by GDPR compliance requirements. Furthermore, consent-based privacy models like the GDPR are difficult to apply to emerging technologies like artificial intelligence, machine learning, and facial recognition. The very regulations Europe is creating to protect its citizens could have the unintended consequence of stifling the growth and creativity of the technology sector. The durability and adaptability of the GDPR will be an important benchmark as we continue to make decisions that shape the future of data privacy.

187. European Union, *Europe Fit for the Digital Age: Commission Proposes New Rules and Actions for Excellence and Trust in Artificial Intelligence*, OFFICIAL WEBSITE OF THE EUROPEAN UNION, https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682 (last visited Apr. 23, 2021).

EXPLAINING OPAQUE AI DECISIONS, LEGALLY

Walter A. Mostowy[†]

I. INTRODUCTION

Consider the case of a man fired by artificial intelligence (AI). He had been working as an Uber driver for two years and was highly rated by riders.¹ Then, one day, Uber's AI flagged his account, and he was terminated.² The only information Uber gave him was that he was accused of "fraudulent activities"; no further explanation was provided.³ Pleading with Uber for more information or a face-to-face meeting had no effect, and instead the man was forced to defend his driver's license before the city government and to find a new job.⁴ Feeling wronged, he sued,⁵ claiming that Uber's automated decision-making violated his rights under Article 22 of the General Data Protection Regulation (GDPR).⁶ When an AI makes an important decision affecting us, are we owed an individualized explanation?⁷ And if so, what form should it take?

The opacity of AI is a novel challenge to accountability and due process. Today, AI is ubiquitous.⁸ Because AI can make decisions quickly, accurately,

DOI: <https://doi.org/10.15779/Z38D21RK1X>

© 2020 Walter A. Mostowy.

† J.D., 2021, University of California, Berkeley, School of Law.

1. Mary-Ann Russon, *Uber Sued by Drivers Over 'Automated Robo-Firing'*, BBC NEWS (Oct. 26, 2020), <https://www.bbc.com/news/business-54698858>.

2. *Id.*

3. *Id.*

4. *See id.*

5. *See id.*

6. Regulation (EU) 2016/679, of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1 [hereinafter GDPR].

7. The presiding court of first instance recently held, though not on Article 22 grounds, that Uber owed the man more information. Rechtbank [Rb.] Amsterdam [District Court, Amsterdam] 11 maart 2021, JAR 2021, 96 m.nt. Rietveld, R.D. (verzoeker 1/Uber B.V.) (Neth.), ECLI:NL:RBAMS:2021:1018, ¶ 4.29; *see also* Rb. Amsterdam 11 maart 2021, ECLI:NL:RBAMS:2021:1019 (verzoeker 1/Ola Netherlands B.V.) (Neth.) ¶¶ 4.41, 4.52 (requiring an explanation of the logic behind an automated decision pursuant to Article 15 in view of Article 22).

8. Amina Adadi & Mohammed Berrada, *Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)*, 6 IEEE ACCESS 52,138 (2018).

and at scale,⁹ it has become indispensable and has come to influence much of our lives, from low-stakes web searches and product recommendations to high-stakes employment, credit scoring, education, and criminal justice.¹⁰ However, AI is not infallible; it can still make mistakes, discriminate,¹¹ offend,¹² or be otherwise unfair or biased.¹³ Unfortunately, today's AI is so complicated, it is difficult to understand its reasoning or identify and fix errors in its decisions.¹⁴ For this reason, AI is often called a “black box”¹⁵—that is, AI is opaque. Research on methods of explaining AI is only just beginning.¹⁶

This is now an urgent problem in light of the GDPR. The European Union's GDPR is at the vanguard of the effort to enforce transparency and accountability in AI decision-making.¹⁷ As part of the GDPR's comprehensive oversight mechanisms, AI controllers are required to provide extensive ex ante

9. See Bryan Casey, Ashkon Farhangi & Roland Vogl, *Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise*, 34 BERKELEY TECH. L.J. 143, 149 (2019); Maja Brkan, *Do Algorithms Rule the World? Algorithmic Decision-making and Data Protection in the Framework of the GDPR and Beyond*, 27 INT'L J.L. & INFO. TECH. 91, 92 (2019).

10. See Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For*, 16 DUKE L. & TECH. REV. 18, 19 (2017); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 1–4 (2014); FRANK PASQUALE, *THE BLACK BOX SOCIETY* 4–5 (2015); Brkan, *supra* note 9, at 92; Will Knight, *The Dark Secret at the Heart of AI*, MIT TECH. REV. (Apr. 11, 2017), <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai>.

11. See, e.g., Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (finding that software used to predict recidivism risks was discriminatory against Black people).

12. See, e.g., Jessica Guynn, *Google Photos Labeled Black People 'Gorillas'*, USA TODAY (July 1, 2015, 2:10 PM), <https://www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465>.

13. See Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson & Harlan Yu, *Accountable Algorithms*, 165 U. PA. L. REV. 633, 680 (2017); Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 68–81 (2019); Allan E. Holder, Note, *What We Don't Know They Know: What to Do About Inferences in European and California Data Protection Law*, 35 BERKELEY TECH. L.J. 1357, 1361[(2020).

14. See *infra* Section II.A and accompanying notes.

15. See, e.g., Adadi & Berrada, *supra* note 8; Edwards & Veale, *supra* note 10, at 18; Knight, *supra* note 10.

16. See *infra* Sections II.B–C and accompanying notes.

17. The European Commission has also recently proposed a comprehensive regulation of AI that, if passed, may prove similarly important for AI. See *Commission Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM (2021) 206 final (Apr. 21, 2021).

explanations—explanations of how the AI system works in general—that are furnished to authorities and auditors. This obligation arises primarily from Articles 13–15, which guarantee the right to “meaningful information about the logic involved”¹⁸ and call for information about the automated system’s design as a whole. Yet scholars still disagree as to whether the GDPR also provides data subjects the right to an ex post explanation: that is, an individualized explanation of a single decision. Moreover, guidance is needed as to what form such explanations should take.¹⁹

This Note argues that Article 22 of the GDPR, which guarantees the “right . . . to contest” automated decisions,²⁰ does indeed establish a right to an ex post explanation of AI decisions. This Note then draws on Article 22’s underlying principle of due process to examine methods of explaining AI decisions and recommends a few types most likely to satisfy Article 22.

This Note is organized as follows. Part II gives an overview of how AI works and what makes AI opaque, explores the technical research into explaining AI, and discusses scholarship providing clues as to what may be desirable in an explanation. Part III gives an overview of the relevant parts of the GDPR, catalogues the scholarship on the “right to explanation,” and argues that Article 22 of the GDPR requires an ex post explanation. The argument examines the GDPR itself, the views of its supervisory authorities, and scholarship on contestation and due process. Finally, Part IV draws on lessons from principles underlying due process, contestation, and cross-examination to conclude that experimentation-based and counterfactual explanations are most aligned with the principles underlying Article 22.

II. AI, EXPLAINABLE AI, AND THEIR LIMITATIONS

Scholars and laypeople alike are now attuned to the growing influence and potential risks of AI. However, both scholarly work and popular media largely lack a substantive understanding of how AI works and what its limitations are. Such an understanding is critical to identifying satisfactory solutions and guiding future research.²¹ In this Part, I hope to help bridge this gap in

18. See GDPR, *supra* note 6, arts. 13–15.

19. Raphaël Gellert, Marvin van Bekkum & Frederik Zuiderveen Borgesius, *The Ola & Uber Judgments: For the First Time a Court Recognises a GDPR Right to an Explanation for Algorithmic Decision-making*, EU LAW ANALYSIS (Apr. 29, 2021, 7:06 AM), <https://eulawanalysis.blogspot.com/2021/04/the-ola-uber-judgments-for-first-time.html>.

20. See GDPR, *supra* note 6, art. 22.

21. Indeed, failure to consider the technical ramifications of regulations places the rights of data subjects at risk. See Antoni Roig, *Safeguards for the Right Not to Be Subject to a Decision Based Solely on Automated Processing (Article 22 GDPR)*, 8 EUR. J.L. & TECH., no. 3, 2017, at 10.

understanding. Section II.A provides a gentle overview of what AI is, how it works, and how it gives rise to a tradeoff between accuracy and interpretability. Section II.B gives an overview of the nascent Explainable AI field of technical research, including both its menu of methods of explanation and its fundamental challenges. Finally, Section II.C discusses legal and scientific scholarship that provides important clues as to which methods of explanation may be most desirable from a human perspective.

A. A GENTLE INTRODUCTION TO AI AND THE ACCURACY–
INTERPRETABILITY TRADEOFF

AI at its core is about understanding and building machines that exhibit intelligence.²² “Intelligence” is the ability to reason one’s way to success; it is exhibited when one applies a body of knowledge to a new context in order to achieve a goal.²³ AI has proved most successful when it automates the process of using data to build a knowledge base. In effect, the computer learns from the data.²⁴ This learning process is called, appropriately enough, machine learning.

Machine learning (ML) powers²⁵ most of what is considered²⁶ AI today. As practitioners adapted AI to many problem domains, they adapted ML by

22. STUART J. RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 1, 18 (3d ed. 2010).

23. This is the rational-actor conception of intelligence. *See id.* at 4. There also exist competing conceptions. *See id.*

24. Some of the earliest efforts at building AIs actually attempted to build a knowledge base by hand, with the idea that the AI could simply deduce useful conclusions from its knowledge. In practice, knowledge built by hand did not make for successful AI. However, in theory, with enough data, such an AI would never need to learn anything new—it would be able to calculate the answer to life, the universe, and everything. (That answer, of course, is forty-two. DOUGLAS ADAMS, *THE HITCHHIKER’S GUIDE TO THE GALAXY* (Del Rey Books ed., Pan Books 2017) (1979).)

25. *See* Brian Fung, *Everything You Think You Know About AI Is Wrong*, WASH. POST (June 2, 2016, 3:30 AM), <https://www.washingtonpost.com/news/the-switch/wp/2016/06/02/everything-you-think-you-know-about-ai-is-wrong> (explaining how ML has exponentially grown, but other kinds of AI have not). Indeed, outside of the technical literature, AI and ML are often conflated or treated as interchangeable. *See, e.g.*, James Vincent, *What Counts as Artificially Intelligent? AI and Deep Learning, Explained*, THE VERGE (Feb. 29, 2016, 3:40 PM), <https://www.theverge.com/2016/2/29/11133682/deep-learning-ai-explained-machine-learning>.

26. Interestingly, what is popularly considered AI has shrunk over time. Whenever a type of AI technology improves enough to enter mainstream use, many users cease to think that technology is intelligent enough to count as AI. This is called the “AI effect.” Michael Haenlein & Andreas Kaplan, *A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence*, 61 CAL. MGMT. REV. 5, 6 (2019). Some now question whether, for example,

inventing many kinds of ML, called models. Each model has a different design and thus introduces different assumptions and restrictions.²⁷ Importantly, models lie along a spectrum of understandability—some models are more readily understandable than others.²⁸ To illustrate, consider the following two ML models, one at each end of the spectrum.

Linear regression²⁹ is a simple type of ML. It represents a best-fit line through a set of training data points. Typically, the points do not all fall precisely on the line; instead, the goal is to minimize the points' collective distance from the line, called the loss. When the model learns, it adjusts the line to reduce the loss. The model can then make predictions by finding points along the best-fit line that match the input data. For example, suppose that a linear regression model is used to predict housing prices based on floor area. The training data might include sales in the past year. This data will not fall along a perfect line; nonetheless, the linear regression can draw a line through the data points that best fits them. Then, given a new floor area number, the model can find the point on the line that matches the floor area number and output the price. There are fancier versions of the model—for example, the model can take into account additional variables, such as year of construction or the number of parks nearby, or it can find a best-fit curve rather than a best-fit line—but the core idea remains the same.

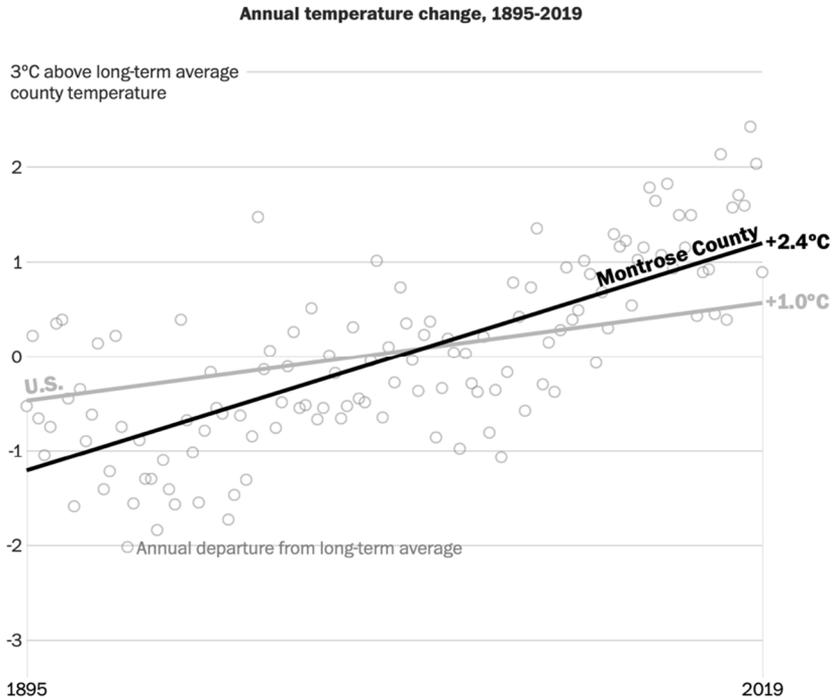
pathfinding, handwriting recognition, and even chess engines exhibit enough “intelligence” to qualify as AI. See RUSSELL & NORVIG, *supra* note 22, at 27. The AI effect parallels the “God of the gaps” concept, in which the role of God is confined to gaps in scientific understanding and thus retreats when science advances. See Robert Larmer, *Is There Anything Wrong with “God of the Gaps” Reasoning?*, 52 INT’L J. PHIL. RELIGION 129, 136 (2002).

27. Assumptions—called “bias” in the technical literature—are required for any form of learning. This is true for humans and machines alike. Without bias, an AI cannot learn or predict beyond the specific data points it is given. A problem fundamental to ML is finding the right types of bias and the right amount of bias to use.

28. See David Bamman, *Interpretability in Human-Centered Data Science*, 2016 CSCW WORKSHOP ON HUMAN-CENTERED DATA SCI. 2–3, available at https://cscw2016hcds.files.wordpress.com/2015/10/bamman_hcds.pdf.

29. See generally RUSSELL & NORVIG, *supra* note 22, at 717–27 (explaining the mathematical mechanics of linear regression).

Figure 1: A Linear Regression for Annual Measurements of Temperature in One U.S. County Versus Time³⁰



Relative to other ML models, linear regression is readily understandable. Indeed, best-fit lines are occasionally used as graphics in popular news media.³¹ When a linear regression takes input data and produces an output, the explanation is that the output falls along the best-fit line. The line, in turn, is determined by the training data in a straightforward way: it is the line that “best fits” the training data. It is easy to predict the model’s behavior simply by looking at the line. See, for example, Figure 1. It is easy to see that the best-fit line is the one labeled “Montrose County” and not the one labeled “U.S.” It is also easy to understand what the model would predict the temperature to be a hundred years in the future. If the model’s behavior is objectionable, it is easy to examine its inner workings: Is the output truly on its best-fit line? Has the

30. Juliet Eilperin, *This Giant Climate Hot Spot Is Robbing the West of its Water*, WASH. POST (Aug. 7, 2020), <https://www.washingtonpost.com/graphics/2020/national/climate-environment/climate-change-colorado-utah-hot-spot/>.

31. See, e.g., *id.* (showing a trend of warming over time as measured in Colorado, United States); The Learning Network, *What’s Going On in This Graph? | Dec. 4, 2019*, N.Y. TIMES, <https://www.nytimes.com/2019/11/26/learning/whats-going-on-in-this-graph-dec-4-2019.html> (last updated Dec. 11, 2019) (comparing various crime rates to the number of undocumented immigrants).

model correctly fit a line to the training data? Are there gaps in the training data? Is the training data well represented by a line? Experts and non-experts alike can grasp these concepts and query the inner workings of the model. As a result, when things go wrong, anyone can analyze the model and find clues as to possible causes.

In contrast, a neural network³² is a complex type of ML that is difficult to understand, even for the experts who built it. A neural network is a type of ML that is loosely based on the brain, in which neurons “activate,” or send a signal through a connecting synapse, thereby causing other neurons to activate in turn. A neural network consists of several layers of neurons. The first layer receives inputs, causing some of them to send activation signals; the next layer receives these signals, causing some of them to activate in turn; and so on, until the final layer receives inputs and activates, thereby providing an output. Each individual neuron is a fairly simple mathematical function: it takes inputs, computes a numerical sum based on an internal configuration, and gives an output. For inputs, the neuron receives signals from all of the neurons in the previous layer. For the configuration, the neuron is configured to assign a different level of importance—a “weight”—to each of these connections, influencing how much heed it pays to any signals it receives. It also has an activation “threshold.”³³ And for the output, if the weighted sum of inputs crosses that threshold, the neuron activates. If the sum goes well beyond the threshold, the neuron can activate more strongly.³⁴

The final layer is the output, and each neuron in the final layer is manually assigned meaning. For example, if the neural network is intended for image recognition, one neuron might be assigned “bird,” another “elephant,” and another “dog.” When the “dog” neuron fires, the network thinks the input

32. For an entertaining and approachable explanation of neural networks, see generally 3Blue1Brown, *But What Is a Neural Network? | Deep Learning, Chapter 1*, YOUTUBE (Oct. 5, 2017), <https://www.youtube.com/watch?v=aircArugnKk> [hereinafter 3Blue1Brown Chapter 1]; 3Blue1Brown, *Gradient Descent, How Neural Networks Learn | Deep Learning, Chapter 2*, YOUTUBE (Oct. 16, 2017), [https://www.youtube.com/watch?v=IHZwWFHWa-w](https://www.youtube.com/watch?v=IHZwWFHWa-w;); 3Blue1Brown, *What is Backpropagation Really Doing? | Deep Learning, Chapter 3*, YOUTUBE (Nov. 3, 2017), <https://www.youtube.com/watch?v=Il3gGewQ5U> [hereinafter 3Blue1Brown Chapter 3]. For a technical approach, see generally RUSSELL & NORVIG, *supra* note 22, at 727–37.

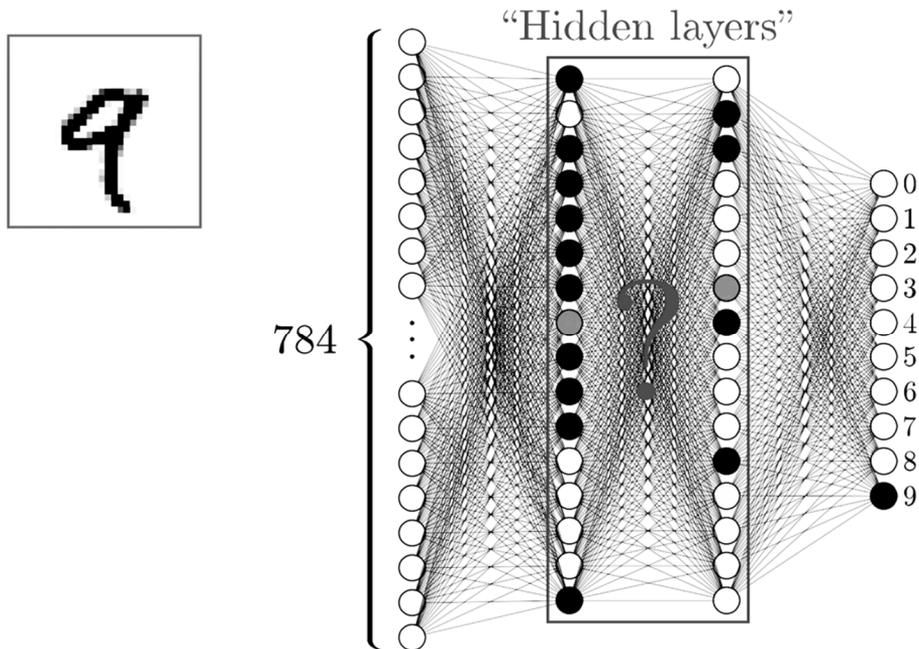
33. In the literature, mostly for reasons of mathematical simplicity, the threshold is indirectly controlled by a parameter unfortunately called “bias.” To avoid nomenclature confusion with the notions of bias more familiar to the general public and discussed in this Note, this Note instead refers to the “bias” parameter as the threshold.

34. There are several different “activation function[s]” that a neural-network architect can choose. See RUSSELL & NORVIG, *supra* note 22, at 729. The one described here is called ReLU. See 3Blue1Brown Chapter 1, *supra* note 32, at 17:30–18:25.

image is a dog. The neuron can fire more strongly to indicate higher confidence in the image being a dog.

See Figure 2 for a visual example. It depicts a neural network used to recognize a handwritten numerical digit in an input image. It has an input layer with 784 neurons, two hidden layers of 16 neurons each, and an output layer with 10 neurons. Each input neuron receives one pixel of the input image. The input image is in the top left. In this case, the output neuron assigned to indicate “9” is activating. Thus, “9” is the output.

Figure 2: A Neural Network Trained to Recognize Handwritten Digits³⁵



As with any other ML model, a neural network “learns” during an automated process called training. To simplify, the neural network training process works as follows. The neural network is given a large data set to train on, such as a database of images of birds, elephants, and dogs. Then, the neural network learns from each individual image, one by one. Suppose the first image is of a dog. When the dog’s image is fed into the first layer, the activations percolate through the in-between layers³⁶ (called the “hidden

35. 3Blue1Brown Chapter 1, *supra* note 32, at 4:06 (color adjusted).

36. This unidirectional percolation is why this type of network is called a “feed-forward” network. See RUSSELL & NORVIG, *supra* note 22, at 729.

layers”) to the output layer, resulting in a classification. Because the neural network is not well trained yet, that classification is probably wrong—perhaps the neural network thinks the image of a dog is “bird.” The neural network then compares this output to the correct output (the output here should be “dog,” not “bird”) and makes small adjustments to all of the neurons’ parameters—the weights and the thresholds—to achieve a slightly better outcome.³⁷ This process is then repeated for the entire training set, again and again, until overall accuracy no longer improves. The hope of this learning process is that each successive layer of neurons learns more sophisticated features.³⁸ For example, perhaps each neuron in the first layer learns to recognize a line or a curve, each neuron in the second layer considers these simple features and learns to recognize a head or a trunk or a wing, and the final layer considers these complex features and learns to recognize a bird or a dog.³⁹

Because neural networks are so complex, they are quite opaque to human understanding. For one thing, the sheer complexity of the design makes it difficult for ordinary people to understand. But neural networks are opaque even to their expert designers. Even when the network seems to be generally performing well, instead of learning to recognize a line or wing, neurons often learn something mysterious that does not seem meaningful to humans.⁴⁰ But because of the sheer complexity of the system, it is unclear whether that mysterious thing is a useful pattern or useless nonsense.

In addition, neural networks are more difficult to examine than logistic regression models are. In a logistic regression, we can ask questions such as whether the model correctly fit a line to the training data, or whether the training data is well represented by a line. What would be the equivalent inquiries for a neural network? Even experts do not know. While it is still possible to examine and find gaps in the training data,⁴¹ there are few tools of inquiry into neural networks.

37. This learning process is called “back-propagation.” See RUSSELL & NORVIG, *supra* note 22, at 733.

38. See 3Blue1Brown Chapter 1, *supra* note 32, at 5:30–8:39.

39. For a visualization, see, for example, Yariv Adan, *Do Neural Networks Really Work Like Neurons?*, START IT UP (Sept. 29, 2018), <https://medium.com/swlh/do-neural-networks-really-work-like-neurons-667859dbfb4f>.

40. See, e.g., 3Blue1Brown Chapter 3, *supra* note 32, at 14:02–16:40.

41. This can result in “uncertainty bias,” in which a risk-averse AI avoids awarding value to people it is uncertain about due to gaps in the training data. See Bryce Goodman & Seth Flaxman, *European Union Regulations on Algorithmic Decision Making and a “Right to Explanation”*, AI MAG., Fall 2017, at 50, 54. It can have other undesirable effects as well. For example, Google blamed gaps in training data for an incident in which its image recognition software classified Black people as gorillas. See Guynn, *supra* note 12.

The opacity of neural networks is no longer a mere theoretical concern; “deep” neural networks, even more complicated than traditional neural networks, have seen massive growth in use in the last decade. A deep neural network is essentially a neural network with more than two hidden layers of neurons.⁴² This added complexity makes the network difficult to train, but this difficulty has been overcome with various techniques, such as reorganizing the neurons and using very large training sets.⁴³ Coupled with today’s increased computing power and larger data sets, deep neural networks come with significant advantages: not only do the added layers in theory mean that the network can learn more complicated relationships, but deep neural networks can more easily learn from unlabeled data⁴⁴—for example, from a set of images that are not known beforehand to be a “bird” or a “dog.” Given a large enough set of images, a deep neural network can discover patterns of its own accord. For example, an early deep neural network trained on millions of YouTube video thumbnails learned to recognize cats.⁴⁵ Deep neural networks are now deployed in many areas, including speech recognition, image recognition, and object detection.⁴⁶ Yet they remain highly complex and difficult for humans to understand.

This opacity of deep AI is a real problem. Opacity not only impedes identifying and fixing errors, it impedes understanding and trust of the model. Deep neural networks have many layers of highly complicated logic, which makes it very difficult to identify the reasoning behind their decisions.⁴⁷ For example, an actual deep neural network performed so well in image recognition, detecting objects in images, that it won an AI competition—but only after that win did scientists eventually realize that the AI didn’t detect a

42. See Michael A. Nielsen, *Why Are Deep Neural Networks Hard to Train?*, NEURAL NETWORKS & DEEP LEARNING (Dec. 26, 2019, 3:26 PM), <http://neuralnetworksanddeeplearning.com/chap5.html>.

43. See, e.g., Quoc V. Le, Marc’Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg S. Corrado, Jeff Dean & Andrew Y. Ng, *Building High-Level Features Using Large Scale Unsupervised Learning*, 29 PROC. INT’L CONF. ON MACHINE LEARNING 507, 507 (2012).

44. See *id.* at 1–2.

45. John Markoff, *How Many Computers to Identify a Cat? 16,000*, N.Y. TIMES (June 25, 2012), <https://www.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html>; Le et al., *supra* note 43.

46. Yinpeng Dong, Hang Su, Jun Zhu & Fan Bao, *Towards Interpretable Deep Neural Networks by Leveraging Adversarial Examples 1* (Aug. 18, 2017) (unpublished manuscript), available at <https://arxiv.org/pdf/1708.05493.pdf>.

47. See Wojciech Samek & Klaus-Robert Müller, *Towards Explainable Artificial Intelligence*, in EXPLAINABLE AI: INTERPRETING, EXPLAINING AND VISUALIZING DEEP LEARNING 5, 12 (Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen & Klaus-Robert Müller eds., 2019).

boat by looking for a boat, but rather by detecting the presence of water.⁴⁸ Similarly, due to the particular training data used, the same AI also accidentally learned to recognize trains by tracks, and horses by copyright watermark.⁴⁹ This AI isn't unique; any opaque AI trained on large datasets is at risk of learning shortcuts like these.⁵⁰ For example, current deep neural networks tend to distinguish wolves from huskies mainly by the presence of snow.⁵¹ In these toy examples, the errors are amusing and harmless. But when human lives or rights are at stake, AIs must not learn shortcuts.⁵²

Unfortunately, this problem will not be solved in the near future. Deep neural networks are the opaque type of ML model, but they remain popular because they remain the most accurate performers. Thus, the opacity that hinders AI's improvement, validation, rectification, and trustworthiness⁵³ will continue to be a problem.

Indeed, there seems to be an inherent tradeoff between how understandable an AI's behavior is and how accurate that AI is.⁵⁴ Simple models like linear regression "can only represent simple relationships but are easy to interpret," whereas powerful, complex models like deep neural networks "can represent a rich class of functions but are hard to interpret."⁵⁵ This tradeoff seems to be a general property of AI models: in order for an AI model to be more understandable, it must be made less complex, but doing so hurts its predictive performance—that is, how accurate it is.⁵⁶ This is because making accurate predictions about complex relationships requires a complex model, so it stands to reason that simpler models, while they may be more understandable to humans, are less capable of accurately capturing these relationships.⁵⁷ Moreover, deep neural network research has focused mainly

48. *Id.* at 7.

49. *Id.*

50. *See* EXPLAINABLE AI: INTERPRETING, EXPLAINING AND VISUALIZING DEEP LEARNING 146 (Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen & Klaus-Robert Müller eds., 2019).

51. Samek & Müller, *supra* note 47, at 7.

52. *See id.* (citing autonomous driving and medical applications as examples).

53. *See* Dong et al., *supra* note 46, at 1.

54. Adadi & Berrada, *supra* note 8, at 52,147; Goodman & Flaxman, *supra* note 41, at 55; Jialei Wang, Ryohei Fujimaki & Yosuke Motohashi, *Trading Interpretability for Accuracy: Oblique Treed Sparse Additive Models*, 21 ACM SIGKDD INT'L CONF. ON KNOWLEDGE DISCOVERY & DATA MINING 1245, 1245 (2015).

55. Goodman & Flaxman, *supra* note 41, at 55.

56. *See* Wang et al., *supra* note 54, at 1245.

57. *See* Adadi & Berrada, *supra* note 8, at 52,147.

on improving accuracy, at the expense of understandability.⁵⁸ As a result, deep neural networks today are both more accurate and more opaque than ever.⁵⁹

This “accuracy–interpretability” tradeoff is a fundamental problem facing AI. Simple models like linear regressions, which are just a best-fit line, are easily understandable, but not complex enough to accurately capture complex relationships in data. Complex models like deep neural networks, which have many neurons, weights, thresholds, and interrelated connections, are incredibly complex and opaque even to experts, but are so accurate that they are used in spite of their opacity. Is it possible to break out of this paradigm and invent an AI that is both highly accurate and understandable?

B. THE BURGEONING FIELD OF EXPLAINING AI

Faced with the accuracy–interpretability tradeoff of AI models and the resulting difficulty in understanding deep AI, AI researchers understood a need for more transparency and birthed a new field of research, now called Explainable AI or XAI.⁶⁰ This nascent field is currently growing at a rapid pace, making incremental progress in many different directions. This Section gives a general overview of the field of XAI, including the many technical types of explanations that have been proposed. It explains how the methods can generally be categorized along two dimensions, scope (global or local) and strategy (intrinsic versus agnostic). It also describes four particularly popular methods of explaining: visualization, knowledge extraction, influence measurement, and example generation. Finally, it describes the fundamental challenges facing the field: the lack of formal theory and the lack of focus on the human factor.

Different types of explanations are intended to serve different goals and may shed light on different aspects of the AI.⁶¹ For example, one explanation might seek to explain the “abstract concept” that an AI has learned, doing so by illustrating a prototypical example of that concept.⁶² Another explanation might seek to explain one particular decision that an AI made, doing so by illustrating which pixels of the input image were most pivotal to the decision.⁶³ Still more ambitious explanations might try to explain an AI’s general

58. *Id.* at 52,140.

59. EXPLAINABLE AI, *supra* note 50, at v; *see also* Goodman & Flaxman, *supra* note 41, at 55–56 (“Neural networks, especially with the rise of deep learning, pose perhaps the biggest challenge—what hope is there of explaining the weights learned in a multilayer neural net with a complex architecture?”).

60. EXPLAINABLE AI, *supra* note 50, at 2.

61. *See* Samek & Müller, *supra* note 47, at 10.

62. *See id.*

63. *See id.* at 10–11.

behavioral strategy, or identify the most representative data points of its training data in order to identify potential biases.⁶⁴

However, surveys of the literature seem to agree that it is useful to categorize XAI research along at least two axes: scope and strategy.⁶⁵ For the first axis, the scope of an explanation is considered to be either global or local.⁶⁶ A global explanation tries to explain the logic behind the model's decision process as a whole in the general case.⁶⁷ In contrast, a local explanation tries to explain the logic behind one particular decision that the model made.⁶⁸ Hybrid approaches combining both global and local explanation qualities are also possible.⁶⁹ Given the high complexity of neural networks, local explanations are currently the favored approach to explaining them.⁷⁰ Although local explanations cannot illuminate an AI's general decision process, they can help earn users' trust in individual decisions.⁷¹ Thus, global and local explanations may be best suited for different purposes; experts are better able to make use of a global explanation, with its complexity and abstractness, while laypeople are better suited for simpler, more concrete local explanations.

For the second axis, the strategy of generating the explanation can be either intrinsic or agnostic.⁷² In an intrinsic approach, the AI model itself is designed to be interpretable, meaning that when it takes input and generates output, either the model itself can be easily examined as an explanation, or it generates some sort of explanation along with the output.⁷³ The latter type is typically brittle, requiring significant work to continue functioning if the inner architecture of the model is changed.⁷⁴ In contrast, in an agnostic approach, a preexisting AI model is treated like a black box that just generates its normal output without explanation; instead, an additional, simplified algorithm

64. *See id.* at 11.

65. *See* Arun Das & Paul Rad, Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey 2 (June 23, 2020) (unpublished manuscript), *available at* <https://arxiv.org/pdf/2006.11371.pdf>; Adadi & Berrada, *supra* note 8, at 52,147; Samek & Müller, *supra* note 47, at 11–12.

66. Das & Rad, *supra* note 65, at 2.

67. Lisa Käde & Stephanie von Maltzan, *Towards a Demystification of the Black Box—Explainable AI and Legal Ramifications*, 2 J. INTERNET L. 3, 5 (2019); Adadi & Berrada, *supra* note 8, at 52,147, 52,151.

68. Käde & von Maltzan, *supra* note 67, at 5; Adadi & Berrada, *supra* note 8, at 52,148.

69. *See id.* at 52,148.

70. *Id.*

71. *See id.* at 52,151.

72. Das & Rad, *supra* note 65, at 2.

73. *See id.* at 3; Käde & von Maltzan, *supra* note 67, at 5.

74. *See* Das & Rad, *supra* note 65, at 3.

separate from the AI model helps to explain the AI's decision after it is made.⁷⁵ This approach is more flexible in that it can be added to a preexisting AI model,⁷⁶ but a significant drawback is that because the approach relies on a simplified algorithm, not the target AI, for an explanation, the explanations generated may be less accurate.⁷⁷ Thus, an intrinsic approach may be most desirable to aim for long-term, but given the deep neural networks already in use, agnostic approaches may be a more practical option that can be added onto the models already deployed.

Turning to concrete methods, four general methods of providing explanations have attracted much of the field's attention: visualization, knowledge extraction, influence measurement, and example generation.⁷⁸ Visualization as an approach has a natural appeal and is thus quite popular.⁷⁹ Visualizations can take many forms.⁸⁰ One form, used on neural networks, attempts to literally illustrate what individual artificial neurons have "learned."⁸¹ However, sometimes these neuron visualizations produce images that are inscrutable to humans.⁸² Another significant drawback is that current methods generally cannot guarantee a close relationship between the visualization and the model; that is, they cannot guarantee great accuracy.⁸³ More generally, visualization techniques often rely on specific datasets, rather than on "real" data, making them ill-suited to production systems.⁸⁴ One interesting variant highlights parts of the input image that are important as it learns to recognize and describe in English the image's contents.⁸⁵ Because of the intuitive appeal of illustrations and visuals, XAI research is likely to continue exploring visualizations for the foreseeable future.

The second popular method is knowledge extraction, which studies the inputs and outputs of a deep neural network in order to construct a

75. *See id.* In the literature, agnostic approaches are sometimes called "post hoc." However, to avoid nomenclature confusion later in this Note, this Note refers to this only as "agnostic."

76. *See id.*

77. *See* Adadi & Berrada, *supra* note 8, at 52,151.

78. *See id.* at 52,149–51.

79. *See id.* at 52,149.

80. *See id.*

81. Dong et al., *supra* note 46, at 1; *see* Adadi & Berrada, *supra* note 8, at 52,149.

82. Adadi & Berrada, *supra* note 8, at 52,153.

83. *See id.* at 52,149.

84. *See* Dong et al., *supra* note 46, at 1.

85. Kelvin Xu, Jimmy Lei Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard S. Zemel & Yoshua Bengio, *Show, Attend and Tell: Neural Image Caption Generation with Visual Attention*, 37 PROC. INT'L CONF. ON MACHINE LEARNING 2048 (2015) (cited by Adadi & Berrada, *supra* note 8, at 52,147).

simplification of the network's behavior.⁸⁶ One variant aims to construct "rules" of behavior that, when added together, approximate the general behavior of the network.⁸⁷ Another variant aims to construct a shallow neural network—that is, a simpler one with fewer layers of neurons—whose behavior approximates that of the deep neural network.⁸⁸ Though knowledge extraction is in its infancy, it may hold the most long-term promise of most accurately explaining the complex relationships learned by complex deep neural networks.

The third popular method is influence measurement, which aims to measure the relative weight (importance) that certain features of the inputs have on the outputs of the neural network.⁸⁹ The most straightforward application of this is feature importance: this puts a number on how much each feature of the input contributes to the ultimate decisions.⁹⁰ Another variant is sensitivity analysis: this makes small changes to the input or to the weights of the neurons and observes how the output changes.⁹¹ This does not amount to an explanation of any particular output, but researchers have used it to identify unimportant inputs and to test the robustness of the model.⁹² Theoretically, it may also be useful as part of a future, more comprehensive explanation technique.⁹³

To understand the difference between feature importance and sensitivity analysis, consider an AI trained to assess creditworthiness, like FICO's Credit Risk Models, often used by lenders in the United States.⁹⁴ The FICO model takes as input a consumer's credit history, such as payment history, amounts owed, the length of credit history, and so on.⁹⁵ The model then outputs a score representing the likelihood that the consumer would pay back a loan.⁹⁶ An analysis of feature importance might take a long list of inputs (consumers' data) and outputs (FICO scores) and run a statistical analysis in order to quantify how much payment history affects the final score, versus how much the debt ratio affects the final score, and so on. In fact, FICO publishes numbers to this effect: payment history is said to account for thirty-five percent of the

86. See Adadi & Berrada, *supra* note 8, at 52,149–50.

87. *Id.* at 52,149.

88. *Id.* at 52,150.

89. *Id.*

90. *Id.*

91. *Id.*

92. *Id.*

93. See *id.*

94. *Id.* at 52,139; Alexandria White, *What Is a FICO Score and Why Is it Important?*, CNBC (Mar. 19, 2021), <https://www.cnbc.com/select/what-is-fico-score>.

95. White, *supra* note 94.

96. *Id.*

score, amounts owed thirty percent, and length of credit history fifteen percent.⁹⁷ This gives a general sense of which factors are important to one's credit history. A sensitivity analysis, by contrast, might take a single consumer's data, observe the output FICO score, then add or subtract a few late payments, or increase or decrease the age of credit lines, to observe how much the score improves or worsens. This may reveal, for example, that in one particular consumer's situation, maintaining credit lines for a few more years will be sufficient to raise their FICO score to an acceptable level. Or, if used as a robustness check, sensitivity analysis can identify a problem with the model if adding an additional late payment paradoxically increases the FICO score. In fact, FICO offers a Credit Scores Estimator tool to the public that allows consumers to play with their own data and perhaps run a sensitivity analysis for themselves.⁹⁸

The fourth popular method is example generation, which seeks to provide an explanation by way of example.⁹⁹ There are two main variations. The first variation selects "prototypes," or sets of data points (inputs and their outputs) from the training data that are representative of that data.¹⁰⁰ This is a bit like automatically selecting, say, fifteen to thirty loan applicants out of five hundred total as a way of demonstrating the prototypical types of people who successfully and unsuccessfully apply for a loan. The second variation is the counterfactual: this method starts with a single data point and endeavors to describe the minimum alterations that would lead to a different output.¹⁰¹ For example, it might consider one of the unsuccessful loan applicants above and propose improvements, such as three fewer late payments, that would have led to the loan application being successful. Like visualizations, example generation has an intuitive appeal, but example generation has the additional advantage of having an immediately practical application for people affected by an AI decision, such as the denial of a loan.

Despite this promising menu of methods to further research and develop, researchers recognize that significant challenges remain ahead for XAI.¹⁰² There are two challenges in particular. First, despite the point of explanations being to benefit humans, XAI research has exhibited a distinct lack of focus

97. *Id.*

98. *FICO Score Estimator*, MYFICO, <https://www.myfico.com/fico-credit-score-estimator/estimator> (last visited May 27, 2021).

99. Adadi & Berrada, *supra* note 8, at 52,150.

100. *Id.*

101. *Id.* at 52,150–51.

102. Samek & Müller, *supra* note 47, at 16.

on human factors.¹⁰³ Explanations often remain at a low abstraction level; in other words, they offer technical details like neuron weights or sets of pixels but fail to synthesize and summarize these details as higher-level human concepts.¹⁰⁴ Indeed, the role that human psychology plays in an AI-explanation system is poorly understood in the field.¹⁰⁵ As a result, most XAI research ignores the intended human beneficiaries and instead relies on the flawed intuitions of the researchers.¹⁰⁶ Second, formal theory is simply lacking. There is no consensus on how to measure the quality of an explanation.¹⁰⁷ More fundamentally, there is no rigorous definition of what an explanation is or what it even means to “explain” or “understand.”¹⁰⁸ Developing these fundamentals will take significant time and effort.¹⁰⁹

Nonetheless, there is good reason to begin deploying AI explanations even in the face of these shortcomings. Explanations of AI are essential for AI to earn people’s trust in their logic and decisions. Experimental studies confirm that explanations impact trust.¹¹⁰ Oversimplified explanations lose users’ trust,¹¹¹ but better explanations can promote the acceptance of unfavorable decisions, help establish informed consent, and build trust and understanding.¹¹² We should choose from the menu available to us today, even if the XAI field has more maturing to do.

In sum, the field of Explainable AI offers a menu of potential methods of generating explanations of AI. They can be global or local, with global explanations perhaps better suited to experts and local ones to laypeople. They can be intrinsic or agnostic, with intrinsic explanations a long-term goal but agnostic ones a more practical option. Visualizing, extracting knowledge, measuring the influence of input features, and generating examples are four popular methods of generating explanations. Explanations can benefit experts and laypeople alike. But with formal theory lacking and no consensus on what it means to “explain” or “understand,” the field is quite unsure which types of explanation from the menu may be best.

103. Adadi & Berrada, *supra* note 8, at 52,153; *see also* Samek & Müller, *supra* note 47, at 17.

104. Samek & Müller, *supra* note 47, at 16.

105. Adadi & Berrada, *supra* note 8, at 52,156.

106. *Id.* at 52,153.

107. *See* Samek & Müller, *supra* note 47, at 15–16 (describing several proposed objective assessments and characterizing objective assessment as an ongoing challenge).

108. Käde & von Maltzan, *supra* note 67, at 5; Samek & Müller, *supra* note 47, at 17; EXPLAINABLE AI, *supra* note 50, at 240; Adadi & Berrada, *supra* note 8, at 52,152–53.

109. *See* Adadi & Berrada, *supra* note 8, at 52,156.

110. *Id.* at 52,154.

111. *Id.*

112. Samek & Müller, *supra* note 47, at 8.

C. WHAT MAKES FOR GOOD EXPLANATIONS?

While XAI offers little guidance as to which of its many types of explanations may be best, legal and scientific scholarship outside that field may offer some clues. Although there is no consensus on what an explanation should look like,¹¹³ some scholarship has begun exploring the solution space in a more abstract sense, proposing theoretical categories of explanations and considering whether they may be desirable. In addition, research in philosophy and the human sciences has examined and discovered general properties of human-generated explanations. Together, this scholarship offers strong clues as to which XAI explanations would be most useful. It also indicates a gap in XAI that urgently needs to be addressed: interactivity.

As a preliminary matter, scholars are in agreement that “full transparency” is not the solution to explaining opaque AI. Most people cannot make use of a mountain of complex technical information.¹¹⁴ Source code is particularly impenetrable.¹¹⁵ Even experts find source code of limited help;¹¹⁶ in contrast to classical computer programs, the decision-making logic of an ML model arises not just from the source code itself, but from the data used to train the model.¹¹⁷ Moreover, the relationships learned by the model are often so complex that they escape human understanding.¹¹⁸ It is well known that humans do not find meaning in a trace of the AI’s complex inner logic from inputs to outputs.¹¹⁹ AI takes into account many more factors than humans can keep track of (this is called the “high-dimensionality” problem), and the relationships learned between those factors are indirect and complicated.¹²⁰ The complexity of the resulting logic far outstrips what human reasoning is capable of.¹²¹

Full transparency also runs into problems regarding privacy, trade secrets, fraud, and abuse.¹²² Training data often contains private information about individuals, so releasing that data may violate those individuals’ privacy.¹²³

113. *See, e.g.*, Kroll et al., *supra* note 13, at 638.

114. *See id.* at 639.

115. *See id.*

116. *Id.* at 638.

117. Goodman & Flaxman, *supra* note 41, at 55; *see also supra* Section II.A and accompanying notes (explaining how training ML models works).

118. *See* Kroll et al., *supra* note 13, at 638.

119. Edwards & Veale, *supra* note 10, at 64.

120. *See* Goodman & Flaxman, *supra* note 41, at 55; Edwards & Veale, *supra* note 10, at 22.

121. *See* Goodman & Flaxman, *supra* note 41, at 55; Edwards & Veale, *supra* note 10, at 22.

122. Kroll et al., *supra* note 13, at 638–39.

123. *See id.* at 639.

Releasing too much data may also expose the decision-making system to fraud, abuse, and harms to market position and trade secrets.¹²⁴ Thus, not only does “more transparency” not solve the problem of explaining opaque AI, but there are legitimate countervailing interests in withholding information.

Instead, legal scholarship suggests that an explanation must be some sort of simplification of the AI’s highly complex logic, the form of simplification perhaps depending on context or purpose. At a minimum, the explanation should probably illuminate the relationship between the particular inputs and the particular output.¹²⁵ However, there may not be a single right form of explanation, but rather different forms suitable in different contexts or for different purposes.¹²⁶ Relevant factors might include the technical sophistication of the recipient, the type of harm at risk, and the set of legal rights at stake.¹²⁷

When considering which forms of explanations might be desirable, several scholars distinguish “model-centric” explanations from “subject-centric” explanations.¹²⁸ In fact, this division is the same as that between global and local explanations recognized in XAI.¹²⁹ A “model-centric” explanation is a global one, explaining the AI’s internal design generally, without regard to a particular decision or its particular inputs.¹³⁰ Because a model-centric explanation does not focus on a particular data subject or decision, the information is more abstract and difficult for non-technical people to understand. However, this type of explanation may remain useful for audiences with technical expertise, such as auditors. In contrast, a “subject-centric” explanation is a local one, because it explicitly centers around one particular decision.¹³¹ Decision subjects more easily find meaning in such explanations, since they directly discuss the relationship between the data and the subject.¹³²

Lilian Edwards and Michael Veale identify four types of subject-centric explanations, which have some overlap with the XAI methods discussed in the

124. *See id.* at 638.

125. *See* Edwards & Veale, *supra* note 10, at 58.

126. EXPLAINABLE AI, *supra* note 50, at 240 (“[T]he discussion on which axioms are desirable is still ongoing and may need future extension depending on the application domain and the recipient’s goals.”).

127. *See* Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189, 213 (2019); Edwards & Veale, *supra* note 10, at 22.

128. *See, e.g.*, Edwards & Veale, *supra* note 10, at 22, 55–58; Bamman, *supra* note 28, at 2.

129. *Cf. supra* Section II.B and accompanying notes (defining global and local explanations).

130. Edwards & Veale, *supra* note 10, at 55.

131. *Id.*

132. *See id.* at 58.

previous Section.¹³³ First, a counterfactual explanation discusses what modifications to the input data would change the ultimate decision.¹³⁴ This is the same as the counterfactual method in XAI.¹³⁵ Second, a case-based explanation identifies a data point in the training data most similar to the subject.¹³⁶ This is similar to the prototypes method, another example-generation method in XAI.¹³⁷ Third, a demographic-based explanation identifies a group of other subjects who received a similar decision.¹³⁸ And fourth, a performance-based explanation discloses the confidence that the decision-maker has in the decision.¹³⁹

As for scientific scholarship, Tim Miller conducted a survey of philosophy, psychology, and cognitive science studying human explanations, finding four major properties of human explanations that machine explanations should seek to reflect.¹⁴⁰ First, human explanations seek contrast; humans understand concepts not just by what they are, but by what they are not.¹⁴¹ Thus, humans want to know not why an outcome happened in isolation, but why one outcome happened instead of another.¹⁴² Second, human explanations focus on one or two causes of an outcome, not an exhaustive list of causes.¹⁴³ Third, human explanations are interactive and incorporate the learner's mental model:¹⁴⁴ that is, they are tuned depending on the learner's previous knowledge, level of reasoning skill, and values held. This is made possible because human explanations are social, typically communicated within an interactive conversation.¹⁴⁵ Fourth, humans care much more about causes than about probabilities.¹⁴⁶ These findings are consistent with additional research suggesting that humans reason about individual outcomes rather than about

133. Compare *id.* at 58, with *supra* notes 78–101 and accompanying text.

134. Edwards & Veale, *supra* note 10, at 58. Edwards and Veale actually call this a “[s]ensitivity-based” explanation. *Id.* However, this Note would invite confusion if it discussed this alongside the sensitivity analysis method, which is a related but somewhat different method. Instead, to avoid nomenclature confusion, this Note uses the name of the XAI method identical to the sensitivity-based explanation, the counterfactual. See *id.*

135. See *supra* note 101 and accompanying text.

136. Edwards & Veale, *supra* note 10, at 58.

137. See *supra* note 100 and accompanying text.

138. Edwards & Veale, *supra* note 10, at 58.

139. *Id.* at 58.

140. Tim Miller, *Explanation in Artificial Intelligence: Insights from the Social Sciences*, 267 ARTIFICIAL INTELLIGENCE 1, 3 (2019); Adadi & Berrada, *supra* note 8, at 52,153.

141. See Miller, *supra* note 140, at 3; Adadi & Berrada, *supra* note 8, at 52,153.

142. See Miller, *supra* note 140, at 3; Adadi & Berrada, *supra* note 8, at 52,153.

143. Miller, *supra* note 140, at 3; Adadi & Berrada, *supra* note 8, at 52,153.

144. Miller, *supra* note 140, at 3; Adadi & Berrada, *supra* note 8, at 52,153.

145. See Miller, *supra* note 140, at 3; Adadi & Berrada, *supra* note 8, at 52,153.

146. See Miller, *supra* note 140, at 3.

the general process of producing outcomes.¹⁴⁷ Computer-generated explanations are most likely to prove useful and meaningful when they reflect these findings.

Miller's findings immediately suggest that, from a human-centered point of view, the most useful types of XAI methods available today are influence measurement and example generation with local scope. Because humans most easily reason about individual outcomes rather than the general decision-making logic, this suggests that explanations of local scope are much more accessible and meaningful to the public. Of the four popular methods of explanation—visualization, knowledge extraction, influence measurement, and example generation—influence measurement and example generation are most in accord with Miller's findings. Counterintuitively, despite the intuitive appeal and the fun factor of visualizations, some types, such as visualizing what a neuron learns, are unlikely to be as useful: they do not contrast outcomes, focus on one cause, or interact with the learner. (Certain visualization techniques, however, such as highlighting pivotal portions of an image, may yet be useful as part of a larger explanation, if they can consistently identify a discrete "cause.") In contrast, influence measurement, particularly sensitivity analysis, can contrast different outcomes and isolate a single cause. Example generation, which includes both prototypes and counterfactuals, is also in significant accord with Miller's findings. Example generation is explicitly inspired by cognitive science;¹⁴⁸ humans often reason based on prototypes,¹⁴⁹ and counterfactuals by definition contrast different outcomes because they literally describe how to arrive at a different one. Because humans reason based on prototypes, Edwards's demographic-based explanation, which provides data subjects with similar outcomes, is also useful.

Sadly, of all the types of explanations discussed so far, none is interactive.¹⁵⁰ All of them assemble some information and unilaterally present it to the target learner, with little or no room for adjustments. Thus, it is difficult to tailor currently available explanations to the learner, a difficulty that XAI research should focus urgently on addressing. XAI cannot be expected to be social—it will not be possible for a long time to build a conversant AI capable of understanding its interlocutor's mental model—but it can certainly incorporate interactivity. Perhaps sensitivity analysis can be most easily adapted to be interactive; the FICO tool, available to the public, allows potential loan applicants to interact with the model, playing with their input data and

147. *Id.*; Adadi & Berrada, *supra* note 8, at 52,153.

148. Miller, *supra* note 140, at 3; Adadi & Berrada, *supra* note 8, at 52,153.

149. Miller, *supra* note 140, at 3; Adadi & Berrada, *supra* note 8, at 52,153.

150. *See* Miller, *supra* note 140, at 34.

observing output credit scores, thereby conducting their own sensitivity analysis. Given the importance of interactivity in human explanations, this sort of experimenting with a model—call it the “experimentation method”—is an excellent way, perhaps even the best way, to learn about and understand complicated AI models.¹⁵¹

To conclude, the state of AI, Explainable AI, and legal thought on explanations can be summarized as follows. Neural networks, with their many neurons, parameters, and interrelated behaviors, are so complicated that they are opaque even to experts. Deep neural networks, currently enjoying widespread use today, are both the most accurate type of AI model available today and the opaqueness, illustrating the accuracy–interpretability tradeoff inherent to the spectrum of AI models available today. XAI, a burgeoning field, is beginning to assemble a menu of ways to generate explanations. Though XAI has fundamental challenges to overcome, such as a lack of formal definitions and lack of focus on human factors and interactivity, there is good reason to begin selecting and using explanation methods today. From a human standpoint, the most useful types of explanations appear to be those of local scope, focusing on an individual decision rather than the general logic; influence measurements, especially a sensitivity analysis, which varies the inputs and observes how the output changes; example generation, both prototypes (to describe the training data) and counterfactuals (to describe how to get a different outcome); and demographic-based explanations (to describe others with similar outcomes). Finally, an area needing urgent attention in XAI is methods of explanation that are interactive. The experimentation method, in which a learner gives a model various inputs and observes how it behaves, may be the most promising method of explanation of all.

III. THE GDPR AND THE RIGHT TO EX POST EXPLANATIONS OF AI DECISIONS

Perhaps due in part to AI models being widely deployed yet lacking mechanisms for explaining their decisions, the European Union passed the most comprehensive regulation of data in the world, the General Data

151. I also recognize potential downsides to this experimentation method. Trade secrets or personal data, to the extent they are used in a model, may be at risk of being exposed. *See* Kroll et al., *supra* note 13, at 638–39 (discussing similar problems posed by “full transparency”). And in some use cases, it may be undesirable to give malevolent actors (such as spammers) an opportunity to learn about and circumvent the model. However, on balance, particularly with strategies in place that mitigate potential downsides, this experimentation method remains an extremely promising method for many use cases.

Protection Regulation (GDPR).¹⁵² Evidence suggests that passage and implementation of the GDPR has begun adding additional motivation to Explainable AI research.¹⁵³

However, in order to understand the relationship between the GDPR and Explainable AI, it is first necessary to examine the GDPR and the rights it guarantees. Section III.A provides an overview of the GDPR articles that bear on automated decisions. Section III.B explains the scholarly debate over whether the GDPR guarantees a right to an explanation of AI decisions. Finally, Section III.C argues that a right to an ex post explanation of an AI's decision—that is, an explanation tailored to that individual decision, rather than a generalized explanation of the AI model—arises from the right to contest automated decisions in Article 22.

A. GDPR BACKGROUND

Though the GDPR grants many enumerated rights, scholars have argued about whether it also grants an unenumerated “right to explanation.”¹⁵⁴ Four articles bear on this question: Articles 13, 14, 15, and 22.¹⁵⁵ These articles differ in which rights they enumerate, when those rights trigger, and whether they explicitly mention some sort of explanation.

Article 13 requires a data controller to provide certain information to a data subject whenever it obtains personal data from that data subject.¹⁵⁶ Whenever that data is collected and might be subject to “automated decision-making,” the data controller must provide the data subject with “meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.”¹⁵⁷ This information is said to be “necessary to ensure fair and transparent processing.”¹⁵⁸ Scholars focus on the phrase “meaningful information about the logic involved”¹⁵⁹ as a kind of explanation to which the data subject has a right. This right seems to trigger upon data collection, presumably before any further processing or decision-making has been done.

Article 14 is ultimately very similar to Article 13, requiring a data controller to provide certain information to a data subject whenever it obtains personal

152. GDPR, *supra* note 6.

153. *See, e.g.*, Samek & Müller, *supra* note 47, at 9; Goodman & Flaxman, *supra* note 41.

154. *See infra* Section III.B.

155. GDPR, *supra* note 6, arts. 13–15, 22.

156. *Id.* art. 13.

157. *Id.* art. 13(2)(f).

158. *Id.* art. 13(2).

159. *Id.* art. 13(2)(f).

data about the data subject from elsewhere.¹⁶⁰ Article 14 is phrased quite awkwardly in that it triggers the right to information when “personal data have not been obtained from the data subject.”¹⁶¹ Instead of triggering whenever data is not collected, which would be nonsensical, the article is meant to trigger when data is collected from elsewhere, such as from a data broker.¹⁶² Just like Article 13, Article 14 requires that whenever such data is collected and might be subject to “automated decision-making,” the data controller must provide the data subject with “meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.”¹⁶³

Article 15 grants a freestanding right rather than one dependent on a timing trigger. It grants data subjects the right to know whether their personal data “are being processed.”¹⁶⁴ If so, and if the data might be subject to “automated decision-making,” the data subject has the right—once again—to obtain “meaningful information about the logic involved” and the “significance and the envisaged consequences.”¹⁶⁵

Article 22, which centers around “[a]utomated individual decision-making,” is quite different.¹⁶⁶ Article 22(1) issues a flat ban on any decision-making “based solely on automated processing” that has “legal effects” or “similarly significant[]” effects on the data subject.¹⁶⁷ Article 22(2) offers exceptions to the ban, such as when the processing is required by contract or when the data subject consents.¹⁶⁸ Even if one of those exceptions applies, Article 22(3) nonetheless requires the data controller to “implement suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.”¹⁶⁹

160. *Id.* art. 14.

161. *Id.*

162. *See* ARTICLE 29 DATA PROT. WORKING PARTY, GUIDELINES ON TRANSPARENCY UNDER REGULATION 2016/679, at 12 (2017).

163. GDPR, *supra* note 6, art. 14(2)(g).

164. *Id.* art. 15(1).

165. *Id.* art. 15(1)(h).

166. *Id.* art. 22.

167. *Id.* art. 22(1); ARTICLE 29 DATA PROT. WORKING PARTY, GUIDELINES ON AUTOMATED INDIVIDUAL DECISION-MAKING AND PROFILING FOR THE PURPOSES OF REGULATION 2016/679, at 19 (2017) [hereinafter GUIDELINES ON AUTOMATED DECISION-MAKING].

168. GDPR, *supra* note 6, art. 22(2); GUIDELINES ON AUTOMATED DECISION-MAKING, *supra* note 167, at 19.

169. GDPR, *supra* note 6, art. 22(3); GUIDELINES ON AUTOMATED DECISION-MAKING, *supra* note 167, at 19.

Notably, the event triggering Article 22(3) rights is different than that triggering Articles 13–15: here, rights trigger upon a decision being made about the data subject.

Recital 71 contains language closely mirroring that of Article 22. Formally, recitals lack the full force of law that articles have, since they are not considered part of the regulation proper.¹⁷⁰ As a result, legislators feel freer to add or move language into recitals during the negotiation process.¹⁷¹ However, recitals are an important factor in interpreting the law, as they are cited when the law itself is vague.¹⁷² Like Article 22, Recital 71 bans any decision-making “based solely on automated processing” that has “legal effects” or “similarly significant[]” effects on the data subject.¹⁷³ The structure of Recital 71 also mirrors that of Article 22: after the ban, it lists exceptions, and then it lists “safeguards” that apply, regardless of whether any of the exceptions apply.¹⁷⁴ On its face, Recital 71 goes further than Article 22, listing among its safeguards not just the right to “challenge the decision,” but also the “right . . . to obtain an explanation of the decision.”¹⁷⁵ However, given Recital 71’s close relation to Article 22, it is no surprise that data protection authorities and courts alike have begun to use Recital 71 to inform their interpretation of Article 22.¹⁷⁶

The natural question, then, is whether a right to an explanation arises from Article 22.

B. THE EVOLVING DEBATE OVER THE “RIGHT TO EXPLANATION”

Scholarship on the right to explanation in the GDPR has quickly evolved since the debate’s genesis in 2017. The initial skepticism over the right’s very existence has already given way to a new debate over the right’s scope.

170. See Kaminski, *supra* note 127, at 193–94; Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT’L DATA PRIVACY L. 76, 77–78 (2017); Edwards & Veale, *supra* note 10, at 21, 50; Brkan, *supra* note 9, at 115; Chris Jay Hoofnagle, Bart van der Sloot & Frederik Zuiderveen Borgesius, *The European Union General Data Protection Regulation: What It Is and What It Means*, 28 INFO. & COMM. TECH. L. 65, 67 (2019).

171. See Edwards & Veale, *supra* note 10, at 50.

172. See Brkan, *supra* note 9, at 115; Kaminski, *supra* note 127, at 194; see also Andrew D. Selbst & Julia Powles, *Meaningful Information and the Right to Explanation*, 7 INT’L DATA PRIVACY L. 233, 235 (2017).

173. GDPR, *supra* note 6, Recital 71; see *id.* art. 22(1).

174. *Id.* Recital 71; see *id.* art. 22.

175. *Id.* Recital 71; see *id.* art. 22(3).

176. See GUIDELINES ON AUTOMATED DECISION-MAKING, *supra* note 167, at 19–20; Rb. Amsterdam 11 maart 2021, JAR 2021, 96 m.nt. Rietveld, R.D. (verzoeker 1/Uber B.V.) (Neth.), ECLI:NL:RBAMS:2021:1018, ¶ 4.11–12.

Early scholarship generally doubted the robustness of the right to explanation, and especially Article 22's role in such a right. In 2017, Bryce Goodman and Seth Flaxman were the first to propose that the GDPR might create a right to explanation of a particular automated decision made about a data subject.¹⁷⁷ This proposal sparked consternation and debate. Sandra Wachter, Brent Mittelstadt, and Luciano Floridi countered that only a "limited right to be informed" of a system's general design arose from Article 15, and that no right of explanation arose from Article 22 at all.¹⁷⁸ Andrew Selbst and Julia Powles disagreed with Wachter et al.'s narrow conception of the right, but nonetheless agreed that no right to explanation arose from Article 22,¹⁷⁹ asserting in support that an Article 15 explanation of the general system sufficed to explain specific decisions.¹⁸⁰ Edwards and Veale also discounted Article 22, arguing that an individual right to explanation would fail to protect against the main algorithmic harm of stigmatizing groups¹⁸¹ and fail to be useful to the individuals harmed.¹⁸² But not all agreed with the powerlessness of Article 22: Isak Mendoza and Lee Bygrave dissented, proposing that a right to explanation could arise from Article 22(3)'s right to contest an automated decision.¹⁸³

However, after the 2017 release of guidelines¹⁸⁴ from the European Data Protection Board (then called the Article 29 Working Party)¹⁸⁵ on automated decision-making under the GDPR, scholarship began to recognize robustness in the right to explanation and instead turned to considering its scope. Gianclaudio Malgieri and Giovanni Comandé argued that Articles 13–15 and 22 should be interpreted "systemically" to require AI "legibility."¹⁸⁶ Antoni Roig argued that a "[r]ight to be informed" was necessary but insufficient,

177. Goodman & Flaxman, *supra* note 41, at 50.

178. Wachter et al., *supra* note 170, at 77–78 (internal quotation marks omitted).

179. Selbst & Powles, *supra* note 172, at 237.

180. *Id.* at 240.

181. Edwards & Veale, *supra* note 10, at 22.

182. *Id.* at 67. Curiously, Edwards and Veale remark in the general case that "an explanation, or some kind of lesser transparency, is of course often essential to mount a challenge," *id.* at 40, yet neglect to apply this reasoning to Article 22(3)'s right to contest a decision.

183. Isak Mendoza & Lee A. Bygrave, *The Right Not to be Subject to Automated Decisions Based on Profiling, in* EU INTERNET LAW: REGULATION AND ENFORCEMENT 77, 93 (Tatiana-Eleni Synodinou, Philippe Jougoux, Christiana Markou & Thalia Prastitou eds., 2017).

184. GUIDELINES ON AUTOMATED DECISION-MAKING, *supra* note 167.

185. *See* GDPR, *supra* note 6, art. 94.

186. Gianclaudio Malgieri & Giovanni Comandé, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, 7 INT'L DATA PRIVACY L. 243, 250 (2017).

instead emphasizing human intervention and external auditing.¹⁸⁷ Maja Brkan argued for interpreting Articles 13–15 and 22 holistically in order to support a right to be informed of the crucial reasons behind an automated decision.¹⁸⁸ Bryan Casey, Ashkon Farhangi, and Roland Vogl, interpreting the guidelines as a call for sweeping regulatory oversight, dismissed the utility of an individual explanation¹⁸⁹ and argued instead for a global explanation of the system, one intended for auditors.¹⁹⁰ Margot Kaminski instead interpreted the guidelines to call for a dual regime with two types of explanation:¹⁹¹ one arising from robust, systemic oversight in Articles 13–15,¹⁹² and a second type arising from a robust system of algorithmic due process, in which Article 22 is central.¹⁹³

Thus, there are two distinct types of explanation at issue: *ex ante* and *ex post*. The right to an *ex ante* explanation arises relatively explicitly, primarily from the right to “meaningful information about the logic involved” in Articles 13–15.¹⁹⁴ Because data subjects can exercise this right before any automated decision is made, such an explanation must be about the system generally (a global or system-centric explanation).¹⁹⁵ The right to an *ex post* explanation arises implicitly, primarily from the right to contest a decision in Article 22. Because an explanation of a decision is needed in order to contest it, such an explanation must be about the specific decision made (a local or subject-centric explanation).¹⁹⁶

C. THE RIGHT TO AN EX POST EXPLANATION ARISES IMPLICITLY

Although scholars now broadly agree that the GDPR guarantees a robust right to an *ex ante* explanation of AI systems, they still debate whether there exists a right to an *ex post* explanation. However, accumulated evidence from supervisory authorities¹⁹⁷ and a recent court case,¹⁹⁸ coupled with an examination of the GDPR itself and of the principles of due process, indicates

187. Roig, *supra* note 21, at 6, 9–10.

188. Brkan, *supra* note 9, at 112.

189. Casey et al., *supra* note 9, at 179–81.

190. *Id.* at 183.

191. Kaminski, *supra* note 127, at 210–11, 217.

192. *Id.* at 215.

193. *Id.* at 204–07.

194. *See* GDPR, *supra* note 6, arts. 13–15.

195. *See supra* notes 128–132 and accompanying text.

196. *See supra* notes 128–132 and accompanying text.

197. GUIDELINES ON AUTOMATED DECISION-MAKING, *supra* note 167.

198. Rb. Amsterdam 11 maart 2021, ECLI:NL:RBAMS:2021:1019 (verzoeker 1/Ola Netherlands B.V.) (Neth.) ¶¶ 4.41, 4.52 (requiring an explanation of an automated decision, pursuant to Article 15 in view of Article 22, to include the most important inputs and their role).

that Kaminski’s argument is most persuasive: there is a dual regime including both ex ante explanations (global explanations, intended for regulators) and ex post explanations (local explanations, intended for due process). The text and structure of the GDPR suggest a dual regime; supervisory authorities view Article 22 as requiring an appeals process with explanations; Article 22’s right to contest requires a right to an ex post explanation, according to the principles of due process; and guidance from the High-Level Expert Group on Artificial Intelligence¹⁹⁹ underscores the connection to due process and the need for ex post explanations.

1. *The Text and Structure of the GDPR Indicate a Dual Regime*

The text of Recital 71 explicitly calls for a right to an ex post explanation, and the separation of Articles 13–15 from Article 22 suggests a dual system of explanation, including both ex post and ex ante explanations.

The text of Recital 71 explicitly says that such a right exists. In straightforward language echoing Article 22, it says that for any decision-making “based solely on automated processing” that has “legal effects” or “similarly significant[]” effects on a data subject, the data subject has the “right . . . to obtain an explanation of the decision.”²⁰⁰ Although Recital 71 may lack the direct legal force of Article 22,²⁰¹ it is first in line to help interpret it.²⁰² Indeed, the European Data Protection Board has already cited Recital 71 to find that Article 22 safeguards should include an explanation of the decision,²⁰³ and courts have already begun using Recital 71 to help interpret Article 22.²⁰⁴ Although the “right . . . to obtain an explanation”²⁰⁵ language has not yet been tested, should Recital 71 continue to be used by supervisory authorities and

199. HIGH-LEVEL EXPERT GRP. ON ARTIFICIAL INTELLIGENCE, ETHICS GUIDELINES FOR TRUSTWORTHY AI 13 (2019) [hereinafter AI-HLEG ETHICS GUIDELINES]; HIGH-LEVEL EXPERT GRP. ON ARTIFICIAL INTELLIGENCE, POLICY AND INVESTMENT RECOMMENDATIONS FOR TRUSTWORTHY AI (2019) [hereinafter AI-HLEG POLICY RECOMMENDATIONS].

200. GDPR, *supra* note 6, Recital 71.

201. *See* Kaminski, *supra* note 127, at 193–94; Wachter et al., *supra* note 170, at 77–78; Edwards & Veale, *supra* note 10, at 21, 50; Brkan, *supra* note 9, at 115; Hoofnagle et al., *supra* note 170, at 67.

202. *See* Brkan, *supra* note 9, at 115; Kaminski, *supra* note 127, at 194; *see, e.g.*, Rb. Amsterdam 11 maart 2021, ECLI:NL:RBAMS:2021:1019 (verzoeker 1/Ola Netherlands B.V.) (Neth.) ¶ 4.38–39 (using Recital 71 to help interpret Article 22).

203. GUIDELINES ON AUTOMATED DECISION-MAKING, *supra* note 167, at 27.

204. *See, e.g.*, Rb. Amsterdam 11 maart 2021, JAR 2021, 96 m.nt. Rietveld, R.D. (verzoeker 1/Uber B.V.) (Neth.), ECLI:NL:RBAMS:2021:1018, ¶ 4.11–12; Rb. Amsterdam 11 maart 2021, ECLI:NL:RBAMS:2021:1019 (verzoeker 1/Ola Netherlands B.V.) (Neth.) ¶ 4.38–39.

205. GDPR, *supra* note 6, Recital 71.

courts to interpret Article 22, it is difficult to avoid the conclusion that the language gives rise to a right to an ex post explanation.

Consistent with this interpretation, the structure of the GDPR suggests a dual regime of both ex ante and ex post explanations. Articles 13–15 comprise Section 2 of the GDPR, “Information and access to personal data.”²⁰⁶ Article 22, on the other hand, sits in Section 4, “Right to object and automated individual decision-making.”²⁰⁷ In other words, Section 2 is about rights regarding the collection of data (ex ante), whereas Section 4 is about rights regarding process following a decision (ex post). These two sections have different rights, different purposes, different timing—and, thus, different explanations. But because each type of explanation can achieve different goals, they combine into a more robust regulatory scheme.

2. *GDPR Supervisory Authorities View Article 22 as Requiring an Appeals Process with an Ex Post Explanation*

The European Data Protection Board (EDPB) and a few individual Data Protection Authorities (DPAs) have released guidance on interpreting Article 22. The EDPB, which released its guidelines in 2017,²⁰⁸ comprises the DPAs of the European Union, each DPA responsible for enforcing the GDPR within its E.U. member state.²⁰⁹ The EDPB’s guidelines therefore represent a consensus on how the DPAs intend to enforce the GDPR.²¹⁰ The guidance from each supports a right to an ex post explanation of AI decisions.

The EDPB’s guidelines plainly state that Article 22 requires a review process. “Article 22(3) requires controllers to implement suitable measures to safeguard data subjects’ rights[,] freedoms[,] and legitimate interests,” and such safeguards must include a “review . . . carried out by someone who has the appropriate authority and capability to change the decision.”²¹¹ Indeed, many member states’ implementing regulations explicitly affirm the right to contest as a safeguard measure.²¹² The guidelines also indirectly suggest that this review

206. GDPR, *supra* note 6, § 2.

207. *Id.* § 4.

208. GUIDELINES ON AUTOMATED DECISION-MAKING, *supra* note 167.

209. *See* GDPR, *supra* note 6, art. 68.

210. *See* Kaminski, *supra* note 127, at 194. The guidelines may also guide courts. *See, e.g.*, Rb. Amsterdam 11 maart 2021, ECLI:NL:RBAMS:2021:1019 (verzoeker 1/Ola Netherlands B.V.) (Neth.) ¶ 4.38 (using the guidelines to help interpret Article 22).

211. GUIDELINES ON AUTOMATED DECISION-MAKING, *supra* note 167, at 27.

212. Gianclaudio Malgieri, *Automated Decision-Making in the EU Member States: The Right to Explanation and Other “Suitable Safeguards” in the National Legislations*, 35 COMPUTER L. & SECURITY REV. 105327, at 21 (2019). Moreover, a few states’ regulations—Hungary, France, the United Kingdom, and Ireland—provide for a right to an explanation, although most do not. *Id.* at 22–23.

process is one part of a dual regime of audits and reviews, recommending that controllers both “provide the auditor with all necessary information about how the algorithm or machine learning system works” and, separately, “allow the data subject to express his or her point of view and contest the decision.”²¹³

Further, the guidelines are clear that the review process must include an explanation of the decision, in order to breathe life into the right to contest. “Recital 71 highlights that in any case suitable safeguards should also include: . . . the right . . . to obtain an explanation of the decision reached . . . and to challenge the decision. The controller must provide a simple way for the data subject to exercise these rights.”²¹⁴ Such an explanation is necessary because “[t]he data subject will only be able to challenge a decision or express their view if they fully understand how it has been made and on what basis.”²¹⁵ The harms that this review process can help address include “errors,” “bias,” and “discrimination.”²¹⁶

The stated purpose of addressing errors, bias, and discrimination implicitly underscores the need for an individualized (ex post) explanation and appeals process. Errors require individualized remedies, because errors will always be a part of any system, no matter how much ex ante testing and oversight exists, and errors that are difficult to find typically harm individuals. Bias and discrimination, though sometimes thought to be mainly rooted out by ex ante review,²¹⁷ in fact require both ex ante review and ex post review to address.²¹⁸ Ex ante review can identify systemic discrimination in many cases, such as if a rule is based on data that is biased or that poorly represents the distribution of data subjects.²¹⁹ But ex ante review cannot detect all types of discriminatory rules.²²⁰ Complex rules can have complex effects, and it is difficult to determine ex ante whether a complex rule will have discriminatory effects.²²¹ This is especially true for AI, whose complex rules often arise from complex relationships between data points in ways humans often cannot determine.²²² As a result, an ex post review—an appeals process—is necessary to remedy

213. *Id.* at 32.

214. *Id.* at 27.

215. *Id.*

216. *Id.* at 27–28.

217. *See, e.g.,* Casey et al., *supra* note 9, at 179–81 (arguing that individual remedial mechanisms are insufficient to protect individuals from discrimination and preferring thorough regulatory oversight).

218. *See* Kroll et al., *supra* note 13, at 680–82.

219. *Id.* (describing several ways discrimination can be built into AI models).

220. *Id.* at 680.

221. *Id.*

222. *See supra* Section II.A and accompanying notes; *see also* Kroll et al., *supra* note 13, at 680.

the unforeseen discriminatory harms, and a necessary part of that process is an ex post explanation of decisions in order to identify the cause of those harms.²²³

In accord with the EDPB's guidance, guidance from individual DPAs echoes that making full use of the right to contest in a review process requires an explanation of the decision. Guidance from the Norwegian DPA, Datatilsynet, states that "the data controller must provide as much information as necessary in order for the data subject to exercise his or her rights. This means that the decision must be explained . . ." ²²⁴ Moreover, "the explanation has to enable the data subject to understand why a particular decision was reached, or what needs to change in order for a different decision to be reached." ²²⁵ In 2019, before the United Kingdom exited the European Union, the United Kingdom's DPA, the Information Commissioner's Office (ICO), released draft guidelines with a strong affirmation of a right to an ex post explanation. ICO's draft guidelines state that Recital 71 "clarif[ies] the meaning and intention" of GDPR articles, ²²⁶ and thus, Recital 71's call for a right "to an explanation of an automated decision after it has been made . . . makes clear that such a right is implicit in Articles 15 and 22." ²²⁷ Echoing the EDPB, ICO's draft guidelines reason that a data subject "need[s] . . . [to be able to receive] an explanation of a fully automated decision to enable their rights to obtain meaningful information, express their point of view[,] and contest the decision." ²²⁸

Moreover, courts may be beginning to agree. Although caselaw is scarce, the District Court of Amsterdam recently held that Article 15, in view of Article 22, requires an explanation of the logic behind an automated decision to include the most important inputs and their role in the automated decision. ²²⁹ The purpose, the court explained, was to allow the data subject to understand the decision's basis in order to verify its correctness and lawfulness. ²³⁰

223. See Brkan, *supra* note 9, at 118.

224. DATATILSYNET, ARTIFICIAL INTELLIGENCE AND PRIVACY 21 (2018), <https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf>.

225. *Id.* at 21–22.

226. INFO. COMM'RS OFFICE & ALAN TURING INST., EXPLAINING DECISIONS MADE WITH AI: PART 1: THE BASICS OF EXPLAINING AI 10 (2019), <https://ico.org.uk/media/about-the-ico/consultations/2616434/explaining-ai-decisions-part-1.pdf>.

227. *Id.*

228. *Id.*

229. Rb. Amsterdam 11 maart 2021, ECLI:NL:RBAMS:2021:1019 (verzoeker 1/Ola Netherlands B.V.) (Neth.) ¶¶ 4.39, 4.41, 4.52.

230. *Id.* ¶ 4.41.

In sum, it is the view of the EDPB and individual DPAs that Article 22 gives rise to the right to an ex post explanation. In their view, an ex post explanation is necessary to breathe life into Article 22's right to contest automated decisions.

3. *The Right to an Ex Post Explanation Arises from Article 22's Right to Contest Automated Decisions*

An examination of the historical purpose of Article 22 and the principles of due process further bolsters the position of the DPAs that an ex post explanation is a necessary part of due process and the right to challenge decisions.

Examining the history of Article 22 requires examining its similar predecessor in the Data Protection Directive,²³¹ which the GDPR replaced.²³² Article 22 guarantees “the right not to be subject to a decision based solely on automated processing . . . which produces legal effects concerning him or her or similarly significantly affects him or her” unless an exception applies, in which case there shall nonetheless be “suitable measures to safeguard the data subject's . . . legitimate interests,” including “the right . . . to express his or her point of view and to contest the decision.”²³³ This language is quite similar to that of Article 15 of the Directive, which says “Member States shall grant the right to every person not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data” unless an exception applies, such as when “there are suitable measures to safeguard his legitimate interests, such as arrangements allowing him to put his point of view.”²³⁴ Thus, the purposes of Article 15 of the Directive surely also animate Article 22 of the GDPR.²³⁵

Examining the legislative history reveals a concern that automatic profiling erodes due process when individuals lack the ability to appeal and make their case. When the European Commission explained an earlier proposal for Article 15, the Commission expressed two concerns.²³⁶ First, it was concerned that automatic profiling “deprives the individual of the capacity to influence

231. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, 1995 O.J. (L 281) 31 [hereinafter Data Protection Directive].

232. See GDPR, *supra* note 6, art. 94; see also Hoofnagle et al., *supra* note 170, at 69–72 (explaining some of the motivations behind replacing the Data Protection Directive with the GDPR).

233. GDPR, *supra* note 6, art. 22.

234. Data Protection Directive, *supra* note 231, art. 15.

235. See Mendoza & Bygrave, *supra* note 183, at 83–84.

236. See *id.*

decision-making processes,” and thus the proposal was “designed to protect the interest of the data subject in participating in the making of decisions which are of importance to him.”²³⁷ Second, the Commission was concerned that “sophisticated software . . . has an apparently objective and incontrovertible character to which a human decision-maker may attach too much weight, thus abdicating his own responsibilities [to investigate].”²³⁸ In other words, due process erodes when individuals lack the ability to appeal and make their case, and this can happen especially when humans give too much deference to decisions made by sophisticated programs. The additional right to “contest” in Article 22 further strengthens the connection to due process, since to contest is not merely to disagree, but to appeal.²³⁹

An examination of the principles of due process reveals the same reasoning: due process requires a right to challenge decisions. E.U. civil procedure is largely unharmonized,²⁴⁰ and to the limited extent it is harmonized, it comprises a complex combination of treaties, principles deriving from the E.U. Charter and the European Convention on Human Rights (ECHR), horizontal secondary legislation, sectoral secondary legislation, and the jurisprudence of the Court of Justice of the European Union and of the European Court of Human Rights (ECtHR).²⁴¹ However, some principles of civil procedure are more harmonized than others. The principles of equality (an adjudicator treating parties equally) and fair play (the parties treating each other fairly) are among the most harmonized.²⁴² From these principles derive several rights: the party’s right to be able to prepare and present a defense, the right to obtain reasons for a decision against the party, and the right to challenge the decision.²⁴³

237. *Proposal for a Council Directive Concerning the Protection of Individuals in Relation to the Processing of Personal Data*, at 29, COM (1990) 314 final (Sept. 13, 1990).

238. *Amended Proposal for a Council Directive on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data*, at 26, COM (1992) 422 final (Oct. 15, 1992).

239. *See* Mendoza & Bygrave, *supra* note 183, at 93.

240. MAGDALENA TULIBACKA, MARGARITA SANZ & ROLAND BLOMEYER, EUROPEAN PARLIAMENTARY RESEARCH SERV., COMMON MINIMUM STANDARDS OF CIVIL PROCEDURE 13 (2016) [hereinafter CIVIL PROCEDURE STUDY].

241. *Id.* at 6.

242. *Id.* at 19–20, 44.

243. *Id.* at 44; *see also* Charter of Fundamental Rights of the European Union art. 41, 2016 O.J. (C 202) 389, 401 [hereinafter E.U. Charter] (guaranteeing that before E.U. institutions, every citizen has the “right . . . to be heard,” which places upon the institution the obligation to “give reasons for its decisions”); *id.* art. 47 (guaranteeing the right to an “effective remedy” by way of a “fair and public hearing”). In addition, ECtHR jurisprudence protects the principle of “equality of arms,” that is, a fair opportunity for each party to present their case. *Dombo*

Further examination reveals that because the right to obtain the reasons for the decision is necessary in order to make use of the right to challenge, the reasoning provided must be *ex post*, not *ex ante*. It is useful to examine another principle of civil procedure, the duty of a court to provide reasons for its decisions.²⁴⁴ In particular, the ECtHR interprets Article 6(1) of the ECHR, which guarantees the right to a “fair and public hearing,”²⁴⁵ as imposing a duty on courts to provide reasons for their decisions.²⁴⁶ This is because the party needs that reasoning, in sufficient detail, in order to make effective use of the right to challenge that decision.²⁴⁷ Although further detail is not harmonized,²⁴⁸ it is easy to see that an *ex ante* explanation would not suffice. An *ex ante* explanation would amount to an explanation of the structure, function, and practices of the E.U. justice system. While such an explanation may help identify systemic issues, providing it as the sole justification for an individual decision would be an obvious affront to justice and due process. It would not allow either the party or an appeals court to examine the decision’s factual findings or legal reasoning. As a result, it would violate the party’s right to prepare and present its defense and to challenge the decision. Only an explanation tailored to the facts and law of the case—an *ex post* explanation—would suffice to make use of the right to challenge.

Thus, due process requires *ex post* explanations. To summarize: the principles of due process include equality and fair play; from these derive the right to challenge a decision; in order to make effective use of that right, there must be a right to obtain the reasons for the decision, i.e., an *ex post* explanation. The right to contest in Article 22, of course, embodies that right to challenge in a specific context, automated processing. Therefore, as this examination of due process shows, and as Kaminski’s analysis of due process recognizes,²⁴⁹ only an *ex post* explanation suffices to satisfy Article 22. Only with an explanation about the decision specifically, rather than about the model generally, can the data subject make effective use of their right to

Beheer B.V. v. The Netherlands, 18 Eur. Ct. H.R. 213, ¶ 33 (1993); CIVIL PROCEDURE STUDY, *supra* note 240, at 44.

244. CIVIL PROCEDURE STUDY, *supra* note 240, at 59.

245. Convention for the Protection of Human Rights and Fundamental Freedoms art. 6(1), Nov. 4, 1950, 213 U.N.T.S. 221 [hereinafter ECHR].

246. H. v. Belgium, 127 Eur. Ct. H.R. (ser. A) ¶ 53 (1987); CIVIL PROCEDURE STUDY, *supra* note 240, at 59.

247. Hirvisaari v. Finland, App. No. 49684/99, 38 Eur. H.R. Rep. 7, ¶ 30 (2001); CIVIL PROCEDURE STUDY, *supra* note 240, at 59.

248. CIVIL PROCEDURE STUDY, *supra* note 240, at 59.

249. Kaminski, *supra* note 127, at 204.

contest the decision. This reasoning echoes the guidance from GDPR regulators discussed earlier.²⁵⁰

4. *AI-HLEG Guidance Elaborates on the Special Need for Ex Post Explanations in a World with Opaque AI*

The European Commission's High-Level Expert Group on Artificial Intelligence (AI-HLEG) is an advisory board set up by the European Commission to provide guidance on AI. In 2019, it released guidelines on the ethics of AI.²⁵¹ Although these guidelines do not directly reflect an official position of the EU, they provide a roadmap for DPAs to use when interpreting and applying the GDPR with regard to AI. The guidelines view AI as a threat that must be addressed, in part by requiring ex post explanations.

The AI-HLEG's guidelines on ethical AI view AI as a threat to the sovereignty of the European Union.²⁵² The ethical guidelines view AI as such a grave threat that AI implicates the E.U. Treaties, the E.U. Charter, and international human rights law.²⁵³ Specifically, the guidelines worry that AI is a threat to human dignity (via objectification), autonomy (via manipulation and surveillance), equality (via discrimination), and even democracy, justice, and the rule of law (via, inter alia, a lack of due process).²⁵⁴

To protect against these harms, the guidelines explicitly call for explainable AI decisions. The guidelines identify "fairness" and "explicability" as principles key to protecting against the harms to equality, rule of law, etc.²⁵⁵ Fairness, the guidelines say, includes the ability to "contest and seek effective redress against decisions made by AI,"²⁵⁶ which in turn requires "decision-

250. See GUIDELINES ON AUTOMATED DECISION-MAKING, *supra* note 167, at 27 ("The data subject will only be able to challenge a decision or express their view if they fully understand how it has been made and on what basis."); INFO. COMM'RS OFFICE & ALAN TURING INST., *supra* note 226, at 10 ("You need to be able to give an individual an explanation of a fully automated decision to enable their rights to obtain meaningful information, express their point of view[,] and contest the decision.").

251. See AI-HLEG ETHICS GUIDELINES, *supra* note 199. The group also released a separate set of AI policy recommendations. AI-HLEG POLICY RECOMMENDATIONS, *supra* note 199.

252. See AI-HLEG ETHICS GUIDELINES, *supra* note 199, at 2, 9–13.

253. See *id.* at 10.

254. See *id.* Moreover, the AI-HLEG Policy Recommendations echo the concern that AI threatens democracy and the rule of law, calling for monitoring the use of AI for lack of due process. AI-HLEG POLICY RECOMMENDATIONS, *supra* note 199, at 40. The AI-HLEG's document defining AI identified two concerns with AI: bias and opacity. HIGH-LEVEL EXPERT GRP. ON ARTIFICIAL INTELLIGENCE, A DEFINITION OF AI: MAIN CAPABILITIES AND DISCIPLINES 5 (2019).

255. See AI-HLEG ETHICS GUIDELINES, *supra* note 199, at 12–13.

256. *Id.* at 13.

making processes [to] be explicable.”²⁵⁷ Without such an ex post explanation, “a decision cannot be duly contested.”²⁵⁸

In sum, all of this evidence, from many sources, cumulatively indicates that the GDPR should and does guarantee a right to an ex post explanation of AI decisions. Recital 71 explicitly calls for it; the structure of the GDPR suggests a dual regime of explanations; the EDPB and individual DPAs view explanations as necessary to breathe life into the right to contest; the legislative history indicates a concern for due process; principles of due process mandate an ex post explanation to make use of the right to challenge decisions; and AI-HLEG guidance calls for due process and explicability. Due process is just one part, but a critical part, of the robust regime of regulatory accountability that the GDPR establishes. In short, Article 22 of the GDPR guarantees the right to an ex post explanation of AI decisions.

IV. WHICH TYPES OF AI EXPLANATIONS BEST SATISFY THE GDPR?

The opacity of AI, particularly the deep neural network ubiquitous today, is a novel challenge to accountability and due process. Its enormous complexity impedes identifying and fixing causes of errors, bias, and discrimination.²⁵⁹ Trying to address this problem, the burgeoning field of Explainable AI offers a menu of potential methods of generating explanations of AI, including visualizing, extracting knowledge, measuring the influence of input features, and generating examples.²⁶⁰ Legal scholarship proposes additional potential methods, such as case-based, demographic-based, and performance-based explanations.²⁶¹ This Note proposes yet another method of explanation, the experimentation method of explanation.²⁶² In the midst of rapid development of these many choices, Article 22 of the GDPR guarantees data subjects a right to an ex post explanation of AI decisions.²⁶³

Thus, one question remains: what form should the ex post explanation take? In other words, how should a practitioner select a method of explanation from the available menu to best satisfy the GDPR?

To recommend solutions and future directions, this Note draws on two lessons from due process. First, the fundamental purpose of contesting a

257. *Id.*

258. *Id.*

259. *See supra* Section II.A and accompanying notes.

260. *See supra* Section II.B and accompanying notes.

261. *See supra* Section II.C and accompanying notes.

262. *See supra* note 151 and accompanying text.

263. *See supra* Section III.C and accompanying notes.

decision is seeking to change it. This is reflected in the principles of due process, particularly the principles of equality and fair play; from them derive the right to challenge a decision, that is, to appeal or ask for review.²⁶⁴ The EDPB views the right to contest in Article 22 as establishing a review process which “must be carried out by someone who has the appropriate authority and capability to change the decision.”²⁶⁵ It follows that, for the right to contest to mean anything, it must be reasonably possible to change the decision. Thus, the data subject must understand what it would take to change the decision. An explanation of an AI decision can help.

Second, contestation requires the ability to check the reliability of the evidence and questioning in person is a key tool for doing so. Although neither evidence law nor appeal mechanisms are harmonized in the European Union,²⁶⁶ this principle can be found at the E.U. level in the context of criminal law. Article 6 of the ECHR, in the context of the right to “a fair and public hearing,” provides criminal defendants the right “to examine or have examined witnesses against him.”²⁶⁷ According to the ECtHR, “The underlying principle is that the defendant in a criminal trial should have an effective opportunity to challenge the evidence against him,” in particular the ability “to test the truthfulness and reliability of [witnesses’] evidence, by having them orally examined in his presence.”²⁶⁸ In other words, a fair trial must allow the defendant to challenge evidence, which necessitates the ability to check reliability of evidence, which is ensured in part via questioning (cross-examination) in person.²⁶⁹ This logic extends naturally beyond the criminal

264. See CIVIL PROCEDURE STUDY, *supra* note 240, at 44.

265. GUIDELINES ON AUTOMATED DECISION-MAKING, *supra* note 167, at 27.

266. CIVIL PROCEDURE STUDY, *supra* note 240, at 46.

267. ECHR, *supra* note 245, art. 6.

268. *Al-Khawaja & Tahery v. United Kingdom*, 2011-VI Eur. Ct. H.R. 191, ¶ 127; see also ELODIE SELIER & ANNE WEYEMBERGH, POLICY DEP’T FOR CITIZENS’ RIGHTS & CONSTITUTIONAL AFFAIRS, CRIMINAL PROCEDURAL LAWS ACROSS THE EUROPEAN UNION—A COMPARATIVE ANALYSIS OF SELECTED MAIN DIFFERENCES AND THE IMPACT THEY HAVE OVER THE DEVELOPMENT OF EU LEGISLATION 69 (2018).

269. U.S. law recognizes similar reasoning. The Sixth Amendment to the U.S. Constitution provides that, “[i]n all criminal prosecutions, the accused shall enjoy the right . . . to be confronted with the witnesses against him.” U.S. CONST. amend. VI. According to the Supreme Court, this “Clause’s ultimate goal is to ensure reliability of evidence.” *Crawford v. Washington*, 541 U.S. 36, 61 (2004).

context to AI,²⁷⁰ aligning well with the finding that human-generated explanations are interactive.²⁷¹

Applying these lessons, there are two types of ex post explanations of AI decisions most aligned with the due-process nature of Article 22: the counterfactual and the experimentation method. The counterfactual is neatly aligned with the fundamental purpose of contesting a decision, seeking to change it. In an Article 22 appeal, wherein the data subject seeks to change an automated decision, they must understand what it would take to change that decision.²⁷² The counterfactual explanation is best suited to help: by definition, it provides an explanation of how to change the AI decision—how to alter the inputs in order to get a different output.²⁷³

The experimentation method of explanation is best suited to check the reliability of the evidence via “questioning” in “person.” A fair review under Article 22 must allow the data subject to challenge the data inputs and the model, which necessitates the ability to check reliability of evidence, which is ensured in part via questioning in person.²⁷⁴ Of course, AI is incapable of being “questioned,” or even being “in person,” in the traditional manner of human cross-examination. But what is important is the interactive inquiry, and while neither popular XAI methods nor methods proposed in legal scholarship are interactive, experimentation is.²⁷⁵ With an experimentation-based explanation, a data subject can experiment with AI—question the AI—by playing with the inputs and observing how the outputs would change.²⁷⁶ This process, akin to cross-examination, provides a check on both the reliability of the data subject’s inputs and the reasonableness of the model’s outputs on those and various other inputs. FICO’s Credit Scores Estimator tool, for example, illustrates how this process could work.²⁷⁷ This experimentation method provides a fuller

270. See, e.g., Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972, 1978 (2017) (“Just as the hearsay dangers are believed more likely to arise and remain undetected when the human source is not subject to the oath, physical confrontation, and cross-examination, black box dangers are more likely to arise and remain undetected when a machine utterance is the output of an inscrutable black box.” (internal quotation marks and footnote omitted)).

271. Miller, *supra* note 140, at 3.

272. See GUIDELINES ON AUTOMATED DECISION-MAKING, *supra* note 167, at 27.

273. See *supra* notes 134–135 and accompanying text.

274. See *supra* notes 267–271 and accompanying text.

275. See *supra* notes 150–151 and accompanying text.

276. See *supra* notes 150–151 and accompanying text.

277. See *FICO Score Estimator*, *supra* note 98. However, the tool has significant limitations, claiming that it “is for informational purposes only and is intended to approximate the FICO Score range based on answers to the questions provided.” *Id.* An ideal tool for experimentation would instead provide access to the data subject’s actual input data, use the real AI model, and provide real outputs rather than ranges.

understanding of the AI model than the counterfactual explanation, but it may need additional research and development to address practical concerns.²⁷⁸

Although these are the two methods best aligned with the underlying principles of Article 22, there remains utility in the other methods of explanation being researched in XAI and proposed in scholarship. For example, because humans often reason based on prototypes,²⁷⁹ explaining using prototypes may provide additional meaning to data subjects. Combining several types of explanation methods into a more comprehensive explanation is desirable; different types can illuminate different parts of the AI model,²⁸⁰ so a combination can provide additional meaning to data subjects and better aid understanding. Not only does this facilitate the data subject's right to contest under Article 22, but it also provides benefits to the AI controller by increasing the data subject's trust in the model.²⁸¹

V. CONCLUSION

Due process is a critical part of the comprehensive package of robust regulation in the GDPR. Ex ante explanations can assist supervisory authorities in identifying systemic problems, but no ex ante scrutiny of highly complex AI can identify all such problems. An ex post review process, providing an ex post explanation, is needed to remedy the unforeseen forms of errors, bias, and discrimination.

Article 22 provides that due process. When an AI makes an important decision, the affected data subject is owed an ex post explanation. It is the view of the EDPB and individual DPAs that Article 22 provides for a review process and an explanation that breathes life into the explicit right to contest. This is necessary to enable data subjects to contest decisions, to help root out unforeseen bias and discrimination, to correct errors, and ultimately to obtain effective remedies. As an analysis of due process demonstrates, the explanation must include the reasons for the decision, i.e., it must be an ex post explanation. It follows that the form of the explanation provided should help the data subject realize their right to contest the decision in a fair review by embodying the principles of due process.

278. See *supra* note 151 and accompanying text.

279. Miller, *supra* note 140, at 3; Adadi & Berrada, *supra* note 8, at 52,153; see *supra* Section II.C and accompanying notes.

280. See Samek & Müller, *supra* note 47, at 10; *supra* Sections II.B–C and accompanying notes.

281. See Samek & Müller, *supra* note 47, at 8; Adadi & Berrada, *supra* note 8, at 52,152; *supra* Sections II.B–C and accompanying notes.

More research in XAI is needed to address fundamental challenges and incorporate human interests, particularly methods of explanation that incorporate interactivity. As researchers and technologists make progress on these challenges and improve the explanatory tools available, the legal obligations of the GDPR will likely, and rightly, strengthen along with them.

The GDPR's right to explanation, backed by the threat of large punitive fines, will incentivize the development of appeals processes and explanations. Moreover, the obligation to explain decisions incentivizes better decision-making in the first place.²⁸² Indeed, perhaps creating these incentives was a goal of the GDPR. In order to comply with Article 22 and overcome practical concerns, scientists and technologists will need to invest more in developing AI explainability. They should do so with a view towards the principles of due process. Not only might their results hold AI to proper account—they might revolutionize AI as we know it.

282. Henry J. Friendly, *Some Kind of Hearing*, 123 U. PA. L. REV. 1267, 1292 (1975).

WHAT WE DON'T KNOW THEY KNOW: WHAT TO DO ABOUT INFERENCES IN EUROPEAN AND CALIFORNIA DATA PROTECTION LAW

Allan E. Holder[†]

I. INTRODUCTION

One day in 2003, an irate man walked into a Target store in the State of Minnesota and demanded to speak to a manager.¹ The man's high school-aged daughter had been sent an advertisement with coupons for baby clothes and accessories, and he wanted to know why the store was encouraging his teenaged daughter to become pregnant.² A few days later, when the store's manager called to apologize to the man, the man in turn apologized profusely to the manager, revealing that his daughter had admitted to him that she was pregnant and due the following August.³ The man's pregnant daughter had not disclosed this information to Target, but Target had ascertained it by analyzing the girl's shopping history with a computer program intended to predict the likelihood of a woman being pregnant.⁴ The concerned man's daughter had likely purchased several out of a list of "twenty-five different products that, when analyzed together, allowed [Target] to . . . guess what trimester she was in—and estimate her due date—so Target could send her coupons when she was on the brink of making new purchases."⁵

Target's conclusion that the young girl was pregnant was an inference. An inference is, at its simplest, information that can be reasonably predicted from pre-existing data.⁶ A person looking out their window to see that the sun is shining brightly and concluding that it is a hot day outside has made an inference about the weather, much like Target made an inference about the young girl's pregnancy. Nowadays, in the rare occasions in which inferences are included in the discourse over data privacy, it is in the context of large

DOI: <https://doi.org/10.15779/Z38MP4VP1V>

© 2020 Allan E. Holder.

[†] J.D., 2021, University of California, Berkeley, School of Law.

1. CHARLES DUHIGG, *THE POWER OF HABIT: WHY WE DO WHAT WE DO IN LIFE AND BUSINESS* 196 (Random House ed., 1st ed. 2012).

2. *Id.*

3. *Id.*

4. *Id.* at 195.

5. *Id.*

6. See Michal Kosinski, David Stillwella & Thore Graepel, *Private traits and attributes are predictable from digital records of human behavior*, 110 *PROC. NAT'L ACAD. SCI.* 5802 (2013).

technology companies whose primary business models depend on using inferences created from data they collect in order to serve their users targeted advertisements.⁷ Nonetheless, the collection of data and subsequent creation of inferences is not a phenomenon created by the growth of technology giants like Facebook and Google. Historically, corporate actors like insurance companies⁸ and government actors like federal agencies⁹ have also availed themselves of the practice of analyzing multiple data sets to infer or predict information about a subject.

In the information technology age, however, technological advancements and the popularization of the internet have enabled the creation of sets of data whose sheer size and level of detail had previously been unimaginable.¹⁰ Companies can now collect information about a user's recent searches, purchases, and current location, as well as a user's preference for books, music, sports, and restaurants (among a host of other kinds of data points).¹¹ In turn, these organizations can analyze the data they acquire to predict, for instance, a user's age, gender, occupation, and education level.¹² These massive data sets, full of information both provided by and gathered about users (with and without their actual knowledge or consent), lend themselves to being processed by computer programs that find relationships between the data points and infer increasingly more detailed and intimate information about the user, as Target did.¹³

Inferences are often then used by organizations to make decisions about particular users, regarding everything from what ads to serve users based on their demographic profile, to creditworthiness, to suitability for employment, to insurance risk.¹⁴ And even though companies have begun to create

7. *See, e.g.*, SHOSHANA ZUBOFF, *THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER* 78 (PublicAffairs ed., 1st ed. 2019).

8. PAM DIXON & ROBERT GELLMAN, *WORLD PRIVACY F., THE SCORING OF AMERICA: HOW SECRET CONSUMER SCORES THREATEN YOUR PRIVACY AND YOUR FUTURE* 16 (2014).

9. *See, e.g., id.* at 62 (“The Health and Human Services Department effectively created a score that ultimately measures how sick a person is.”); *id.* at 76 (“...the [Department of Homeland Security’s] program collects data about passengers and links the data with other sources of information to establish a risk score for each passenger. The Transportation Security Administration uses the scores to screen passengers.”).

10. VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, *BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK* 6 (2013).

11. *See* Kosinski, Stillwell & Graepel, *supra* note 6, at 5802.

12. *Id.*

13. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 10, at 156.

14. *See* FED. TRADE COMM’N, *DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY* i, 19 (2014) (“The FCRA covers the provision of consumer data by

increasingly sophisticated, sensitive inferences and make decisions based on them, no organization can ever be one hundred percent certain of the veracity of an inference, and sometimes the inferences drawn are flat-out wrong.¹⁵ Given inferences' increased sophistication and sensitivity, lack of certainty regarding their accuracy, and their role in decision-making, inferences have become increasingly important to both data-reliant organizations and privacy advocates.

In recent years, the European Union and several U.S. states have enacted new privacy legislation in order to grant users control over how their data is collected and used. The European General Data Protection Regulation (GDPR) came into effect on May 25, 2018, with the intent of acknowledging the rapidly developing field of consumer technology and its benefits for economic and social relations, while also establishing and protecting the principle that “the processing of personal data should be designed to serve mankind.”¹⁶ While the GDPR grants data subjects the rights to know about, rectify, delete, object to the processing of, and transfer personal data that a data controller might hold about them, the regulation significantly curtails subjects' rights when it comes to inferences.¹⁷ In fact, there are gaps in the GDPR's jurisdiction over inferences, some of which have begun to be addressed only recently through jurisprudence from the European Court of Justice, and some of which remain unaddressed.¹⁸

In 2018, the State of California became the first U.S. state to enact a comprehensive data privacy law, setting an initial standard for what comprehensive data protection could look like in the United States and spurring a number of other states to follow its lead.¹⁹ The California Consumer Privacy Act (CCPA) was passed with the goal of granting consumers similar

consumer reporting agencies where it is used or expected to be used for decisions about credit, employment, insurance, housing, and similar eligibility determinations”), (“In developing their products, the data brokers use not only the raw data that they obtain from their sources, such as a person's name, address, home ownership status, age, income range, or ethnicity (‘actual data elements’), but they also derive additional data (‘derived data elements’).”).

15. Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data*, 2019 COLUMBIA BUS. L. REV. 494, 509 (2019).

16. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L119) 1, Recital 4 [hereinafter GDPR].

17. Wachter & Mittelstadt, *supra* note 15, at 499.

18. *Id.*

19. See Sarah Rippey, US State Comprehensive Privacy Law Comparison, IAPP, <https://iapp.org/resources/article/state-comparison-table/> (last visited Apr. 18, 2021).

rights to those of the GDPR when consumer data is held by businesses.²⁰ Enforcement of the CCPA began on January 1, 2020, and the statute called for the California Attorney General's Office to adopt regulations for the CCPA's enforcement.²¹ In November 2020, voters in California approved passage of the California Privacy Rights Act, supplementing the CCPA with additional GDPR-like provisions.²² While the California data privacy regime attempts to grant consumers rights with respect to inferences by classifying them as "personal information," this Note argues that language within the CCPA complicates interpretation of all subsequent regulations and statutes in the California regime by limiting certain rights to information directly collected from a consumer.²³

While both the European Union and California have acknowledged the importance of inferences, they have only done so to an extent, and inferences' patchy legal status makes for an uncertain and confusing regulatory treatment regarding this type of information. The current gaps in the regulation of inferences left by these two major data privacy regimes are worrisome because the existence of these gaps constitutes suboptimal protection of inferences, going against the most basic justifications for the existence of each data protection regime. This Note argues that comprehensive regulation of inferences should be a top priority for legislators, regulators, and policy makers around the world as they undertake regulation of data privacy. This is because the subject of inferences touches directly on the very concerns over personal autonomy, safety, and dignity that historically justify the recognition of the privacy rights of individuals. The importance of the regulation of inferences is not underscored only by historical precedent. This Note also argues that modern data collection and inference creation practices create unprecedented risks to vulnerable populations that necessitate the strongest protections over personal data available under every data protection regime.

This Note proceeds as follows. Part II provides an overview of contemporary data collection practices and how they fuel the economic model that drives large technology organizations and a significant portion of the digital economy. Part III explains the current legal status of inferences under the GDPR in the European Union and under the California data privacy framework. In Part IV, I argue that inferences are personal data and users ought to have the highest degree of control available over them. I further

20. AB-375, 2017-2018 Assemb., Reg. Sess. (Cal. 2018).

21. CAL. CIV. CODE § 1798.185.

22. Erin Illman, Junaid Odubeko, Steve Snyder & Bradley Aaron Boulton Cummings, *Steps for Proactive CPR/A Compliance*, BLOOMBERG LAW (2020) (on file with the author).

23. See *infra* Section III.B.

contend that inferences' high level of sophistication and their close relation to identified or identifiable subjects makes them equivalent to the sort of information that the statutes in question cite as the original justification for the regulation of information privacy and protection of personal data. Moreover, inferences leave subjects vulnerable to the same kind of harms—to dignity, reputation, and personality—as other data that have been labeled personal information and over which users have been given rights. In addition to these harms, inferences by their nature increase the likelihood that an invasion of privacy could result in immediate or near-immediate threats to the physical security, autonomy, livelihood, or future prospects of the user. To illustrate the latter point, I draw on some experiences of women, LGBTQ people, and people of color to provide examples of the kind of novel invasions of privacy made possible by inferences that could uniquely affect marginalized groups. In Part V, I present a few anticipated arguments from stakeholders who might argue against strong user rights and protections over inferences. I also attempt to identify the actors who will be involved in resolving these conflicts as well as arguments that they ought to consider in adjudicating them.

II. INFERENCES AND THEIR IMPORTANCE TO THE DIGITAL ECONOMY

In the context of digital data, and for the purposes of this Note, inferences are defined as “information relating to an identified or identifiable natural person created through deduction or reasoning rather than mere observation or collection from the data subject,” the data subject being the user to whom the information pertains.²⁴ In the aforementioned Target example, the company collected information about the teenage girl's purchases and then made the inference that the girl was pregnant. Just as Target sought to benefit from the creation of inferences based on a customer's information, actors of all stripes have come to depend on data collection and subsequent inference creation to better market their products and services, catalyzing the emergence of an entire economy based on the digital collection of user data.

Inferences are used primarily in two contexts: profiling and scoring.²⁵ Profiling is the practice of “using data from various sources to infer something about an individual, based on qualities of others who appear statistically similar.”²⁶ A business might, for instance, scan a list of customers that

24. Wachter & Mittelstadt, *supra* note 15, at 515.

25. See FED. TRADE COMM'N, *supra* note 14, at 19; see generally Dixon & Gellman, *supra* note 8.

26. Article 29 Data Protection Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, at 7.

purchased camping gear in the last year, identify that they were all men from a certain postal code whose credit cards had a certain range of credit limits, and then use this information to put similar customers in the company's larger database in a category called "Customers Interested in Buying Camping Gear."²⁷ By way of example, a company called Recorded Future "captures historical data on consumers and companies across the [i]nternet," compiling it into profiles "to predict the future behavior of those consumers and companies."²⁸

Scoring is the related practice of using inferences to rank users based on certain traits and behaviors.²⁹ For instance, some companies turn their analyses of customer data and interests "into marketing scores that . . . rank clients' customers on the basis of how likely they are to respond to particular marketing efforts or to make a purchase, their presence on the web or their influence over others, or other metrics."³⁰ Political campaigns have ranked television viewers by likelihood of voting for a candidate based on what they watch on cable television and when.³¹

Profiling and scoring can be and have been used by all manner of data-reliant organizations to help make determinations about users, concerning everything from ad targeting to employability, insurance risk, and creditworthiness.³² However, over the course of the past two decades, technological advancements have helped usher in what scientists have called the era of Big Data, which could be defined (somewhat imperfectly) as a new period in the history of data wherein "the volume of information ha[s] grown so large that the quantity being examined no longer fit[s] into the memory that computers use for processing."³³ The popularization of certain technologies, both digital and physical, has allowed for the collection of massive and unprecedented quantities and types of user data. These large, comprehensive data sets have in turn enabled the creation of more complex and sophisticated inferences about users, increasing the ease and frequency with which scores are assigned, or profiles are created. In the physical realm, the manufacturers of everyday items humans depend on, such as cellphones and laptops (but also clothing, lightbulbs, microwaves, and toothbrushes), have begun augmenting

27. FED. TRADE COMM'N, *supra* note 14, at 19.

28. *Id.* at 9.

29. See Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 2 (2014).

30. FED. TRADE COMM'N, *supra* note 14, at iii.

31. See Alice E. Marwick, *How Your Data Are Being Deeply Mined*, N.Y. REV. BOOKS (Jan. 9, 2014).

32. Citron & Pasquale, *supra* note 29, at 4.

33. MAYER-SCHÖNBERGER & CUKIER, *supra* note 10, at 6.

and enhancing these products “through the use of emerging technologies—sensors, actuators, wireless connection, and embedded processing,” all of which are able to collect data about users.³⁴ Through the chips that allow smartphones to make phone calls and connect to other accessories, cellular service providers³⁵ and even retailers³⁶ are able to track and log the physical locations of users with great precision. As we move further into the Internet of Things (IoT) era, suddenly our toothbrushes can log what times of day we brush our teeth and for how long,³⁷ our bathroom scales can keep a record of our weight fluctuations,³⁸ and our umbrellas can determine where we walked to and from in the rain.³⁹

In the digital sphere, large technology services have developed the ability to follow users around the internet and observe their online habits, in order to add this data to the set of data they acquire directly by requesting it from the user. By way of example, Facebook’s “Like” button plug-in, which “take[s] the form of a snippet of code to be added to a page,”⁴⁰ allows Facebook to recognize users across sites where the button has been encoded regardless of whether the user has “Liked” something or not.⁴¹ Five months after the launch of the Like button in 2010, two million websites throughout the web had added

34. DAVID ROSE, *ENCHANTED OBJECTS: DESIGN, HUMAN DESIRE, AND THE INTERNET OF THINGS* 47 (Simon & Schuster ed., 1st ed. 2014).

35. See Nancy K. Oliver, *Location, Location, Location: Balancing Crime Fighting Needs and Privacy Rights*, 42(3) U. BALTIMORE L. REV. 485, 490 (2013).

36. See Michael Kwet, *In Stores, Secret Surveillance Tracks Your Every Move*, N.Y. TIMES (June 14, 2019), <https://www.nytimes.com/interactive/2019/06/14/opinion/bluetooth-wireless-tracking-privacy.html>.

37. See *Why Switch To A Bluetooth Electric Toothbrush?*, ORAL-B, <https://oralb.com/en-us/why-switch/> (last visited Jan. 12, 2021) (“Oral-B’s latest electric toothbrushes connects to the Oral-B app on your phone. The results? You’ll get real-time feedback on your brushing. You’ll know if you’re brushing too hard, if you’ve brushed long enough and even if your brushing habits have improved over time.”).

38. See *Body Composition Smart Scales*, WITHINGS, <https://www.withings.com/us/en/scales> (last visited Jan. 12, 2021) (“Data from every weigh-in automatically syncs to your smartphone via the free Health Mate app, available for iOS and Android.”).

39. See Harry Hu, *HAZ: The World’s 1st Motorized Smart Umbrella*, INDIEGOGO, <https://www.indiegogo.com/projects/haz-the-world-s-1st-motorized-smart-umbrella#/> (last visited Feb. 21, 2020) (explaining the ability to perform location tracking with the umbrella and a smartphone app in order to locate the umbrella).

40. Tom Simonite, *Facebook’s Like Buttons Will Soon Track Your Web Browsing to Target Ads*, MIT TECH. REV. (Sept. 16, 2015), <https://www.technologyreview.com/s/541351/facebook-like-buttons-will-soon-track-your-web-browsing-to-target-ads/>.

41. Cotton Delo, *Facebook To Use Web Browsing History for Ad Targeting*, ADAGE (June 12, 2014), <https://adage.com/article/digital/facebook-web-browsing-history-ad-targeting/293656>.

the plugin to their pages.⁴² Google engages in similar practices by “us[ing] web cookies to track browsing behaviour online by [users’] IP address to deliver targeted ads.”⁴³ In 2007, Google acquired the advertising network DoubleClick, and with it, the massive database of user web-browsing records DoubleClick compiled by relying on “non-personally-identifiable information” to create user profiles.⁴⁴ While at first DoubleClick’s anonymous browsing records were separate from Google’s user profiles (which include personally identifiable information), in 2016 Google changed its policies and indicated that browsing records may in the future be combined with what the company already knows from a user’s activities on Gmail and other Google services.⁴⁵

Whereas in the past it was more common for companies to rely primarily on information collected directly from users with their awareness or consent, in the age of Big Data, companies can procure information about users without any action on the user’s part. Unburdening the user from having to actively provide data allows data-reliant organizations to collect a significantly greater number of data points that the user creates in the normal course of their daily activities. As humans have become more and more reliant on the internet and connected devices, Big Data has demanded the development of novel approaches to analyzing data due to the sheer quantity now available to be analyzed. It has also allowed analysts and organizations to do more with data than they had been able to before—“to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more.”⁴⁶ In short, it has allowed for the creation of highly sophisticated inferences.

Inferences have been and continue to be used by organizations to contribute to decisions made about users. In more recent times, however, internet companies like Facebook and Google have based their business

42. Leena Rao, *Five Months In, 2 Million Websites Using Facebook’s New Social Plugins*, TECHCRUNCH (Sept. 29, 2010), <https://techcrunch.com/2010/09/29/five-months-in-2-million-websites-using-facebooks-new-social-plugins/>.

43. Olivia Solon, *Google’s ad tracking is as creepy as Facebook’s. Here’s how to disable it*, THE GUARDIAN (Oct. 21, 2016), <https://www.theguardian.com/technology/2016/oct/21/how-to-disable-google-ad-tracking-gmail-youtube-browser-history>.

44. See Suzanne Monyak, *Google Changed a Major Privacy Policy Four Months Ago, and No One Really Noticed*, SLATE (Oct. 21, 2016), <https://slate.com/technology/2016/10/google-changed-a-major-privacy-policy-and-no-one-really-noticed.html>.

45. Julia Angwin, *Google Has Quietly Dropped Ban on Personally Identifiable Web Tracking*, PROPUBLICA (Oct. 21, 2016), <https://www.propublica.org/article/google-has-quietly-dropped-ban-on-personally-identifiable-web-tracking>.

46. See MAYER-SCHÖNBERGER & CUKIER, *supra* note 10, at 6.

models on using inferences to serve users targeted ads.⁴⁷ The companies take advantage of massive amounts of data about their billions of users to then score and profile them. They subsequently sell access to their users' attention to advertisers that have products or services that could appeal to certain types of user, or that certain types of user are more likely to purchase.⁴⁸ The revenue models of companies like Facebook and Google, which rely on selling users' attention to advertisers, effectively behoove them to collect as much information about users as they can in order to infer increasing amounts of information. With these inferences, they can then provide the most precise access to advertisers and remain competitive in their particular market sectors.

Data points that are usable for inferences are collected not only by large technology companies, but also by government agencies, banks, physical retailers of varying sizes, product manufacturers, and a myriad other kinds of organizations.⁴⁹ The Big Data economy has spawned companies called data brokers, whose main purpose is to "collect consumers' personal information and resell or share that information with others."⁵⁰ The existence of data brokers indicates that information collected from a user, with or without the user's awareness, is highly likely to end up in the hands of companies or organizations with which the user never intended to share this information.⁵¹ It is foreseeable, then, for a health insurance company to preemptively learn that a person suffers from a chronic health condition, based on the user's shopping history showing that they purchase a certain medical aid every month and the user having searched for home remedies for certain symptoms. Additionally, data brokers themselves make inferences about consumers and use them to sort them into profiles, which they then provide to companies in order to serve consumers targeted ads.⁵²

Large technology companies make use of artificial intelligence to process the enormous data sets that they compile from various sources, identify relationships between seemingly unrelated data points, and create inferences about users. Facebook, for instance, makes use of a type of artificial

47. See generally Zuboff, *supra* note 7; Caitlin Dewey, *98 Personal Data Points that Facebook Uses to Target Ads to You*, WASH. POST (Aug. 19, 2016), <https://www.washingtonpost.com/news/the-intersect/wp/2016/08/19/98-personal-data-points-that-facebook-uses-to-target-ads-to-you/>.

48. Kurt Wagner, *This Is How Facebook Uses Your Data for Ad Targeting*, RECODE (Apr. 11, 2018), <https://www.vox.com/2018/4/11/17177842/facebook-advertising-ads-explained-mark-zuckerberg>.

49. See generally Dixon & Gellman, *supra* note 8.

50. FED. TRADE COMMISSION, *supra* note 14, at i.

51. *Id.*

52. *Id.* at 47.

intelligence called “deep neural networks” to process the large amounts of user data it collects in order to improve its ad targeting.⁵³ YouTube uses the same type of artificial intelligence to fuel its recommendations engine, which makes inferences about what kind of videos users will want to see based on a user’s YouTube activity history, past searches, and demographic information.⁵⁴

Deep neural networks are a type of artificial intelligence modeled after human brains so that computers can learn things, adapt the lessons they learn to new information, and make decisions in a human-like manner.⁵⁵ Engineers at large data-reliant organizations depend on the massive sets of data available to them to train neural networks. By way of illustration, Facebook could feed neural networks all the user data they have as well as the relationships within the data set, and by providing human validation of the outputs that the machine correctly produces, the model can slowly “learn” the correct relationships between the data points so that it is eventually able to make “predictions” about users based only on input data. Neural networks are so complex that their exact functioning is very difficult to understand, even for the experts who build them, and their complexity also makes it difficult to audit how they create sometimes very accurate relationships among data points.⁵⁶

In this Note, and in most of the literature on machine learning, a “prediction” is meant to refer to an educated guess on the part of the artificial intelligence about unknown data based on known data. When the data set used to train a deep neural network consists of or includes personal data, the unknown facts could be future events that the artificial intelligence deems likely to happen (“User will likely purchase item X in the next seven days”), or current true facts about a user for which the artificial intelligence simply does not have confirmation (“User is a resident of California”). Both types of predictions can be viewed as inferences: the former about a user’s propensity to a certain kind of behavior, and the latter about a certain fact about the user being true.

53. Cade Metz, *Building AI Is Hard—So Facebook Is Building AI That Builds AI*, WIRED (2016), <https://www.wired.com/2016/05/facebook-trying-create-ai-can-create-ai/>.

54. Paul Covington, Jay Adams & Emre Sargin, *Deep Neural Networks for YouTube Recommendations*, RECSYS’16: PROC. 10TH ACM CONF. ON RECOMMENDER SYSTEMS 191, 192 (2016).

55. Bernard Marr, *What Are Artificial Human Networks – A Simple Explanation for Absolutely Anyone*, FORBES (Sept. 24, 2018), <https://www.forbes.com/sites/bernardmarr/2018/09/24/what-are-artificial-neural-networks-a-simple-explanation-for-absolutely-anyone/#23ad3ba12457>.

56. Walter A. Mostowy, Note, *Explaining Opaque AI Decisions: How to Satisfy the GDPR’s Right to an Ex Post Explanation*, 35 BERKELEY TECH. L.J. 1291 (2020).

As previously mentioned, deep neural networks and other forms of artificial intelligence make educated guesses about facts still uncertain to them, meaning that, barring a user confirming the veracity of an influence, data controllers are in possession of a host of inferences about a user that might range from being slightly inaccurate to flat-out erroneous. This poses dangers particularly when data-reliant organizations use inferences to make decisions that might alter the lives of the users, as will be explained in Part V of this Note.

One would expect a type of information that is essentially the driver of the modern digital economy—and which major data controllers invest so much to develop—to be stringently regulated. However, the widespread use of inferences and the enactment of comprehensive data protection statutes are such recent developments that the regulation of inferences so far remains imperfect, incomplete, and uncertain.

III. THE CURRENT LEGAL STATUS OF INFERENCES

Comprehensive data protection law in the European Union and the State of California both address the existence and especially sensitive nature of inferences. However, both regimes fail to fully protect inferences from misutilization, resulting in uncertainty over their treatment in both regimes.

A. THE EUROPEAN UNION

The European Union began enforcement of the GDPR on May 25, 2018.⁵⁷ It was originally adopted to replace the 1995 Data Protection Directive, which provided a framework on which E.U. member states could base their national data protection laws.⁵⁸ The GDPR was enacted as a binding regulation in order to create a standardized data protection regime across all E.U. member states, and partly also to “address contemporary privacy challenges, such as those posed by the [i]nternet, social media, mobile apps, cloud computing, ‘big data,’ and behavioral marketing.”⁵⁹

57. Nikhil Kalyanpur & Abraham Newman, *Today, A New E.U. Law Transforms Privacy Rights for Everyone. Without Edward Snowden, It Might Never Have Happened*, WASH. POST (May 25, 2018), <https://www.washingtonpost.com/news/monkey-cage/wp/2018/05/25/today-a-new-eu-law-transforms-privacy-rights-for-everyone-without-edward-snowden-it-might-never-have-happened/>.

58. See GDPR, *supra* note 16, Recitals 9, 10, art. 94.

59. W. Scott Blackmer, *GDPR: Getting Ready for the New EU General Data Protection Regulation*, INFOLAWGROUP (May 5, 2016), <https://web.archive.org/web/20180514111300/https://www.infolawgroup.com/2016/05/articles/gdpr/gdpr-getting-ready-for-the-new-eu-general-data-protection-regulation/>; see also GDPR, *supra* note 16, Recital 6.

The GDPR grants data subjects a number of rights over their “personal data,” defined as “any information relating to an identified or identifiable natural person.”⁶⁰ Under the GDPR, data subjects have the following rights:

- to basic information about the personal data collected regardless of its source (Art. 13–14);
- to access the data that has been collected (Art. 15);
- to rectify inaccurate data (Art. 16);
- to have personal data permanently erased (Art. 17);
- to restrict the processing of personal data (Art. 18);
- to transfer personal data between controllers (Art. 20);
- to object to processing of personal data, including for the purposes of direct marketing (Art. 21); and
- to not have decisions made about the data subject on the basis of automated processing of personal data (Art. 22).⁶¹

The regulation further earmarks certain special categories of personal data that data subjects have additional rights over, such as data “revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership,” as well as “genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person’s sex life or sexual orientation.”⁶² Processing of these special categories of personal data is prohibited save for a number of exceptions, including when the data subject has granted explicit consent,⁶³ when they have manifestly made the data public,⁶⁴ or when the processing is necessary for reasons of substantial public interest.⁶⁵

While the GDPR itself does not designate inferences as personal data (or even mention inferences at all), the views of the Article 29 Data Protection Working Party (hereinafter, “the Working Party”) and recent European Court

60. GDPR, *supra* note 16, art. 4(1).

61. GDPR, *supra* note 16.

62. *Id.* art. 9(1).

63. *Id.* art. 9(2)(a).

64. *Id.* art. 9(2)(e).

65. *Id.* art. 9(2)(g).

of Justice (ECJ) jurisprudence provide a sketch of the current (somewhat convoluted) legal approach to inferences in the European Union.⁶⁶

The Working Party adopted the broadest definition of personal data. The positions of the Working Party and its successor, the European Data Protection Board (EDPB), are recommendations for the practical application of European data protection laws, and have no binding legal effect.⁶⁷ The GDPR established the EDPB to replace the Working Party as soon as the regulation came into effect in May 2018, and the EDPB endorsed all of the Working Party's recommendations.⁶⁸ According to one of the Working Party's position papers, personal data is created by any data processing whose content, purpose, or result relates to an identifiable person directly or indirectly.⁶⁹ The Working Party's test makes it so that data can still be classified as personal data even if it does not describe an identifiable person, or even if it will not be used to make a decision about a person, so long as it has the potential to impact "an identifiable person's rights and interests."⁷⁰ Consequently, a data inference that does not include an identified data subject but still has the potential to affect the data subject's life is personal data in the view of the Working Party. While the Working Party's views, as previously mentioned, are merely recommendations, they have been influential in ECJ data protection jurisprudence and are expected to continue to play a role in the court's reasoning in these matters.⁷¹

In contrast to the Working Party, the ECJ has generally adopted more limited interpretations of the concept of personal data, and its decisions are legally binding across the European Union. In two recent cases, *YS, M and S v. Minister voor Immigratie, Integratie en Asiel*⁷² and *Peter Nowak v. Data Protection*

66. Wachter & Mittelstadt, *supra* note 15, at 498.

67. Tim Wybitul, *GDPR Guidance – European Data Protection Board Adopts Art. 29 Working Papers*, HOGAN LOVELLS CHRONICLE OF DATA PROTECTION (May 2018), <https://www.hldataprotection.com/2018/05/articles/international-eu-privacy/gdpr-guidance-european-data-protection-board-adopts-art-29-working-papers/>.

68. *Id.*

69. Article 29 Data Prot. Working Party, *Guidelines on the Right to Data Portability*, 16/EN, WP242rev.01, at 9–11 (Dec. 13, 2016), https://ec.europa.eu/newsroom/document.cfm?doc_id=44099.

70. *See* Wachter & Mittelstadt, *supra* note 15, at 518. While the Working Party has not given examples, the reading of the text implies that, for instance, information that was inferred, anonymized, and aggregated can be considered personal data if it is likely to become re-identified as a consequence of an inference attack.

71. *See* Wybitul, *supra* note 67; Wachter & Mittelstadt, *supra* note 15, at 498 n.2.

72. Joined Cases C–141 & 372/12, *YS, M and S v. Minister voor Immigratie, Integratie en Asiel*, 2014 E.C.R. I-2081 [hereinafter *YS, M and S*].

Commissioner,⁷³ the ECJ shined some light on its interpretation of what “personal data” encompasses, but the decisions split on some significant points in a way that leaves the legal status of inferences unresolved.

First, the cases split over whether “personal data” includes “opinions, reasoning, and assessments that underlie” final decisions—in other words, the intermediate inferences that lead to a final, inferred result.⁷⁴ The ECJ in *Nowak* held that any comments made by an examiner with respect to the exam answers of a data subject—which constitute inferences in that they are assessments of the data subject’s knowledge and competence in the exam’s field, based on information provided by the data subject⁷⁵—are personal data.⁷⁶ In *YS, M and S*, the ECJ held that an explanatory legal analysis included in the file relating to a data subject’s application for a residence permit is not personal data even though it may contain personal data.⁷⁷ The ECJ in this case characterized the legal analysis at issue as “information about the assessment and application by the competent authority of [the relevant law] to an applicant’s situation,” therefore designating it as an inference created from subject-provided data.⁷⁸ However, the court still concluded that facts about a data subject, “such as the applicant’s name, date of birth, nationality, gender, ethnicity, religion and language,” constitute personal data in the context of an application for a residence permit.⁷⁹

The court seems to contradict itself, stating in *Nowak* that inferences in the form of assessments, opinions, or reasonings are classified as personal data under the GDPR, while appearing to state the contrary in *YS, M and S*. The key lies partly in the court’s implied argument in *Nowak*. The court proposes that if having the right to correct, erase, or block processing of certain information would serve the GDPR’s purpose of “guaranteeing the protection of the [data subject]’s right to privacy with regard to the processing of data relating to [them],” then the data subject should have a right to *access* information about them in order to exercise those very rights.⁸⁰ If so, then in order for the data subject to have the right to access the information, the information needs to be granted the designation of personal data.⁸¹

73. Case C–434/16, *Peter Nowak v. Data Prot. Comm’r*, 2017 E.C.R. I-582 [hereinafter *Nowak*].

74. Wachter & Mittelstadt, *supra* note 15, at 537–38.

75. *Nowak*, *supra* note 73, ¶ 43.

76. *Id.* ¶ 62.

77. *YS, M and S*, *supra* note 72, ¶ 39.

78. *Id.* ¶ 40.

79. *Id.* ¶ 38.

80. *Nowak*, *supra* note 73, ¶¶ 56–57.

81. *Id.*

In *Nowak*, the ECJ found that the examinee had a legitimate interest in being able to, for instance, correct potential mix-ups where another examinee's answers, or comments by the examiner (i.e., inferences), are mistakenly ascribed to the data subject.⁸² The ECJ also found a legitimate interest of the examinee in being able to prevent the sharing or publication of their answers to the examination.⁸³ Consequently, the examinee deserves the right to protect their right to privacy by accessing the information, making the information personal data.

In contrast, in *YS, M and S*, the court held that the legal analysis pertaining to a decision of residency, considered separately from the facts about the data subject, cannot be checked for accuracy, as it is an application of law to facts.⁸⁴ It also cannot be corrected, given that any attempts at correction would constitute an appeal to an administrative decision, a procedure that the GDPR does not provide for.⁸⁵ Extending a right to access to the legal analysis would therefore not serve the GDPR's purpose of helping the data subject in protecting their right to privacy by correcting, erasing or blocking processing of information about them. Since the data subject does not have a claim to a right to access the legal analysis, the information is not personal data.

Both cases left unresolved the issue of whether final inferences (the data that results from processing) are personal data. Sandra Wachter and Brent Mittelstadt indicate that it is highly likely that final inferences will be considered personal data, given that that in this regard the court seemed to rely on the views of the Working Party, which include output data in their definition of personal data.⁸⁶

In other respects, the ECJ's position conflicted with that of the Working Party. The ECJ's view is that the rights of data subjects over inferences are limited, and from the two rulings it becomes clear that whether the rights apply, according to Wachter and Mittelstadt, "must be interpreted according to the purposes for which the data was collected."⁸⁷ The ECJ rulings in the cases "clarify that the remit of data protection law is not to assess the accuracy of the reasoning behind decisions and assessments, or the accuracy of the decisions and assessments themselves," and the court grounds its holding on the fact that, in the GDPR itself, there are broad exemptions to the right to

82. *Id.* ¶ 54.

83. *Id.* ¶ 50.

84. *YS, M and S*, *supra* note 72, ¶ 45.

85. *Id.* ¶ 46.

86. Wachter & Mittelstadt, *supra* note 15, at 538.

87. *Id.* at 538.

access.⁸⁸ The rights to correct, erase, or block processing of inaccurate data seem to exist simply so that data subjects can verify that the initial data that will be processed to create inferences and/or make decisions about the data subjects (like the examination data in *Nowak* and the pre-legal analysis facts in *YS, M and S*) are complete and accurate. Data subjects do not have the same rights over the data resulting from the processing because such data in some cases will not be processed further and, in some cases like in *YS, M and S*, processing does not impinge on the data subject's right to privacy.

As it stands, the legal status of inferences under the GDPR is puzzling, and certainty regarding said status is in its infancy. While the Working Party embraces a broad definition of the concept of personal data that would include inferences, data controllers in the European Union are bound only by ECJ jurisprudence, which seems to imply that inferences deserve personal data status only as long as data subjects have an interest in correcting, erasing, or blocking processing of them.

B. CALIFORNIA

Given the relative recency of the California data privacy framework, there have been no significant public debate or litigation pertaining to the subject of inferences. A reading of the regime's statutes and the attorney general's subsequent regulations, however, reveals that the legal status of inferences under the framework is very uncertain: perhaps more so than in Europe, and in ways that invite future amendments and litigation regarding this unique type of data.

The California Consumer Privacy Act (CCPA) was passed on June 28, 2018, and its enforcement began on January 1, 2020.⁸⁹ The legislation designated the California Attorney General as its primary enforcer,⁹⁰ and also required them to enact certain regulations to make the legislation effective.⁹¹ On November 3, 2020, the California electorate voted to approve the California Privacy Rights Act (CPRA), which amended key provisions of the CCPA to mirror some of the provisions of the GDPR.⁹² The CPRA created a new agency called the California Privacy Protection Agency ("the Agency"), which is slated to take over from the California Attorney General as privacy

88. *Id.* at 539–40.

89. *See* CAL. CIV. CODE § 1798.100.

90. *See id.* § 1798.155(b).

91. *Id.* § 1798.185(a).

92. Illman, Odubeko, Snyder & Boulton Cummings, *supra* note 22; *see also*, California Privacy Rights Act of 2020, Proposition 24, 1879 (19-0021A1), Amends Consumer Privacy Laws—Initiative Statute, https://www.oag.ca.gov/system/files/initiatives/pdfs/19-0021A1%20%28Consumer%20Privacy%20-%20Version%203%29_1.pdf.

enforcer “beginning the later of July 1, 2021, or six months after the Agency provides notice to the Attorney General that it is prepared to begin rulemaking” under the CPRA.⁹³ Taken as a whole, the CCPA, the regulations enacted by the California Attorney General pursuant to the CCPA, and the CPRA constitute the California data privacy framework.

The CCPA grants consumers a number of rights over “personal information,” defined in the statute as “information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household.”⁹⁴ The statute goes on to list categories of information that shall be considered personal information under the statute if they are capable of being “reasonably linked, directly or indirectly,” with a particular consumer or household.⁹⁵ These categories include unique or semi-unique identifiers (like real names and aliases, but also IP addresses and postal addresses), biometric information, internet activity information, geolocation data, employment information, education information, and “[c]haracteristics of protected classifications under California or federal law.”⁹⁶ Notably, the final category that shall be considered personal information according to the CCPA is “[i]nferences,” thereby designating as personal information any information mentioned in the previous categories that has been combined or processed “to create a profile about a consumer reflecting the consumer’s preferences, characteristics, psychological trends, predispositions, behavior, attitudes, intelligence, abilities, and aptitudes.”⁹⁷ Further, the act defines “infer” or “inference” as “the derivation of information, data, assumptions, or conclusions from facts, evidence, or another source of information or data.”⁹⁸

While the CCPA’s definition of inference and its designation of inferences as personal information are seemingly comprehensive enough to secure robust protection of the sensitive information that inferences can carry, their effectiveness is impeded by the statutory language that actually grants consumers rights over their data. The CCPA grants consumers six rights regarding businesses’ use of consumers’ personal information, four of which closely mirror rights that the GDPR grants users over their personal data. However, when the CCPA grants consumers the right to access,⁹⁹ delete,¹⁰⁰ or

93. CPRA Sec. 21, § 1798.185(d).

94. CAL. CIV. CODE § 1798.140(o)(1).

95. *See id.*

96. *Id.* § 1798.140(o)(1)(A)–(J).

97. *Id.* § 1798.140(o)(1)(K).

98. *Id.* § 1798.140(m).

99. *Id.* § 1798.100.

100. *Id.* § 1798.105.

request disclosure of information collected,¹⁰¹ the right only applies to information *collected* about a consumer. The right to request disclosure of information sold similarly grants the consumer the right to compel a business to disclose the categories of information “collected” about a consumer, before allowing the consumer to demand a business disclose the categories of personal information sold.¹⁰² The CCPA appears to internally contradict itself, creating uncertainty as to the existence and breadth of rights consumers have over inferences.

The statute’s own definition of inference-creation as “derivation” of data from existing information does not fit comfortably within any of those activities enumerated in the statute’s definition of “collection.” The statute defines “collection” as the “buying, renting, gathering, obtaining, receiving, or accessing [of] any personal information pertaining to a consumer by any means,” and it advises that the definition “includes receiving information from the consumer, either actively or passively, or by observing the consumer’s behavior.”¹⁰³ The creation of inferences, however, is most accurately characterized as creating or predicting information about a consumer, activities that could be read to be forms of “gathering” or “obtaining” information about a consumer. Such an interpretation of the statute is likely to be challenged given the imprecise fit of inference-creation into those terms. This uneasy conceptual fit would leave open the matter of whether inference creation falls under “collection” of personal information until courts rule on it, the CCPA is amended, or the enforcer of the California framework provides further clarifying regulations.

The aforementioned right to disclosure of information sold is unique, in that it grants consumers a glimpse into the inferences a business may have made about them. As previously stated, the consumer has the right to obtain certain details about the personal information that has been collected about them, in which case inferences would not be included. However, the statute also grants the consumer the right to know “the categories of personal information that the business sold about the consumer,”¹⁰⁴ which expands the scope of the information subject to disclosure to any personal information about the consumer that has been sold. Since the scope is not limited only to personal information that has been collected, the consumer may have the right to know the categories of information created or derived by the business, which would include inferences.

101. *Id.* § 1798.110.

102. *Id.* § 1798.115.

103. CAL. CIV. CODE § 1798.140(e).

104. *Id.* § 1798.115(a)(2).

The CCPA's right to opt-out allows a consumer "at any time, to direct a business that sells personal information about the consumer to third parties not to sell the consumer's personal information."¹⁰⁵ It therefore seems to wholly cover inferences by virtue of not being limited to personal information that has been collected. However, in its subdivision detailing compliance obligations for businesses, the CCPA specifies that when a consumer exercises their right to opt-out, a business shall "refrain from selling personal information collected by the business about the consumer."¹⁰⁶ Inferences are therefore included in the scope of the right of the consumer, but left out of the scope of the obligation on the business, making it ultimately unclear whether inferences are or are not covered by the CCPA's opt-out procedure.

The CCPA offers confusing treatment of inferences in its right to nondiscrimination as well. The statute's right to nondiscrimination protects consumers from receiving different treatment from a business—including denial of goods or services, differences in price or quality of goods or services, or the suggestion thereof—as a result of having exercised one of the previously mentioned rights.¹⁰⁷ The right also allows businesses to "offer financial incentives, including payments to consumers as compensation, for the collection of personal information, sale of personal information, or deletion of personal information."¹⁰⁸ In effect, the right allows businesses to buy from consumers their rights to have personal information about them be collected, sold, or deleted. It does not, however, explicitly allow businesses to buy the right to derive or create personal information about consumers. The lack of this specific allowance for businesses has two possible implications. At its least damaging, the CCPA once again fails to recognize that businesses come into possession of personal information about consumers by means other than pure collection, and thus businesses are not encouraged to compensate consumers for the creation of sensitive inferences about them that constitute personal information. At its most troublesome, the statute simply does not recognize the rights of consumers over information created or derived about them as a valuable right or as a right that exists at all. In either case, businesses are made to believe that their creation of sensitive inferences is of vastly less importance than their collection and sale of personal information.

It is not only the core rights of consumers that are limited to personal information collected by businesses. Several additional sections of the CCPA are similarly limited, including subsections dedicated to compliance obligations

105. *Id.* § 1798.120(a).

106. CAL. CIV. CODE § 1798.135(a)(4).

107. *Id.* § 1798.125(a).

108. *Id.* § 1798.125(b)(1).

for businesses, applicability of the Act, and exemptions, thereby making them inapplicable to inferences.¹⁰⁹ The CCPA's recurring references to information "collected from the consumer" in defining the rights it confers hints at the possibility that the drafters of the statute failed to consider that derived or inferred information can be among the most sensitive personal information about a consumer, and among the most valuable information for businesses to access.

The California Attorney General's guidance, instead of providing clarification on the uncertainty over collected information, also relies on the limited concept of information "collected from the consumer" by businesses to attempt to explain how businesses are to comply with the statute.¹¹⁰ For instance, the CCPA explains in detail businesses' obligations with respect to the consumers' right to access, which requires that businesses notify the consumer of their personal information collection practices.¹¹¹ The Attorney General's regulations call this the "notice at collection" requirement, which they define as "the notice given by a business to a consumer at or before the time a business collects personal information from the consumer."¹¹² The language in the regulations still limits the notice requirement to information collected, and it further proscribes behavior of businesses based only on collection, prohibiting them from collecting—but not creating—any types of information not included in the notice at collection.¹¹³ Businesses are also prohibited from collecting information from consumers—but not creating information about them—if the notice at collection is not given.¹¹⁴ These limitations in the Attorney General's regulations exclude inferences.

In some instances, the regulations might give consumers rights over inferences where the CCPA does not, like in their requirements that businesses' privacy policies "explain that a consumer has the right to request that the business disclose what personal information it collects, uses, discloses, and sells"¹¹⁵ or that privacy policies explain to the consumer their right to opt out of the sale of their personal information.¹¹⁶

109. *See, e.g., id.* § 1798.175 (on Applicability); *id.* § 1798.135 (on Compliance Obligations); *id.* § 1789.145 (on Exemptions).

110. *See generally* CAL. CODE REGS. TIT. 11, DIV. 1, CH. 20 [hereinafter "CCPA Regulations"].

111. CAL. CIV. CODE § 1798.100(b).

112. CCPA Regulations § 999.301(l).

113. *See Id.* § 999.305(a)(5).

114. *Id.* § 999.305(a)(6).

115. *Id.* § 999.308(c)(1)(a).

116. *Id.* § 999.308(c)(3)(a).

The CPRA amended the CCPA to significantly change and expand the rights that the California framework grants to consumers.¹¹⁷ Notable among the changes is the creation of a new category of data, “sensitive personal information,” which consists of any personal information that reveals any of a number of personal data points such as social security numbers, racial or ethnic origin, financial account information, and biometric information.¹¹⁸ Perhaps drawing from the GDPR’s treatment of its special categories of data, the CPRA stipulates that consumers should be able to control the use of their sensitive personal information because its unauthorized use or disclosure “creates a heightened risk of harm to the consumer.”¹¹⁹ The CPRA also carves out special provisions aimed at addressing inferences by stating that the collection or processing of sensitive personal information for inference-creation purposes obligates businesses to comply with additional requests from consumers to limit or disclose use.¹²⁰

The types of data that the drafters of the CPRA chose to designate as sensitive personal data are similar to the sort of information that inferences can reveal about consumers. However, the rights afforded to users over their sensitive personal information, including the aforementioned rights introduced by the CPRA, appear to also still be limited by the fact that they only apply to data “collected,” and not created or derived, from the consumer. The newly created California Privacy Protection Agency is tasked with issuing new regulations to clarify the California framework in advance of the CPRA’s effective date of January 1, 2023.¹²¹ As of early 2021, however, the text of the recently passed statute seems to leave the legal status of inferences as uncertain as it was in the CCPA and the California Attorney General’s regulations.

The fact that the CCPA includes and defines the concept of an inference and inference-making, as well as the fact that the CPRA acknowledges the significance of inference-creation, implies that the drafters of the California framework acknowledge the sensitive nature of inferences and the importance of protecting them by granting consumers rights over them. At some point in

117. Bret Cohen, Tim Tobin & Aaron Lariviere, *Understanding the new California Privacy Rights Act: How businesses can comply with the CPRA*, HOGAN LOVELLS ENGAGE (Nov. 25, 2020), <https://www.engage.hoganlovells.com/knowledgeservices/news/understanding-the-new-california-privacy-rights-act-how-businesses-can-comply-with-the-cpra?nav=FRbANEucS95NMLRN47z%2BeeOgEFCt8EGQ71hKXzqW2Ec%3D&key=BcJlhLtdCv6%2FJTDZxvL23TQa3JHL2AIGr93BnQjo2SkGJpG9xDX7S2thDpAQsCconWHAwe6cjTn0ksYg%2Fo%2FRPN3OjGmtaEjr&uid=iZAX%2FROFT6Q%3D>.

118. See CPRA Sec. 14, § 1798.140(ae).

119. CPRA Sec. 3(A)(2).

120. See CPRA Sec. 10, §§ 1798.121(a), (d).

121. Cohen, Tobin & Lariviere, *supra* note 117.

the drafting process, that intention was either pushed to the side or forgotten, since key portions of the California framework grant consumers rights using the CCPA's own defined concepts in ways that make the different documents contradict themselves. Ultimately, the legal status of inferences under the framework remains up in the air. The resulting uncertainty means that there is still plenty of room for the California framework to be amended to remedy this uncertain status, as well as a chance that certain contradictory portions of the documents will be litigated in order to address the uncertainties.

The legal status of inferences in California is contradictory and uncertain, leaving this sensitive data without adequate protection. As previously seen, it is not only the case that inferences are inadequately protected in California, which has come to be seen as a floor for privacy protections in the United States since the enactment of the CCPA. E.U. statutes, policy, and jurisprudence have also left significant questions unanswered regarding the legal status of sensitive personal information such as inferences. The patchy and uncertain legal and regulatory framework surrounding inferences poses a number of significant dangers for global and regional data protection regimes generally, and specifically for users and their rights to control their personal information.

IV. USERS DESERVE THE STRONGEST POSSIBLE RIGHTS WITH REGARD TO INFERENCE

Inferences are particular in that they allow data controllers and businesses to know information about users' personal lives without having to ask for such data. The way highly sophisticated inferences are created about a person depends on a phenomenon that is near invisible to the lay user (the collection of massive amounts and types of data) and takes advantage of a technology that sounds inconceivable, like something out of a science fiction film (deep neural networks, or computers that can be trained to think like humans). Because of these reasons, it is understandable how the drafters of recent data protection statutes and regulation could have neglected to attach due importance to inferences.

As previously mentioned, however, inferences can contain or reveal some of the most sensitive information about a person's life, like their pregnancy status, sexual orientation, or race. Were these sorts of sensitive inferences to end up in the hands of bad actors—or even organizations that are not supposed to have access to certain types of information—users could experience tangible, adverse consequences in their everyday lives.

Current and future data protection statutes and regulations should grant users the strongest possible rights over inferences made about them. Under

the GDPR, all inferences made with personal data as their starting point should be afforded the designation of personal data, with special consideration given to the fact that many inferences contain those special categories of personal information, the processing of which is prohibited save for a few exceptions.¹²² With regard to the California framework, all consumer rights whose application is limited to information “collected from” the user or “maintained by” the business should be expanded to cover information “inferred” or “created” based on the user’s personal data, as well. Further, all language in the framework’s statutes should be clarified in such a way that inferences are clearly covered.

These changes should be implemented for two reasons, which I elaborate upon in the following Sections. Firstly, both the GDPR and the CCPA state they were enacted to protect the very kind of sensitive information that makes up inferences. In fact, inferences are among the type of information that first signaled the need for the protection of personal data in general in the early days of the information technology boom. Secondly, the sophisticated nature of modern inferences—and the information that data-reliant organizations gain access to because of inferences—mean that invasions of privacy involving inferences can, unprecedentedly, result in immediate or near-immediate threats to the physical integrity, autonomy, and livelihood of data subjects.

A. DATA PROTECTION STATUTES EXIST PRECISELY TO PROTECT THE SORT OF INFORMATION THAT INFERENCES CONSTITUTE

It could be argued that inferences are such a hyper-specific type of data, or so pivotal to the modern internet economy, that European and California lawmakers can be forgiven for instituting frameworks that treat inferences so haphazardly. However, to dive into the justifications for the existence of these data protection regimes—and even the history of data protection as a field—is to realize that inferences and other similar types of data are the animating issue for the enactment of these laws and regulations.

In its Recital 6, the GPDR acknowledges that “[r]apid technological developments . . . have brought new challenges for the protection of personal data,” and that the “scale of the collection and sharing of personal data has increased significantly.”¹²³ It goes on to acknowledge that modern technology “allows both private companies and public authorities to make use of personal data on an unprecedented scale in order to pursue their activities,” and that while data should flow freely within the European Union, the GDPR must

122. See *supra* text accompanying notes 61–63.

123. GDPR, *supra* note 16, Recital 6.

ensure a high level of protection of personal data.¹²⁴ The unprecedented use and scale of collection of personal information mentioned are undoubtedly a reference to, among other modern data practices, the creation and use of inferences.

The GDPR's preamble also notes that the 1995 Data Protection Directive, while necessary, had several weaknesses that resulted in differing levels of protection of personal data.¹²⁵ It also notes that "[e]ffective protection of personal data throughout the [European] Union requires the strengthening . . . of the rights of data subjects and the obligations of those who process and determine the processing of personal data."¹²⁶ In Recital 13, the GDPR lays the foundation for the new data protection regime it will establish by stating that, in light of the aforementioned points, the GDPR is necessary to grant all natural persons across the European Union equal levels of legally enforceable rights and ensure consistent monitoring of the processing of personal data.¹²⁷

The creation of highly sophisticated inferences thanks to massive, widespread data collection is exactly the kind of phenomenon the GDPR concerns itself with and mentions as its reason for being in its Recital 6.¹²⁸ Such a clear indication of the type of data that necessitated stronger data rights for users implies that inferences, due to their nature, should be afforded the strongest protections under the GDPR.

The California Legislature's language in the CCPA's preamble is striking in that it seems to heavily allude to inferences. In the CCPA bill's preambular § 2, the California Legislature acknowledges that the right to privacy is among the inalienable rights granted by the state's constitution.¹²⁹ It goes on to find and declare, "As the role of technology and data in the every daily lives of consumers increases, there is an increase in the amount of personal information shared by consumers with businesses," and that "California law has not kept pace with these developments" and their implications for personal privacy."¹³⁰ Remarkably, § 2(e) acknowledges the unprecedented kinds of personal information that businesses may now have access to by stating: "[Businesses] may know where a consumer lives and how many children a consumer has, how fast a consumer drives, a consumer's personality, sleep habits, biometric and health information, financial information, precise

124. *Id.*

125. *See id.* Recital 9.

126. *Id.* Recital 11.

127. *See id.* Recital 13.

128. *See id.* Recital 6.

129. AB-375, 2017-2018 Assemb., Reg. Sess. (Cal. 2018) § 2(a).

130. *Id.* § 2(d).

geolocation information, and social networks, to name a few categories,” all information that can be derived or inferred from data that businesses routinely collect.¹³¹ In the following subsection, the CCPA speaks to the potential consequences of the mishandling of sensitive personal information: “The unauthorized disclosure of personal information and the loss of privacy can have devastating effects for individuals, ranging from financial fraud, identity theft, and unnecessary costs to personal time and finances, to destruction of property, harassment, reputational damage, emotional stress, and even potential physical harm.”¹³² The preamble closes by stating that, in light of the aforementioned facts, it is the intent of the Legislature to “further Californians’ right to privacy” by enacting the CCPA.¹³³

Mentions within the CCPA of the increased amounts of personal information that consumers share as a result of modern technology are a reference to the widespread collection of personal data that animates not only the modern digital economy, but also all modern data protection. That clause, paired with the specific mention of several potential data points that are often inferred rather than requested (such as a consumer’s personality, sleep habits, and social connections) points toward the fact that inferences were one of, if not *the* primary motivation behind efforts to beef up California consumers’ rights over their personal data. As previously mentioned, the CCPA ends up contradicting itself in ways that result in a lackluster protection of inferences, which might have been the result of a rushed drafting process and legislative trajectory for the CCPA.¹³⁴ However, the fact that California’s opening salvo in addressing data protection justifies this effort by name-checking inferences bolsters the idea that the CCPA should be amended to remedy the patchy protection of this sensitive type of data.

But the significance of data inferences was recognized much earlier than these recent data protection efforts in Europe and California. There is precedent for the stronger protection of inferences within the early history of data privacy rights and data protection, when experts and advocates were first noticing the implications of information technology on personal privacy. Technological developments that facilitated the processing of data—and

131. *Id.* § 2(e).

132. *Id.* § 2(f).

133. *Id.* § 2(i).

134. *See generally* Issie Lapowsky, *California Unanimously Passes Historic Privacy Bill*, WIRED (June 28, 2018), <https://www.wired.com/story/california-unanimously-passes-historic-privacy-bill/> (“The so-called California Consumer Privacy Act of 2018 (AB 375) was introduced late last week by state assemblymember Ed Chau and state senator Robert Hertzberg, in a rush to defeat a stricter privacy-focused ballot initiative that had garnered more than 600,000 signatures from Californians.”).

ultimately facilitated the creation of inferences—are among the factors that first animated the regulation of personal data in the United States. Paul Schwartz and Daniel Solove explain that numerous American privacy laws turn on the concept of personally identifiable information (PII), for which there is no uniform definition.¹³⁵ PII can be said to equate to the concept of personal data, and Schwartz and Solove indicate that it “first became an issue in the 1960s with the rise of the computer,” which not only allowed entities both public and private to collect more information, but also to process that information in unprecedented ways.¹³⁶ Computers removed the limiting factors for how data could be stored and retrieved, and “permitted information to be searched and organized by *multiple attributes* rather than simply through a single index, as, for example, a person’s first and last name.”¹³⁷ This technological development “changed the way information could be linked to an individual,”¹³⁸ and facilitated the practice of inference-making, albeit a rudimentary version of the practice conducted by humans instead of artificial intelligence. The new data handling capabilities subsequently “required Congress to confront the issue of the kinds of information that should matter for information privacy law.”¹³⁹ Following this unprecedented expansion of the level of access to personal information, Congress—albeit with much delay—responded by starting to enact privacy legislation that strongly protected PII by making the presence of PII the trigger for privacy protections. According to Schwartz and Solove, Congress enacted the Cable Communications Policy Act of 1984 (“the Cable Act”) and included this new approach to PII in response to how technological advances made it theoretically possible for consumers to send information to broadcasters and television operators via their televisions, utilizing a technology called “videotex.”¹⁴⁰ Policymakers were concerned that “by collecting these data, the cable operator would be able to construct detailed profiles about viewing choices” and derive information about viewers’ interests from them.¹⁴¹

The ease with which PII could be collected and processed triggered a rethinking of the legislative and policymaking approaches to privacy protection precisely because of concerns that technology made PII more vulnerable to exploitation and the creation of new information based on it, in the form of

135. Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 N.Y.U.L. REV. 1814, 1816 (2011).

136. *Id.* at 1820.

137. *Id.*

138. *Id.*

139. *Id.* at 1821.

140. *Id.* at 1826.

141. *Id.*

inferences. Lawmakers have once, in the past, acknowledged the threat to privacy posed by technological advances that made it easy to collect and derive more information about individual persons. It follows that it is appropriate and necessary for lawmakers to implement changes to new privacy legislation that are responsive to AI-facilitated inference-making, which, like computers were in the 1960s, is the novel data-handling practice that threatens personal privacy in contemporary times.

Granting users more protection over inferences is a policymaking approach that is supported not only by present justifications for data protection regimes in the very text of the GDPR and CCPA statutes, but also by the history of data protection policymaking itself. These sources show that when technological advancements bring about uncertainty as to the protection of sensitive personal information, legislative and policy action aimed at adapting to the changes is logical and appropriate. The stronger protection of inferences did not come slowly in response to then-present harms. In the 1980s, potential and anticipated technologies like videotex, as well as their potential harms, were enough to justify strong protection of PII in the Cable Act. Contemporary policy- and lawmakers should mimic their predecessors and consider the potential harms that sophisticated, machine-learning-powered inferences could represent in the future when finding justifications for the strong protection of inferences in contemporary data protection regimes.

B. INVASIONS OF PRIVACY INVOLVING INFERENCES MAY INVOLVE HARMS TO INTERESTS OTHER THAN THE REPUTATIONAL OR DIGNITARY

When justifications for privacy regulation are offered, they often focus on the individual interests that privacy rights seek to protect. According to Paul Schwartz and Karl-Nikolaus Peifer, William Prosser conceived of the four modern privacy torts (intrusion upon seclusion, public disclosure of private facts, false light, and appropriation) as intended to protect the rights-holder from offensive behavior, attacks against their reputation, mental distress, and exploitation of their image for financial gain.¹⁴² The forefathers of privacy law, Samuel Warren and Louis Brandeis, thought of the right to privacy as protecting a person's "inviolable personality," which Edward Bloustein

142. 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, 1941–42 (2010).

expanded upon to mean “the individual’s independence, dignity, and integrity.”¹⁴³

The aforementioned justifications within the text of the GDPR and CCPA, by their mention of technological advancements and the unprecedented types of information businesses can access, hint at efforts to protect the same dignitary and reputational interests that Prosser, Warren, and Brandeis enumerated, but with an eye towards accounting for other interests that might be unforeseeably invaded. By virtue of the way they are created, the types of sensitive information they can include about users, and the way they are used by data-reliant organizations, inferences can result in invasions of privacy that violate interests outside of the ones suggested by most privacy scholarship to date. In fact, invasions of privacy involving inferences can result in immediate or near-immediate threats to the physical integrity, autonomy, civil rights, and even financial prospects of members of certain vulnerable populations.

The incident that introduces this Note is a prime example. The young woman whose pregnancy was involuntarily disclosed to her family by Target suffered an invasion of privacy as a result of an inference. Target’s disclosure by implication of the young girl’s pregnancy status, while not illegal, goes against established medical consensus regarding a minor’s rights to not involve her parents in matters relating to her pregnancy.¹⁴⁴ The fact that she was robbed of the decision of when and how to inform her family of her pregnancy also represents the compromise of her autonomy, and depending on her family dynamic, her physical safety might also have been imperiled.

Social networking platforms like Facebook, which collect massive amounts of user data, are able to make inferences about users’ sexual orientation, putting LGBTQ people’s physical and emotional safety at risk.¹⁴⁵ Until recently, Facebook made it possible for advertisers to target users based on

143. *Id.* at 1943–45; Edward J. Bloustein, *Privacy as an Aspect of Human Dignity: An Answer to Dean Prosser*, 39 N.Y.U.L. REV. 962, 963 (1964).

144. See AAP Committee on Adolescence, *The Adolescent’s Right to Confidential Care When Considering Abortion*, 139(2) PEDIATRICS e20163861, 2 (2017) (“The American Medical Association, the Society for Adolescent Health and Medicine, the American Public Health Association, the American College of Obstetricians and Gynecologists, the AAP, and other health professional organizations have reached a consensus that a minor should not be compelled or required to involve her parents in her decision to obtain an abortion, although she should be encouraged to discuss the pregnancy with her parents and/or other responsible adults.”).

145. See Carter Jernigan & Behram F.T. Mistree, *Gaydar: Facebook Friendships Expose Sexual Orientation*, 14(10) FIRST MONDAY, <https://www.firstmonday.org/article/view/2611/2302> (last visited Jan. 14, 2021).

their sexual orientation.¹⁴⁶ In 2012, a British teen was kicked out of his home after his parents found out he was gay via Facebook.¹⁴⁷ The teen had not submitted any information to Facebook about his sexual orientation, nor had he joined any gay groups, yet claimed that Facebook had begun displaying gay-interest ads on his profiles, which were seen by his parents when he accidentally left his computer unattended.¹⁴⁸

According to Human Rights Watch, there are at least seventy countries around the world where same-sex relations are criminalized, and nine where certain “forms of gender expression that target transgender and gender nonconforming people” are also crimes.¹⁴⁹ In these countries, LGBTQ people face punishments ranging from prison, to corporal punishment, to the death penalty.¹⁵⁰ In a jurisdiction where certain sexual or gender minorities are criminalized, a misplaced LGBTQ-related targeted ad or recommendation on a platform that has inferred a user’s sexual orientation could present harms ranging in severity from that experienced by the British teenager to threats to the physical integrity and lives of LGBTQ people. According to Human Rights Watch, during the anti-gay purge perpetrated by local authorities in Chechnya, Russia, in 2017, one of the ways in which police would identify gay men to capture and torture was to search captured victims’ cell phones, “looking for contacts of other men who might be gay.”¹⁵¹ If presence in an LGBTQ person’s contact list is enough of a proxy for authorities in queerphobic regimes, inference-based relevant advertisements and recommendations in an individual’s social media profile—like Instagram’s “Suggested For You,” which shows suggestions for “similar profiles” to a particular user’s¹⁵²—could be as well.

146. Alex Kantrowitz, *Facebook Has Blocked Ad Targeting By Sexual Orientation*, BUZZFEED NEWS (Mar. 21, 2018), <https://www.buzzfeednews.com/article/alexkantrowitz/facebook-has-blocked-ad-targeting-by-sexual-orientation>.

147. See Kenneth C. Werbin, Mark Lipton & Matthew J. Bowman, *The Contextual Integrity of the Closet: Privacy, Data Mining and Outing Facebook’s Algorithmic Logics*, 2(1) QUEER STUD. IN MEDIA & POPULAR CULTURE 37 (2017).

148. *Id.*

149. #Outlawed: “*The Love That Dare Not Speak Its Name*,” HUMAN RIGHTS WATCH, http://internap.hrw.org/features/features/lgbt_laws/ (last visited Jan. 14, 2021).

150. *Id.*

151. “*They Have Long Arms and They Can Find Me*,” *Anti-Gay Purge by Local Authorities in Russia’s Chechen Republic*, HUMAN RIGHTS WATCH (May 26, 2017), <https://www.hrw.org/report/2017/05/26/they-have-long-arms-and-they-can-find-me/anti-gay-purge-local-authorities-russias>.

152. See *People Are Getting Suggestions to Follow Other People After They Follow Me on Instagram. How Do I Turn This Off?*, INSTAGRAM, <https://help.instagram.com/530450580417848> (last visited Apr. 18, 2021).

At a minimum, the potential for such targeted content on the internet adds a persistent burden to the online activities of LGBTQ users, in that they must either always attempt to monitor what information certain platforms know about them, or they must keep from accessing LGBTQ-related content on the internet entirely for fear that platforms will infer their sexual orientation. LGBTQ people have historically depended on the relative anonymity that the internet provides in order to access important educational content about their sexual orientation or gender identities without fear of adverse consequences. The existence of inferences over which the user has little control can effectively prevent users from extracting value from the wealth of information online.

In some instances, the use of inferences by certain types of organization can be illegal. In her landmark study *Discrimination in Online Ad Delivery*, Latanya Sweeney exhibits as part of her problem statement how Google Images has learned to associate, perhaps from user data it has collected or acquired, certain proper names with certain races.¹⁵³ Image searches for “Latanya” and “Latisha” return results for Black women, while searches for “Kristen” and “Jill” return images of white women.¹⁵⁴ Google Images’s inference-making is not illegal, but a problem arises when platforms that collect large amounts of user information, armed with knowledge such as Google Images’s, can make inferences about users’ race and other traits and include them in profiles about the users. These profiles may be shared with data brokers who will themselves share them with business who use these profiles to target advertisements to the user. An investigation by ProPublica showed that Facebook allowed advertisers to exclude users based on their race or gender, even in cases of ads where exclusion from viewing could be illegal, like housing ads.¹⁵⁵ A credit card company hoping to advertise a new product to a specific audience could find, on Facebook, the ability to exclude users based on race, a practice that in the United States is outlawed by the Equal Credit Opportunity Act (ECOA).¹⁵⁶

153. Latanya Sweeney, *Discrimination in Online Ad Delivery*, DATA PRIVACY LAB (2013), <http://dataprivacylab.org/projects/onlineads/1071-1.pdf>.

154. *Id.*

155. Terry Parris, Jr. & Julia Angwin, *Facebook Lets Advertisers Exclude Users by Race*, PROPUBLICA (Oct. 28, 2016), <https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race>.

156. 15 U.S.C. § 1691(a)(1); *see also* CONSUMER FINANCIAL PROTECTION BUREAU, EQUAL CREDIT OPPORTUNITY ACT (ECOA) EXAMINATION PROCEDURES (2015), https://files.consumerfinance.gov/f/documents/201510_cfpb_ecoa-narrative-and-procedures.pdf (“... a creditor may not advertise its credit services and practices in ways that would tend to encourage some types of borrowers and discourage others on a prohibited basis. In addition, a creditor may not use prescreening tactics likely to discourage potential applicants on a prohibited basis.”).

Users, however, have no way of knowing what ads are being hidden from them due to certain traits, and as such are unaware of these violations to their civil rights. Credit-lending is not the only context in which such exclusionary targeting is possible; the practice is possible in the housing and employment contexts as well. The threat to a users' financial status and prospects that these practices represent amounts to digital redlining, a term that harkens back to the sort of practices legislation like ECOA were passed to prevent. Christopher Gilliard defined digital redlining as "the creation and maintenance of technology practices that further entrench discriminatory practices against already marginalized groups."¹⁵⁷ As more aspects of everyday life, especially the management of personal finances, are handled online, digital redlining can perpetuate and in many cases worsen economic inequality, especially as it affects people of color and other historically disadvantaged communities.

These threats made possible by the combination of inference creation and user targeting represent types of harms that not only implicate new types of interests but could also disproportionately affect marginalized groups and some of society's most vulnerable populations. New types of data that can result in new types of harm justify unprecedented levels of protection for such data, bolstering the argument for strong user control over inferences created about them. If users are to be granted stronger control over their inferences, both the European and California data privacy regimes ought to be amended to fully incorporate inferences into the definition of sensitive personal data so as to trigger the strongest protections afforded by both statutes, as previously mentioned.

V. PROPOSALS FOR THE STRONG PROTECTION OF INFERENCES MAY STILL LEAVE SOME ISSUES UNRESOLVED

In this Note, I endeavored to lay a foundation for how to think about the issue of inferences while keeping in mind the sheer importance, power, and value of this type of data in current times. The muddled state of the data protection of inferences, however, still raises a host of unresolved questions that could adversely impact any attempt at stronger protection of inferences, including the approach I propose in this Note.

Firstly, data-reliant businesses could lobby against the stronger protection of inferences by arguing that the inferences are created by using technologies

157. *Banking on Your Data: the Role of Big Data in Financial Services: Hearing before the Comm. on Financial Services Task Force on Financial Technology*, 116th Cong. 4 (2019) (statement of Dr. Christopher Gilliard, PhD, Professor of English, Macomb Community College and Digital Pedagogy Lab Advisor).

the controllers themselves have developed to serve their own needs, and which these businesses consider to be trade secrets. Their argument would be, essentially, that inferences are created by the sweat of the businesses' brows, and, as such, any rights granted to the user over inferences—especially the right to access—risk disclosure of the trade secret and a threat to the business's ability to profit off of it. Sonia Katyal has written on the perils of such reasoning, contending that trade secrecy in the context of consumer technology is closely intertwined with civil rights, and arguing that transparency should be incorporated into trade secrecy law in recognition of the modern threats algorithms can represent for individual's civil rights.¹⁵⁸ Katyal puts forth as a remedy to this issue the federal whistleblowing protections in the Defend Trade Secret Act (DTSA) of 2016, by which whistleblowers who disclose source code they suspect leads to biased decision-making are protected, and so is the trade secret while the claims are investigated.¹⁵⁹ However, depending on whistleblowers to disclose issues related to the handling of sensitive personal data is an inefficient and insufficient remedy given the scale at which data collection and inference creation are occurring. Katyal remarks as such in the context of trade secrets by making reference to such a solution's administrative costs as well as the difficulty of initial detection of malfeasance.¹⁶⁰

Secondly, data-reliant businesses might argue that, since inferences are information about a user that the controller creates (instead of collecting it), preventing the creation of inferences or limiting their use restricts businesses' speech and violates their First Amendment rights. Legal challenges based on this argument could hinge on whether it can be determined that protecting the privacy of data subjects by granting them rights over inferences is a substantial state interest. In *Sorrell v. IMS Health*, the Supreme Court held that restricting the use of doctors' personal information to target marketing at them was a content-based restriction, and, since it did not advance a substantial state interest, it violated the speaker's First Amendment rights.¹⁶¹ Additionally, in *Central Hudson Gas & Electric v. Public Service Commission*, the Court instituted a test for whether restrictions on commercial speech violate the First Amendment.¹⁶² The *Central Hudson* test states that if (1) the speech at issue relates to lawful activity and is not misleading, (2) the government interest is

158. Sonia Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 120 (2019).

159. *Id.* at 130.

160. *Id.* at 140.

161. *Sorrell v. IMS Health Inc.*, 131 S. Ct. 2653 (2011).

162. *Central Hudson Gas & Electric Corp. v. Pub. Serv. Comm'n*, 447 U.S. 557, 557 (1980).

substantial, (3) regulation of the speech directly advances the government interest, and (4) the regulation is “no more extensive than necessary,” the restriction is lawful.¹⁶³ In the case of inferences, granting users the right to prevent processing of their sensitive personal information for ad-targeting purposes might constitute a content-based restriction on businesses’ commercial speech. Provided that the inferences relate to lawful activities and do not contain misleading or inaccurate facts about users, the constitutionality of any stronger protections will hinge on courts. Judges will have to decide, hopefully with access to the academic literature on inferences, whether users having control over inferences and their use is a substantial government interest, and whether and to what extent restrictions on data-reliant businesses’ use of inferences advance such an interest.

Lastly, data-reliant organizations might object to the classification of inferences as personal data, which would trigger stronger protections under both the GDPR¹⁶⁴ and the California data privacy frameworks.¹⁶⁵ Since the inferences that these organizations create are often predictions about the likelihood of a fact being true at present,¹⁶⁶ they can argue that any one set of inferences or profile they create are not about a specific user, but rather a hypothetical individual whose data points and created inferences are very similar to the user’s, and therefore the user has no rights over the information. This argument could be weakened by two factors. On the one hand, in making this argument these organizations might create more issues and legal exposure for themselves, given that the argument implies decisions of all levels of importance are being made about specific users on the basis of potentially inaccurate data. Additionally, even if these data-reliant organizations are not a hundred percent certain of the inferences they draw unless they seek confirmation from the user, the inference is often made on the basis of data collected from or about a user. Even if any identifiers are removed from both collected data and inferences—in a process called de-identification or anonymization—contemporary data sets can include inferences that are so numerous, unique, detailed, or sophisticated that they run a high risk of being re-identified.¹⁶⁷ The risk of re-identification is significant, and it increases as

163. *Id.*

164. *See* GDPR, *supra* note 16.

165. *See* CAL. CIV. CODE § 1798.100.

166. *See supra* discussion in Part II.

167. *See generally* Luc Rocher, Julien M. Hendrickx & Yves-Alexandre de Montjoye, *Estimating the Success of Reidentifications in Incomplete Datasets Using Generative Models*, 10:3069 NATURE COMMUNICATIONS 1 (2019) (“...we find that 99.98% of Americans would be correctly re-identified in any dataset using 15 demographic attributes. Our results suggest that even heavily sampled anonymized datasets are unlikely to satisfy the modern standards for

these organizations gain access to more and different kinds of data points.¹⁶⁸ In the age of Big Data, different types of organizations collect or derive countless different or overlapping types of data points depending on the information that might be of importance to them, making users vulnerable to re-identification attacks, and as such underscoring the need for the protection of inferences.¹⁶⁹

These issues are only three out of numerous possible ones, and they are likely exponentially more complex than this space allows. My hope is that each will garner enough attention from data protection scholars to merit further in-depth study. That is also my wish for the issue of inferences at large.

VI. CONCLUSION

The current treatment of inferences under European and Californian data protection statutes is confusing. This uncertainty harms all users, but especially historically vulnerable populations.

Everything from the collection of the data that drives inferences, to their creation, to their use is shrouded in a veil of secrecy and nebulousness. Nonetheless, inferences are the same type of data that the field has historically concerned itself with the most, only collected, organized, and handled in unprecedented ways that can bring about new kinds of harms. As daunting as the task of addressing this issue seems, legislators, policymakers, and regulators must strive to demystify and address the issue if they are strongly committed to the protection of the most sensitive types of personal data and, by extension, the most vulnerable people in our increasingly digital societies and economies.

anonymization set forth by GDPR and seriously challenge the technical and legal adequacy of the de-identification release-and-forget model.”).

168. *See id.* at 2 (“With population uniqueness increasing fast with the number of attributes available, our results show that the likelihood of a re-identification to be correct, even in a heavily sample dataset, can be accurately estimated, and is often high.”).

169. Khaled El Emam, Elizabeth Jonker, Luk Arbuttle & Bradley Malin, *A Systematic Review of Re-Identification Attacks on Health Data*, 6(12) PLOS ONE 1, 2 (2011).

DEFINING THE PRIVACIES OF LIFE: LOWER COURT TRENDS IN THE WAKE OF *CARPENTER*

Tiffany Chen[†]

I. INTRODUCTION

When the Supreme Court announced its *Carpenter v. United States* decision in June 2018,¹ many scholars and journalists lauded the opinion as a groundbreaking victory for privacy. The decision “chose to bring the Fourth Amendment into the digital future and protect against growing technologically enhanced police surveillance powers,” wrote Professor Andrew Ferguson.² *Slate* writer Mark Joseph Stern perceived *Carpenter* as a “far-reaching decision,” even an “earthquake in Fourth Amendment law,” one that could “dramatically . . . expand[] the scope of the Fourth Amendment” and “provide[] vital new protections to the vast majority of Americans.”³ Numerous other commentators similarly predicted that *Carpenter* would have wide-ranging consequences for digital privacy; American Civil Liberties Union staff attorney Nathan Freed Wessler, for example, remarked that this “groundbreaking update to privacy rights” “open[ed] the door to the protection of many other kinds of data generated by popular technologies.”⁴

The last year,⁵ however, has seen little of the “widespread implications” so anticipated by observers.⁶ Thus far, lower courts have repeatedly

DOI: <https://doi.org/10.15779/Z38RF5KG63>

© 2020 Tiffany Chen.

† J.D., 2021, University of California, Berkeley, School of Law.

1. *Carpenter v. United States*, 138 S. Ct. 2206 (2018).

2. Andrew Ferguson, *Future-Proofing the Fourth Amendment*, HARV. L. REV. BLOG (June 25, 2018), <https://blog.harvardlawreview.org/future-proofing-the-fourth-amendment>.

3. Mark Joseph Stern, *A Historic Victory for Privacy*, SLATE (June 22, 2018, 11:41 AM), <https://slate.com/news-and-politics/2018/06/carpenter-v-united-states-supreme-court-rules-fourth-amendment-protects-cell-phone-location-records-in-an-opinion-by-chief-justice-john-roberts.html>.

4. Nathan Freed Wessler, *The Supreme Court's Groundbreaking Privacy Victory for the Digital Age*, FREE FUTURE (June 22, 2018, 2:30 PM), <https://www.aclu.org/blog/privacy-technology/location-tracking/supreme-courts-groundbreaking-privacy-victory-digital-age>.

5. This Note was written in 2020, so “the last year” refers to the year 2019.

6. Megan L. Brown, Matthew J. Gardner, Kathleen E. Scott & Vesna K. Harasic-Yaksic, *Carpenter v. United States: The Supreme Court's Recent Decision Will Have Widespread Implications for the Collection of Digital Information by Law Enforcement*, WILEY REIN LLP NEWS & INSIGHTS (June 25, 2018), https://www.wiley.law/alert-Carpenter_v_United_States_The_Supreme_Courts_Recent_Decision_Will_Have_Widespread_Implications

emphasized *Carpenter*'s own admission that its "decision [wa]s a narrow one," and generally declined to extend the holding to data outside of the cell site location information (CSLI) discussed in *Carpenter*.

These lower courts were correct not to broaden *Carpenter* in the last year, because these cases' fact patterns have not yet presented digital technologies at their most invasive. However, as law enforcement's surveillance techniques become increasingly sophisticated and intrusive, courts should heed *Carpenter*'s warnings and find Fourth Amendment violations whenever a "seismic shift[] in digital technology" invades the "privacies of life."⁷

This Note will analyze the last year's post-*Carpenter* decisions and highlight emerging surveillance technologies that may lead to Fourth Amendment violations in the near future. Parts II and III will provide background for *Carpenter*, with Part II discussing the Fourth Amendment as applied to the digital age, and Part III describing the legal and factual lead-up to the decision. Part IV will summarize the *Carpenter* ruling itself. Part V will then analyze lower court decisions announced in the year since, covering cases on police use of telephone pole cameras, internet protocol (IP) addresses, Global Positioning System (GPS) devices, and home Internet of Things (IoT) devices. The penultimate Part VI will introduce a new law enforcement practice on the horizon—big data database tracking—and argue that courts should hold that this technique violates the Fourth Amendment. Finally, Part VII will conclude with some observations on other actors in the digital privacy space who may contribute to *Carpenter*'s broadening in the near future.

II. THE FOURTH AMENDMENT

The Fourth Amendment of the U.S. Constitution grants citizens "the right . . . to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures."⁸ The Amendment further states that this right "shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized."⁹ As the *Carpenter* Court noted, the Amendment was adopted for two key purposes: first, to "secure 'the privacies of life' against 'arbitrary power'"; and second, "to place obstacles in the way of a too permeating police surveillance."¹⁰

7. *Carpenter*, 138 S. Ct. at 2214.

8. U.S. CONST. amend. IV.

9. *Id.*

10. *Carpenter*, 138 S. Ct. at 2214 (quoting *Boyd v. United States*, 116 U.S. 616, 630 (1886) and *United States v. Di Re*, 332 U.S. 581, 595 (1948)).

Courts' Fourth Amendment analyses generally rest upon the *Katz* standard, which asks “whether a person invoking its protection can claim . . . a ‘reasonable’ . . . ‘expectation of privacy’ that has been invaded by government action.”¹¹ In order to constitute a Fourth Amendment violation under this test, an act must satisfy two elements. First, an aggrieved individual must demonstrate that they had a subjective expectation of privacy that was infringed by the act; second, this subjective expectation must be “one that society is prepared to recognize as ‘reasonable.’”¹²

Of course, as Justice Scalia was quick to note in *Kyllo v. United States*, “in the case of the search of the interior of homes”—an activity that falls squarely within the “persons, houses, papers, and effects” language of the Fourth Amendment—“there is a ready criterion, with roots deep in the common law, of the minimal expectation of privacy that . . . is acknowledged to be *reasonable*.”¹³ Thus, if the government action clearly constitutes an intrusion into citizens' homes and does not use technology “in general public use,” then the conduct will, per *Kyllo*, be considered a search, and will not require a full *Katz* analysis.¹⁴

In the last couple of centuries, technological improvements have, as Professors Susan Freiwald and Stephen Smith observed, “inevitably present[ed] new tools for the criminally minded” and created novel questions for Fourth Amendment applicability.¹⁵ As a result, the legislature and judiciary have repeatedly, to borrow Professor Orin Kerr's term, pursued “equilibrium-adjustment”¹⁶—i.e., readapted the law to address these new technologies, from telegraph messaging, to pole cameras, to GPS devices. The advent of location tracking devices in the late twentieth century—including CSLI collection, the basis of the Supreme Court's landmark *Carpenter* ruling—constituted one such development.

III. THE PATH TO *CARPENTER*

In the 1970s, a pair of Supreme Court cases—*United States v. Miller* and *Smith v. Maryland*—articulated the concept at the heart of *Carpenter*: the third

11. *United States v. Knotts*, 460 U.S. 276, 280 (1983) (quoting *Smith v. Maryland*, 442 U.S. 735, 740 (1979)).

12. *Smith v. Maryland*, 442 U.S. 735, 740 (1979) (quoting *Katz v. United States*, 389 U.S. 347, 361 (1967)).

13. *Kyllo v. United States*, 533 U.S. 27, 34 (2001).

14. *Id.* at 34, 40.

15. Susan Freiwald & Stephen Wm. Smith, *The Carpenter Chronicle: A Near-Perfect Surveillance*, 132 HARV. L. REV. 205, 205 (2018).

16. Orin S. Kerr, *An Equilibrium-Adjustment Theory of the Fourth Amendment*, 125 HARV. L. REV. 476, 476 (2011).

party doctrine.¹⁷ *Miller* concerned a criminal defendant's bank records; his banks had supplied law enforcement with these materials upon receiving grand jury subpoenas.¹⁸ In light of the fact that "all of the documents obtained . . . contain[ed] only information voluntarily conveyed to the banks," the Court reasoned that the defendant had no legitimate expectation of privacy in his bank records.¹⁹ To support this conclusion, the Court articulated a concise definition of the third party doctrine: "[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."²⁰ Given this doctrine, then, the Court held that the government's actions did not constitute a search under the Fourth Amendment.²¹

The Court upheld this doctrine a few years later in *Smith*.²² Here, the third party in question was a telephone company: law enforcement officials installed a pen register at the company's central offices without a warrant to track the numbers that a criminal suspect used to call a robbery victim.²³ Citing *Miller*, the Court again applied the third party doctrine and found that the police had not conducted a Fourth Amendment search.²⁴ "When [the defendant] used his phone," the Court observed, "[he] voluntarily conveyed numerical information to the telephone company and . . . assumed the risk that the company would reveal to police the numbers he dialed."²⁵

Given this doctrine, one can understand why courts and law enforcement officials at first believed that warrantless retroactive CSLI collection did not violate the Fourth Amendment. After all, cell phone users did turn their location information over to third parties, i.e., their wireless carriers.²⁶ Thus, starting in the late 1990s, law enforcement agencies began compelling service providers to supply retroactive CSLI based on showings required by 18 U.S.C. § 2703(d) (a provision of the Stored Communications Act which, by

17. See *United States v. Miller*, 425 U.S. 435, 443 (1976); *Smith*, 442 U.S. at 744.

18. *Miller*, 425 U.S. 435, 437–38 (1976).

19. *Id.* at 442.

20. *Id.* at 443.

21. *Id.* at 446.

22. See *Smith v. Maryland*, 442 U.S. 735, 744 (1979).

23. *Id.* at 737–38.

24. *Id.* at 744.

25. *Id.*

26. See *id.*; *United States v. Miller*, 425 U.S. 435, 443 (1976).

this time, mandated a “specific and articulable facts” threshold far less stringent than the probable cause standard for warrants).²⁷

By the late 2000s, however, several federal magistrate judges had begun to reject these orders; after all, neither Congress nor the Supreme Court had weighed in on the new CSLI surveillance regime.²⁸ On appeal, circuit courts did not split and instead uniformly found no reasonable expectation of privacy in CSLI records in light of the third party doctrine.²⁹

These decisions carried on for several years, until, in 2017, the Supreme Court disregarded the lack of circuit split and granted *certiorari* to review a Sixth Circuit decision on the issue: *Carpenter v. United States*.³⁰

IV. THE *CARPENTER* DECISION

In *Carpenter*, law enforcement obtained over one hundred days of defendant Carpenter’s CSLI from his wireless carriers under the Stored Communications Act.³¹ In total, the police gathered 12,898 location points recording Carpenter’s movements, averaging 101 data points per day.³² According to the government, these CSLI records “clinched the case”; at trial, an FBI agent produced maps showing Carpenter’s cell phone close to four of his charged robberies.³³ Carpenter was eventually convicted of all but one of his firearm counts.³⁴

On appeal, the Sixth Circuit stayed true to the circuit courts’ previous decisions and affirmed Carpenter’s sentence under the third party doctrine.³⁵ After all, the court explained, “any cellphone user who has seen her phone’s signal strength fluctuate must know that, when she places or receives a call, her phone ‘exposes’ its location to the nearest cell tower.”³⁶

The Supreme Court granted *certiorari* for the case and disagreed with the Sixth Circuit. “Technology has enhanced the Government’s capacity to encroach upon areas normally guarded from inquisitive eyes,” the Court noted, and judges must continue to adjust Fourth Amendment readings so as

27. Freiwald & Smith, *supra* note 15, at 212.

28. *Id.*

29. *Id.* at 215–16. *See* Smith v. Maryland, 442 U.S. 735, 744 (1979); *Miller*, 425 U.S. at 443.

30. Freiwald & Smith, *supra* note 15, at 216.

31. *Carpenter v. United States*, 138 S. Ct. 2206, 2212 (2018).

32. *Id.*

33. *Id.* at 2213.

34. *Id.*

35. *Id.*

36. *United States v. Carpenter*, 819 F.3d 880, 888 (6th Cir. 2016).

not to leave citizens “‘at the mercy of advancing technology.’”³⁷ Here, CSLI collection represented a “seismic shift[] in digital technology,” one that gave law enforcement “detailed, encyclopedic, and effortlessly compiled” information that effectively amounted to “tireless and absolute surveillance.”³⁸ The data disclosed in this case, the Court warned, “provide[d] an intimate window into a person’s life, revealing not only his particular movements, but through them his ‘familial, political, professional, religious, and sexual associations’”—i.e., the same “privacies of life” that the Fourth Amendment was originally adopted to protect.³⁹ Also, the Court carefully pointed out, this information was not truly voluntarily given; cell phones were now “indispensable to participation in modern society,” and “apart from disconnecting the phone from the network, there [wa]s no way to avoid leaving behind a trail of location data.”⁴⁰ Given the uniquely intrusive nature of CSLI, then, the Court refused to extend the third party doctrine to this novel technology and held that accessing seven or more days of CSLI records constituted a Fourth Amendment search.⁴¹

Importantly, however, the Court tempered this finding with a caveat. “[O]ur decision today,” it emphasized, “is a narrow one.”⁴² The majority was careful to list the types of data that the decision did *not* address: real-time CSLI; “tower dumps” showing specific cell sites’ data over a given time period; “conventional surveillance techniques . . . such as security cameras”; business records containing location information; and collections involving national security or foreign policy.⁴³ However, the Court did implicitly acknowledge that its decision extended to GPS data.⁴⁴ At various turns, it observed that the “accuracy of CSLI [wa]s rapidly approaching GPS-level precision,” and that “CSLI data [wa]s less precise than GPS information,” effectively stating that GPS information collection was even more likely to constitute a Fourth Amendment search than CSLI access.⁴⁵

In the immediate aftermath of *Carpenter*’s June 2018 decision, many scholars perceived the case as a landmark ruling that would drastically change

37. *Carpenter v. United States*, 138 S. Ct. 2206, 2214 (2018) (quoting *Kyllo v. United States*, 533 U.S. 27, 35 (2001)).

38. *Id.* at 2216–19.

39. *Id.* at 2217.

40. *Id.* at 2220.

41. *Id.* at 2217.

42. *Id.* at 2220.

43. *Id.*

44. *Id.* at 2210; *see also* *United States v. Jones*, 565 U.S. 400 (2012) (finding that the government’s installation and tracking of a GPS device on the defendant’s car constituted a Fourth Amendment search).

45. *Carpenter*, 138 S. Ct. at 2210.

the landscape of digital Fourth Amendment cases. “Oceans of ink have been spilled by those worried about how the dramatic expansion of technologically fueled corporate surveillance of our private lives automatically expands police surveillance, too,” Professor Paul Ohm wrote in December 2018.⁴⁶ As such, *Carpenter* was “the opinion most privacy law scholars and privacy advocates ha[d] been awaiting for decades.”⁴⁷ Others agreed and predicted that *Carpenter* would provoke significant changes to the Fourth Amendment legal landscape. For *New York Times* reporter Adam Liptak, the case “ha[d] implications for all kinds of personal information held by third parties, including email and text messages, internet searches, and bank and credit card records”; for Sidley Austin attorneys Christopher Fonzone, Kate Heinzelman, and Michael Roberts, the decision “ha[d] potentially dramatic consequences not only for the government, but also for private industry holders of data.”⁴⁸ Evidently, then, stakeholders had high hopes that lower courts would soon broaden the *Carpenter* decision to other forms of digital data outside of CSLI. Such has not been the case, however, in the year since the decision.

V. LOWER COURT DECISIONS IN THE YEAR SINCE *CARPENTER*

In 2019, lower courts have taken the Supreme Court’s word and largely interpreted *Carpenter* as a “narrow” decision.⁴⁹ These rulings have been reasonable thus far because their fact patterns have not yet implicated *Carpenter*-level privacy issues, but police officers will soon use—indeed, have already used—alarmingly invasive techniques that *will* amount to the same “tireless and absolute surveillance” so feared by *Carpenter*.⁵⁰ As this Part will delineate, the technologies used in the last year’s cases each have the potential to breed practices that expose the “privacies of life” and violate the Fourth Amendment.⁵¹

46. Paul Ohm, *The Many Revolutions of Carpenter*, 32 HARV. J.L. & TECH. 358, 362 (2019).

47. *Id.*

48. Adam Liptak, *In Ruling on Cellphone Location Data, Supreme Court Makes Statement on Digital Privacy*, N.Y. TIMES (June 22, 2018), <https://www.nytimes.com/2018/06/22/us/politics/supreme-court-warrants-cell-phone-privacy.html>; Christopher Fonzone, Kate Heinzelman & Michael Roberts, *Carpenter v. United States: A Revolution in Fourth Amendment Jurisprudence?*, 4 PRATT’S PRIVACY & CYBERSECURITY LAW REPORT 283, 283 (Nov./Dec. 2018).

49. *See Carpenter*, 138 S. Ct. at 2220.

50. *See id.* at 2218.

51. *See id.* at 2214.

Future courts should continue to address *Carpenter*'s two core inquiries: (1) whether the methods used in the case constituted one of the "seismic shifts in digital technology" that *Carpenter* so feared, and (2) whether the activities at hand invaded the same "privacies of life" that *Carpenter* was so concerned with protecting.⁵² In answering these questions, judges should readily adopt the *Carpenter* Court's forward-looking approach: in Kerr's words, the ruling "seem[ed] more interested in where the technology [wa]s thought to be going" than the specific facts of the case.⁵³ Thus, going forward, judges should remain vigilant and extend *Carpenter* where necessary.

Each Section of this Part will discuss a type of digital data that has been addressed by courts in the past year: (A) pole cameras, (B) GPS devices, (C) IP addresses, and (D) home smart technology. In each Section, this Note will first summarize the last year's decisions for that data type and then consider the data type's future privacy implications.

A. POLE CAMERAS

First and perhaps most predictably, courts have largely remained loyal to the Supreme Court's assertion that *Carpenter* did not "call into question conventional surveillance techniques and tools, such as security cameras."⁵⁴ For example, in *United States v. Kay*, the U.S. District Court for the Eastern District of Wisconsin found that, "unlike the new technology addressed in *Carpenter*," law enforcement's pole camera surveillance did not constitute a Fourth Amendment search.⁵⁵ The defendant argued that such camera footage constituted the same "too permeating police surveillance" feared by the *Carpenter* Court, and he emphasized that he had a reasonable expectation that his home would not be constantly monitored by the police.⁵⁶ In response, the court stressed *Carpenter*'s status as a "limited decision" on a "new phenomenon" and observed that pole cameras had been used for decades.⁵⁷ It added that these cameras remained stationary in public spaces, so they were "unlikely to provide the same 'intimate window' into the person's life [that *Carpenter*'s CSLI collection could], revealing his 'political, professional,

52. *See id.*

53. Orin Kerr, *Understanding the Supreme Court's Carpenter Decision*, LAWFARE BLOG, (June 22, 2018, 1:18 PM), <https://www.lawfareblog.com/understanding-supreme-courts-carpenter-decision>.

54. *See Carpenter*, 138 S. Ct. at 2220.

55. *United States v. Kay*, No. 17-CR-16, 2018 U.S. Dist. LEXIS 141615, at *7 (E.D. Wis. Aug. 21, 2018).

56. *Id.* at 5.

57. *Id.* at 7 (quoting *Carpenter*, 138 S. Ct. at 2216).

religious, and sexual associations.’”⁵⁸ As such, the court concluded that law enforcement’s actions had not violated the Fourth Amendment.⁵⁹

The same district court ruled similarly in *United States v. Tirado*, which the court reconsidered in light of the *Carpenter* decision.⁶⁰ Here, too, the defendants challenged law enforcement’s use of pole camera surveillance outside of their residences.⁶¹ The court again emphasized the *Carpenter* Court’s note that its opinion did not “call into question conventional surveillance . . . such as security cameras,” and again asserted that the pole cameras at issue had been in use for decades.⁶² Further, though the defendants had drawn on *Carpenter*’s fears to argue that the surveillance “permit[ted] a detailed chronicle of a person’s activities,” the court was unconvinced.⁶³ After all, the court reasoned, the cameras never captured footage from inside the homes themselves; thus, the court found that defendants had “fail[ed] to explain how such surveillance provides the same aggregate amount of a person’s life, revealing his ‘political, professional, religious, and sexual associations’ [as CSLI could].”⁶⁴

It is worth noting, however, that a state court recently deviated from *Kay* and *Tirado* and extended *Carpenter* to a case where law enforcement conducted pole camera surveillance for an extended, continuous period of time. In November 2019’s *People v. Tafoya*, the Colorado Court of Appeals found that police’s long-term, constant use of a pole camera directed at a suspect’s house constituted a Fourth Amendment search.⁶⁵ In this case, officers streamed and recorded footage of the area around the defendant’s home, including regions behind his privacy fence, for over three months.⁶⁶ The court acknowledged that many other courts, like the *Kay* and *Tirado* court, did not consider “the nature, continuity, and extended duration of police observation” relevant to this Fourth Amendment analysis and likely would not have found a Fourth Amendment search here.⁶⁷ Nevertheless, this court disagreed; the judge here considered these factors, especially the duration of observation, to be “extremely relevant.”⁶⁸ Here, then, the

58. *Id.* (quoting *Carpenter*, 138 S. Ct. at 2217).

59. *Id.*

60. *United States v. Tirado*, No. 16-CR-168, 2018 U.S. Dist. LEXIS 141605, at *7 (E.D. Wis. Aug. 21, 2018).

61. *Id.* at 7.

62. *Id.* at 5–6 (quoting *Carpenter*, 138 S. Ct. at 2220).

63. *Id.* at 7.

64. *Id.*

65. *People v. Tafoya*, 2019 COA 176 (Colo. App. Nov. 27, 2019).

66. *Id.* ¶ 6.

67. *Id.* ¶ 33.

68. *Id.* ¶ 35.

surveillance's three-month length infringed on the suspect's reasonable expectation of privacy.⁶⁹ As the court observed, even if a neighbor could peer through the suspect's privacy fence and see all that this pole camera revealed, it would still be highly improbable that he or she would stand in place for three months; similarly, a helicopter or drone would not be able stay in the air above the backyard for three months.⁷⁰ The court therefore concluded that the warrantless, three-month-long surveillance of the defendant's home curtilage violated the Fourth Amendment.⁷¹

Given current pole camera surveillance practices, lower courts have been correct in their general reluctance to extend *Carpenter* to these "conventional surveillance techniques."⁷² Today, most pole cameras are installed by the government and remain stationary in public spaces, so the recordings available from these devices do not usually amount to *Carpenter*'s "detailed chronicle of a person's physical presence compiled every day, every moment."⁷³ The *Kay* and *Tirado* cases both heavily emphasized these facts in their discussions.⁷⁴ To the *Tafuya* court, of course, this surveillance does rise to *Carpenter*-level intrusiveness when conducted constantly for over three months; this distinction is reasonable, because three months' worth of footage reveals far more than a day's worth.⁷⁵ These opinions were thus all understandable given the present nature of pole cameras.

However, law enforcement has begun to augment these cameras with increasingly sophisticated technologies, and soon even these seemingly harmless "conventional" cameras may provide "an intimate window into a person's life."⁷⁶ For instance, camera surveillance could quickly constitute an intrusion on the "privacies of life" if these cameras are equipped with facial recognition technology.⁷⁷ Such a case will likely appear in the near future, since the United States is, in the words of *MIT Technology Review*'s Angela Chen, "smack in the middle of an era when cameras on the corner can

69. *See id.*

70. *Id.* ¶¶ 47–48.

71. *Id.* ¶ 51.

72. *See Carpenter v. United States*, 138 S. Ct. 2206, 2220 (2018).

73. *See id.*

74. *See id.*; *United States v. Kay*, No. 17-CR-16, 2018 U.S. Dist. LEXIS 141615, at *7 (E.D. Wis. Aug. 21, 2018); *United States v. Tirado*, No. 16-CR-168, 2018 U.S. Dist. LEXIS 141605, at *7 (E.D. Wis. Aug. 21, 2018).

75. *See Tafuya*, 2019 COA 176.

76. *See id.*

77. *See Carpenter*, 138 S. Ct. at 2220.

automatically recognize passersby.”⁷⁸ While facial recognition is not yet a ubiquitous feature in pole cameras, law enforcement in regions such as Orlando, Florida, and Washington County, Oregon, have already piloted this software on street surveillance cameras.⁷⁹

One can only imagine the numerous “privacies of life” that will be intruded upon in a future where nearly every pole camera is equipped with facial recognition technology. Armed with this software, the police could effortlessly follow a person through the same “trips to the psychiatrist, the plastic surgeon, the abortion clinic, the AIDS treatment center, the strip club, [and] the criminal defense attorney” that *Jones* feared GPS data would expose.⁸⁰ Indeed, *Carpenter’s* warning about CSLI might soon apply to camera surveillance, too: “Unlike the nosy neighbor who keeps an eye on comings and goings, they are ever alert, and their memory is nearly infallible.”⁸¹ So-called traditional technologies like pole cameras therefore may themselves experience “seismic shifts in technology” and lend themselves to Fourth-Amendment-violating practices in the near future.⁸²

Additionally, the fact, emphasized by *Kay*, that pole cameras have been used for decades should have no bearing on Fourth Amendment analyses.⁸³ After all, any technology, no matter how invasive, will eventually become a “conventional” technique that has been in use for decades.⁸⁴ The focus instead should be on whether such cameras would intrude upon the sensitive privacies *Carpenter* wished to protect.

Thus, the *Kay*, *Tirado*, and *Tajofa* courts were reasonable in their decisions applying *Carpenter* to pole camera surveillance in the last year. Going forward, courts should remain cautious and keep *Carpenter’s* key inquiries in mind as police officers begin to implement facial recognition software.⁸⁵

78. Angela Chen, *This is How You Kick Facial Recognition Out of Your Town*, MIT TECH. REV. (Oct. 4, 2019), <https://www.technologyreview.com/s/614477/facial-recognition-law-enforcement-surveillance-private-industry-regulation-ban-backlash>.

79. Matt Cagle & Nicole Ozer, *Amazon Teams Up with Government to Deploy Dangerous New Facial Recognition Technology*, FREE FUTURE (May 22, 2018 10:00AM), <https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazon-teams-government-deploy-dangerous-new>.

80. See *Carpenter*, 138 S. Ct. at 2220; *United States v. Jones*, 565 U.S. 400, 415 (2012) (quoting *People v. Weaver*, 12 N.Y.3d 433, 441–42 (2009)).

81. See *Carpenter*, 138 S. Ct. at 2219.

82. See *id.* at 2214.

83. See *United States v. Kay*, No. 17-CR-16, 2018 U.S. Dist. LEXIS 141615, at *7 (E.D. Wis. Aug. 21, 2018).

84. See *Carpenter v. United States*, 138 S. Ct. 2206, 2220 (2018).

85. See *id.* (quoting *Northwest Airlines, Inc. v. Minnesota*, 322 U.S. 292, 300 (1944)).

B. GPS DEVICES

Lower courts have followed *Carpenter*'s observations on GPS data's intrusiveness and extended the ruling to GPS information collection. For example, in *United States v. Diggs*, the U.S. District Court for the Northern District of Illinois held that detectives' collection of GPS data violated the Fourth Amendment.⁸⁶ Law enforcement officials did not themselves install the GPS tracking device in this case.⁸⁷ Instead, they accessed retroactive GPS information spanning over a month from a device installed on the defendant's vehicle by a previous owner.⁸⁸ The court identified this data as "fit[ting] squarely within the scope of the reasonable expectation of privacy identified by the *Jones* concurrences and reaffirmed in *Carpenter*"; the records did, after all, supply the same level of "detailed, encyclopedic, and effortlessly compiled" information as the *Carpenter* CSLI, and certainly "provide[d] an intimate window into a person's life."⁸⁹ The court thus found that the government had indeed infringed upon the defendant's reasonable expectation of privacy and violated the Fourth Amendment.⁹⁰

This court properly extended *Carpenter* to law enforcement's collection of GPS tracking data. The *Diggs* court was correct to observe that GPS information "fit[s] squarely within the scope of the reasonable expectation of privacy."⁹¹ In fact, as the *Carpenter* Court commented, GPS data is currently even *more* precise than CSLI—GPS can locate an individual within fifteen feet, while CSLI can only estimate to a range of one-half to two miles—and so supplies an even more "detailed, encyclopedic, and effortlessly compiled" record of a person's movements.⁹²

C. IP ADDRESSES

Thus far, courts have also declined to extend the *Carpenter* ruling to IP address information collection. Judges comparing IP addresses to CSLI have focused on: (1) the fact that IP addresses are not logged as frequently as CSLI, and (2) the additional steps that law enforcement agents must take even after receiving an IP address to determine the user's identity and location.

86. *United States v. Diggs*, 385 F. Supp. 3d 648, 655 (N.D. Ill. 2019).

87. *Id.* at 650.

88. *Id.*

89. *Id.* at 653 (quoting *Carpenter*, 138 S. Ct. at 2216–18) (internal quotations omitted).

90. *Id.*

91. *Id.*

92. *See Carpenter*, 138 S. Ct. at 2216–18 (internal quotations omitted); *see also id.* at 2225 (Kennedy, J., dissenting).

In *United States v. Hood*, the First Circuit determined that the police's IP address information collection did not implicate the specific concern noted in *Carpenter* and so did not constitute a Fourth Amendment search.⁹³ In this case, law enforcement accessed the recent IP addresses associated with an account on Kik, a smartphone messaging application.⁹⁴ Officials then gathered location information on the IP addresses from the digital communications providers controlling them and subsequently found the defendant.⁹⁵ The defendant challenged this warrantless IP address data collection, contending that the act was analogous to the CSLI in *Carpenter* since it allowed officials to access his exact location when he logged on to Kik.⁹⁶ The defendant further argued that “[t]he notion that anytime one accesses the internet from their cell phone, they are effectively providing the police a specific record of their whereabouts, [wa]s in direct contrast to society's expectations.”⁹⁷ However, the First Circuit was unconvinced and declined to find the police's activities to be a Fourth Amendment search for two reasons. First, unlike CSLI, which was recorded every time a person received a call, text message, or email and even during automatic application updates, IP address data was only generated when a user “ma[de] the affirmative decision to access a website or application.”⁹⁸ Second, the IP address data “d[id] not itself convey any location information” and was “merely a string of numbers associated with a device,” while CSLI immediately revealed a person's location “without any independent investigation.”⁹⁹

The First Circuit continued in this vein in another decision the same month: *United States v. Morel*.¹⁰⁰ Here, image-hosting site Imgur provided law enforcement with the IP address associated with certain images uploaded to Imgur's servers; officials then learned the IP address owner's identity from Comcast.¹⁰¹ The defendant—the owner of the IP address—called on *Carpenter* in his argument, arguing that the ruling had “effected a sea change [sic] in the law of reasonable expectation of privacy, and he [wa]s the beneficiary of that change.”¹⁰² The court rejected this contention, noting the

93. *United States v. Hood*, 920 F.3d 87, 94 (1st Cir. 2019).

94. *Id.* at 88–89.

95. *Id.*

96. *Id.* at 91–92.

97. *Id.* (internal quotations omitted).

98. *Id.* at 92.

99. *Id.*

100. *United States v. Morel*, 922 F.3d 1, 4 (1st Cir. 2019).

101. *Id.* at 9.

102. *Id.* at 8.

same two reasons mentioned by the *Hood* opinion. “IP address information of the kind and amount collected here,” the court concluded, “simply does not give rise to the concerns identified in *Carpenter*.”¹⁰³

The U.S. District Court for the District of Rhode Island reached the same holding in *United States v. Monroe*.¹⁰⁴ Here, law enforcement gathered the IP addresses of devices that had downloaded certain illicit files via a Georgia-based internet file sharing service (FSS).¹⁰⁵ Agents then learned the IP addresses’ owner’s—i.e., the defendant’s—identity from an internet service provider.¹⁰⁶ The court, in a discussion similar to the First Circuit’s observations in *Hood* and *Morel*, asserted that the collected IP address information “was not an ‘exhaustive chronicle’ of [the defendant’s] physical or digital activities,” since it “c[ould] only provide ‘the location at which one of any number of computer devices may be deployed, much like a telephone number can be used for any number of telephones.’”¹⁰⁷ As such, the IP addresses themselves did not immediately reveal the user’s identity and served only as “one link held by a third party in a chain of information that may lead to a particular person,” with none of *Carpenter*’s “minutely detailed, historical portrait of ‘the whole of [a person’s] physical movements.’”¹⁰⁸ The court therefore ruled that the defendant did not have a reasonable expectation of privacy in the data collected, so the police’s activities did not count as a Fourth Amendment search.¹⁰⁹

Courts have properly declined to apply *Carpenter* in their IP address data opinions thus far, but they may need to find Fourth Amendment violations as fact patterns change in the near future. *Hood*, *Morel*, and *Monroe* each distinguished IP addresses from CSLI by arguing that IP addresses were not logged nearly as frequently as CSLI, and that IP addresses were multiple steps removed from the user’s identity and location.¹¹⁰ These decisions were appropriate given current law enforcement’s limited access to IP address information in these instances. However, if law enforcement’s IP address surveillance becomes more comprehensive, courts should carefully reevaluate

103. *Id.* at 9.

104. *United States v. Monroe*, 350 F. Supp. 3d 43, 49 (D.R.I. 2018).

105. *Id.* at 44.

106. *Id.*

107. *Id.* at 48.

108. *Id.* at 49.

109. *Id.*

110. *See Hood*, 920 F.3d at 91–92; *United States v. Morel*, 922 F.3d 1, 8 (1st Cir. 2019); *United States v. Monroe*, 350 F. Supp. 3d 43, 48–49 (D.R.I. 2018).

the distinctions they drew in *Hood*, *Morel*, and *Monroe*—IP addresses could actually expose just as much of the “privacies of life” as CSLI.¹¹¹

Police officers could feasibly access IP address information via two different avenues. In the first, more limited method, law enforcement officials could discover an illicit online behavior—e.g., a forum post, a file upload, an instant message—and reach out to the platform (and, later, the relevant internet service provider) for information on the specific user involved. Officers thus may gain IP addresses, physical addresses, and names through this practice, but their entire query is restricted to the particular illegal incident they uncovered. In the second, far broader approach, police officers could target the entire history of a criminal suspect’s online activity. They can gather the suspect’s IP addresses from their internet service provider, translate the IP addresses into a domain name, and then identify the websites the suspect was visiting in a given time period.

In the last year, lower courts’ Fourth Amendment IP address rulings have involved only the first technique. Police officers in these cases accessed solely the narrowest information on a particular user in a particular situation. Law enforcement only collected the recent IP addresses associated with a specific Kik account in *Hood*; the IP addresses associated with specific images uploaded to Imgur in *Morel*; and the IP addresses of devices that had downloaded specific files via an FSS in *Monroe*.¹¹² Because these fact patterns addressed isolated moments in a person’s online actions and not their internet history in its entirety, none of these law enforcement practices amounted to the “tireless and absolute surveillance” discussed in *Carpenter*.¹¹³ Thus, these lower courts were correct not to broaden *Carpenter*, because the police’s actions were not nearly invasive enough to implicate the “privacies of life.”¹¹⁴

Lower courts *should* extend *Carpenter*, however, if law enforcement begins to adopt the second method and track an IP address’s entire browsing history. The courts will also need to reexamine their CSLI-versus-IP-address distinctions in the process.

First, the *Hood*, *Morel*, and *Monroe* courts use the wrong framing when they contend that IP address information is not as privacy-intrusive as CSLI because it is recorded less frequently than CSLI. Such a comparison focuses on how often IP addresses reveal the *physical* location of users; however, the proper equivalent analogy would be how often IP addresses reveal users’

111. *Carpenter v. United States*, 138 S. Ct. 2206, 2217 (2018).

112. *See Hood*, 920 F.3d at 88–89; *Morel*, 922 F.3d at 8; *Monroe*, 350 F. Supp. 3d at 48–49.

113. *See Carpenter*, 138 S. Ct. at 2218.

114. *See id.* at 2217.

virtual location. Just as a person’s CSLI “tracks nearly exactly the movements of its owner” in the physical world, IP addresses track nearly exactly the movements of their owners in the *virtual* world—every website a user visits sees that user’s IP address.¹¹⁵ To again cite *Jones*’s concerns, IP addresses could thus easily expose the online equivalents of a user’s “trips to the psychiatrist, the plastic surgeon, the abortion clinic, the AIDS treatment center, the strip club, [and] the criminal defense attorney”; for example, these addresses could show that the user visited the websites of any of these places or visited forum pages covering sensitive topics related to criminal defense or strip clubs.¹¹⁶ Through this *virtual* location lens, then, IP addresses reveal location just as, if not more, frequently than CSLI does, and could no doubt invade the “privacies of life” as well.¹¹⁷

The courts’ other distinction that IP addresses are merely “one link . . . in a chain of information that may lead to a particular person” is similarly misguided.¹¹⁸ These cases place too heavy an emphasis on the few additional steps law enforcement would need to take to ascertain an IP address owner’s identity. Even more critically, the courts miss the fact that *Carpenter* still found a search even after acknowledging that the government could only “deduce a detailed log of Carpenter’s movements” from CSLI “*in combination with* other information.”¹¹⁹ The additional steps are often negligible. Once law enforcement obtains an IP address, it need only reach out to the service provider controlling that address to learn the address’s owner. Indeed, the Office of the Privacy Commissioner of Canada recently reported that one could “build a detailed profile of a person or group associated with the IP address” by merely “carrying out . . . a simple test” of searching through public databases online, “no special equipment or software . . . needed.”¹²⁰ Additionally, *Monroe*’s contention that IP address data “c[ould] only provide ‘the location at which one of any number of computer devices may be deployed’ ” ignores the reality that many public IP addresses are shared by a small number of people.¹²¹ Setting aside businesses and academic institutions—where internet users are probably less likely to engage in

115. *See id.* at 2218.

116. *See* United States v. Jones, 565 U.S. 400, 416 (2012) (quoting *Weaver*, 12 N.Y.3d at 441–42).

117. *See Carpenter*, 138 S. Ct. at 2220.

118. *See* United States v. Monroe, 350 F. Supp. 3d 43, 49 (D.R.I. 2018).

119. *See Carpenter*, 138 S. Ct. at 2218 (emphasis added).

120. Technology Analysis Branch of the Office of the Privacy Commissioner of Canada, *What an IP Address Can Reveal About You*, OFF. OF THE PRIVACY COMM’R OF CAN. (May 2013), https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2013/ip_201305.

121. *See Monroe*, 350 F. Supp. 3d at 48.

incriminating activities in the first place—home internet routers (which each have unique public IP addresses) usually have only a handful of regular users.¹²² IP address information therefore is not merely one “link” in a long chain of clues leading to an individual; in many cases, it serves as a crucial, identifying segment of the chain.¹²³

To conclude, then, while the *Hood*, *Morel*, and *Monroe* courts appropriately declined to broaden *Carpenter* to cover the fact patterns at issue, their reasoning may not hold water as police officials begin to track IP addresses’ entire browsing histories. In today’s increasingly digital age—when citizens spend perhaps the same amount of time surfing the internet as they do moving around in the physical world—IP address data has the potential to expose as much information as CSLI.

D. HOME IoT DEVICES

Courts have yet to consider home IoT device data, but the last year’s opinions on government access to home public utility data provide some guidance on how courts may handle IoT in the future. For example, the Seventh Circuit’s *Naperville Smart Meter Awareness v. City of Naperville* decision concerned a city government’s collection of home electricity usage information.¹²⁴ The City of Naperville in this case entered all electricity-enabled Naperville homes into a mandatory smart-meter program that collected residents’ energy-usage data at fifteen-minute intervals; the City then stored these records for up to three years.¹²⁵

The *Naperville* court likened the program to law enforcement’s CSLI collection in *Carpenter*.¹²⁶ In the court’s words:

If a person does not—in any meaningful sense—‘voluntarily “assume the risk” of turning over a comprehensive dossier of physical movements’ by choosing to use a cell phone [as in

122. See Bradley Mitchell, *How Many Devices Can Connect to One Wireless Router*, LIFEWIRE (Apr. 15, 2020), <https://www.lifewire.com/how-many-devices-can-share-a-wifi-network-818298> (observing that most home networks use a single wireless access point); Richard Fry, *The Number of People in the Average U.S. Household is Going Up for the First Time in Over 160 Years*, PEW RESEARCH CENTER (Oct. 1, 2019), <https://pewresearch.org/fact-tank/2019/10/01/the-number-of-people-in-the-average-u-s-household-is-going-up-for-the-first-time-in-over-160-years> (finding that there was an average of 2.63 people per household in the United States in 2018). Given these statistics, a single household router likely has around 2.63 regular users on average.

123. See *id.* at 49.

124. See *Naperville Smart Meter Awareness v. City of Naperville*, 900 F.3d 521, 521 (7th Cir. 2018).

125. *Id.* at 524.

126. *Id.* at 527.

Carpenter], . . . it also goes that a home occupant does not assume the risk of near constant monitoring by choosing to have electricity in her home.¹²⁷

After all, “a choice to share data imposed by fiat is no choice at all.”¹²⁸ Additionally, it noted, the “technology-assisted data collection” at issue, i.e., the smart-meter measurement of electricity usage, was both “not in general public use” and “at least as rich as that found to be a search in *Kyllo*.”¹²⁹ As such, the court concluded, the City’s activities were a search under the Fourth Amendment.¹³⁰

Importantly, though, and unlike the other cases in this Note, the City here “conduct[ed] the search with no prosecutorial intent.”¹³¹ Only public utility employees, not law enforcement, accessed the data.¹³² This fact, the court emphasized, “lessen[ed] an individual’s privacy interest.”¹³³ The government, meanwhile, did have a significant, legitimate interest in this information (for the sake of providing cheaper energy, promoting power efficiency, and helping grid stability), and only collected usage data at fifteen-minute intervals.¹³⁴ After balancing the lowered privacy interest against the government’s legitimate interests and limited practices, the Seventh Circuit held that the act, while a search, was still a *reasonable* search that did not violate the Fourth Amendment.¹³⁵ Thus, *Naperville* provided a rather convoluted, inconclusive application of *Carpenter* principles.¹³⁶

However, the *Naperville* opinion did cast light on *Carpenter*’s potential applicability to IoT device data.¹³⁷ In its acknowledgement of the search’s reasonableness, the *Naperville* court warned that its finding “depend[ed] on the particular circumstances of this case,” and if “a city [were] to collect the data at shorter intervals, [the court’s] conclusion could change.”¹³⁸ “Likewise,” it added, “[the court’s] conclusion might change if the data was more easily accessible to law enforcement or other city officials outside the

127. *Id.*

128. *Id.*

129. *Id.* at 526 (quoting *Kyllo v. United States*, 533 U.S. 27, 40 (2001)) (internal quotations omitted).

130. *Id.*

131. *Id.* at 527.

132. *Id.* at 528.

133. *Id.*

134. *Id.* at 526.

135. *Id.*

136. *See id.*

137. *Id.* at 529.

138. *Id.*

utility.”¹³⁹ From these statements, then, the Seventh Circuit implied that, in the future, *Carpenter* may well extend to, say, cities’ collection of electricity usage at ten-minute intervals or law enforcement’s warrantless access to fifteen-minute-interval electricity usage data.¹⁴⁰

As law enforcement inevitably begins to investigate IoT data in the future, courts should follow *Naperville*’s lead and keep *Kyllo* in mind.¹⁴¹ With the recent advent of smart home devices such as Google Home and Amazon Echo, this issue will likely arise in courts soon. Once homes become fully wired with these devices, courts should continue to read police collection of transactional data (e.g., records showing when certain lights in the home are turned on or off) as Fourth Amendment searches that require a warrant.

One could easily argue that home IoT transactional data, like the *Naperville* smart meter program’s electricity usage information, is also “at least as rich as that found to be a search in *Kyllo*.”¹⁴² Law enforcement agents in *Kyllo* merely used thermal images to detect infrared radiation from inside *Kyllo*’s house.¹⁴³ Transactional usage data pulled from an IoT home device, meanwhile, could expose *Kyllo*-like infrared lights and much more; from the information, officials could glean when residents arrived at home, entered and left rooms, and went to sleep.¹⁴⁴ Further, as Justice Scalia stated, “in the case of the search of the interior of homes”—which certainly includes the collection of home device data—“. . . there is a ready criterion . . . of the minimal expectation of privacy . . . acknowledged to be *reasonable*.”¹⁴⁵

Additionally, home IoT data could be likened to an equivalent of a person’s cell phone in their home. Like the cell phone, which *Carpenter* described as “almost a ‘feature of human anatomy’ . . . [that] tracks nearly exactly the movements of its owner,”¹⁴⁶ a device tracking which lights are turned on in a home can reveal exactly which room a person was in at any time (indeed, it may even be more precise than a cell phone in these instances, since many people do not keep their cell phones with them at all times while at home). In light of *Kyllo* and even *Carpenter*, then, a court should readily conclude that access to home devices’ transactional records would

139. *Id.*

140. *Id.*

141. *See id.* at 526; *Kyllo v. United States*, 533 U.S. 27, 40 (2001).

142. *See Naperville*, 900 F.3d at 526 (quoting *Kyllo*, 533 U.S. at 40) (internal quotations omitted).

143. *See Kyllo*, 533 U.S. at 29–30.

144. *See id.*

145. *See id.* at 34.

146. *See Carpenter v. United States*, 138 S. Ct. 2206, 2218 (2018).

violate a citizen's reasonable expectation of privacy and so would constitute a Fourth Amendment search.

In response, law enforcement agencies may assert that people with home IoT technologies have voluntarily assumed the risk of having their data turned over to the police. While this claim may hold water today, it will likely become less and less effective as home IoT devices become increasingly ubiquitous. Fifty-nine percent of American adults surveyed in a 2018 Forrester Research report said they were interested in using a smart home device; as *New York Times* reporter Janet Morrissey observed in January 2019, “the soaring popularity of smart speakers . . . is starting to move the ‘Smart Home’ into mainstream America.”¹⁴⁷ As the number of smart-device-wired homes grows, citizens will have increasingly compelling arguments that these devices, like the cell phones of *Carpenter*, are “‘such a pervasive and insistent part of daily life’ that [owning] one is indispensable to participation in modern society.”¹⁴⁸ Thus, in the near future, those with home IoT will likely be able to convincingly claim that they, like the residents of Naperville, do not “‘voluntarily ‘assume the risk’ of turning over a comprehensive dossier of physical movements’” by having smart homes.¹⁴⁹

As these smart home devices increase in prevalence, the government may invoke *Kyllo*'s requirement that the technology not be “in general public use.”¹⁵⁰ However, this “general public use” qualification is a problematic standard for reasonable privacy assessments; indeed, it may, in the words of Derek Conom in the *Willamette Law Review*, prove to be a “potentially troublesome sliding scale of privacy that depends on how fast technology goes into general public use.”¹⁵¹ As Justice Stevens in his *Kyllo* dissent observed: “[T]his [general public use] criterion is somewhat perverse because it seems likely that the threat of privacy will grow, rather than recede, as the use of intrusive equipment becomes more readily available.”¹⁵² This concerning standard pulls the emphasis away from ordinary citizens' privacy expectations and instead focuses on potential privacy invaders' actions. The

147. Janet Morrissey, *In the Rush to Join the Smart Home Crowd, Buyers Should Beware*, N.Y. TIMES (Jan. 22, 2019), <https://www.nytimes.com/2019/01/22/business/smart-home-buyers-security-risks.html>.

148. See *Carpenter*, 138 S. Ct. at 2220 (quoting *Riley v. California*, 134 S. Ct. 2473, 2484 (2014)).

149. See *Naperville Smart Meter Awareness v. City of Naperville*, 900 F.3d 521, 527 (7th Cir. 2018) (quoting *Carpenter*, 138 S. Ct. at 2220).

150. See *Kyllo v. United States*, 533 U.S. 27, 40 (2001).

151. Derek T. Conom, *Sense-Enhancing Technology and the Search in the Wake of Kyllo v. United States: Will Prevalence Kill Privacy?*, 41 WILLAMETTE L. REV. 749, 761 (2005).

152. See *Kyllo*, 533 U.S. at 47 (Stevens, J., dissenting).

Kyllo majority claimed the *Katz* test supported this public use requirement—presumably because it believed people should expect police to employ widely-used surveillance technologies—but the majority failed to consider a future in which law enforcement uses technologies that citizens simply cannot prepare for or protect themselves against.¹⁵³ In *Naperville*'s words, “under *Kyllo* . . . even an extremely invasive technology can evade the warrant requirement if it is ‘in general public use.’”¹⁵⁴ Citizens should not be at the whim of the general public’s use of a surveillance technology, and courts should instead draw a hard line so that officials’ handling of “extremely invasive technolog[ies]” would still constitute a Fourth Amendment search.¹⁵⁵

Thus, even as smart home devices increase in prevalence over the coming years, courts should remain wary of law enforcement’s access to these technologies’ transactional data and continue to find that collection of such information constitutes a Fourth Amendment search.

VI. FUTURE TECHNOLOGIES RELEVANT TO *CARPENTER*

Law enforcement has begun to turn to big data surveillance in its tracking of citizens, and courts should be cautious and protective of citizens’ privacy when adjudicating big-data-related Fourth Amendment cases. Big data mining—defined by Professor Adam Frank as “the machine-based collection and analysis of astronomical quantities of information”—can capture astonishingly intricate profiles of individuals’ behaviors.¹⁵⁶ As they consider these cases, courts should keep the dangerously invasive potential of big data in mind, even if the particular facts at issue do not yet involve extremely sophisticated technology. The *Carpenter* Court, after all, “seem[ed] more interested in where the technology [wa]s thought to be going,” and *Kyllo* asserted that Fourth Amendment rulings must take heed of more intrusive systems that are already in use.¹⁵⁷

As an example, U.S. Immigration and Customs Enforcement (ICE) officials have begun to mine previously unrelated computer databases for data on every realm of citizens’ lives. As a recent *New York Times* article reported: “[T]he business of deportation, like so much else in the modern

153. See *Katz v. United States*, 389 U.S. 353 (1967); *Kyllo*, 533 U.S. at 34–35.

154. See *Naperville*, 900 F.3d at 527.

155. See *id.*

156. Adam Frank, *A Brave New World: Big Data’s Big Dangers*, NATIONAL PUBLIC RADIO 13.7 COSMOS & CULTURE (June 11, 2013 2:41 PM), <https://www.npr.org/sections/13.7/2013/06/10/190516689/a-brave-new-world-big-datas-big-dangers>.

157. Kerr, *supra* note 53; see *Kyllo*, 533 U.S. at 36.

world, has been transformed by the power of big data.”¹⁵⁸ Under “relentless pressure from the White House to deport people,” ICE agents have begun to “suck[] up terabytes of information from hundreds of disparate computer systems, from state and local governments, from private data brokers and from social networks . . . fusing little bits of stray information together into dossiers.”¹⁵⁹ While immigrants are the main targets of these particular searches, “it’s an increasingly trivial exercise to track any of us.”¹⁶⁰ These two passages perfectly capture the precise facts that so concerned the Supreme Court in *Carpenter*.

First, this all-inclusive brand of tracking certainly “provide[s] an intimate window into a person’s life, revealing . . . his ‘familial, political, professional, religious, and sexual associations,’” like CSLI.¹⁶¹ Indeed, in today’s increasingly digital world, it arguably reveals an even more “comprehensive dossier,” since it exposes every aspect of a person’s online presence.¹⁶² Also, as with CSLI, citizens cannot be said to have voluntarily consented to giving others this aggregated information. The internet is just as “indispensable to participation in modern society” as cell phones; in particular, social media is now inextricably tied to people’s abilities to contact friends, read news, find jobs, and learn about the world around them.¹⁶³ Further, internet users probably do not quite understand the intrusive potential of big data, and while they may be aware of each individual social media platform’s privacy implications, they likely have not yet imagined just how much an aggregation of *all* their social media accounts can expose about their personal behaviors and preferences.

Second, this type of all-encompassing tracking is just as “effortlessly compiled” as the CSLI in *Carpenter*.¹⁶⁴ The Supreme Court particularly feared the reality that CSLI was “remarkably easy, cheap, and efficient” and could happen “with just the click of a button.”¹⁶⁵ The *New York Times* noted that ICE’s immigrant tracking practices are “increasingly trivial.”¹⁶⁶ For example, dozens of ICE officers have accounts on a “flexible search” computer

158. McKenzie Funk, *How ICE Picks Its Targets in the Surveillance Age*, N.Y. TIMES MAGAZINE, <https://www.nytimes.com/2019/10/02/magazine/ice-surveillance-deportation.html> (last updated Oct. 3, 2019).

159. *Id.*

160. *Id.*

161. *See* *Carpenter v. United States*, 138 S. Ct. 2206, 2217, 2220 (2018).

162. *See id.*

163. *See id.* at 2220.

164. *See id.* at 2216.

165. *See id.* at 2218.

166. Funk, *supra* note 158.

interface known as Driver and Plate Search on which they run tens of thousands of searches “again and again, nearly every day at all times of day.”¹⁶⁷ Like CSLI, these databases do not require officials to engage in lengthy stakeouts, search through piles of documents, or even leave their desks; instead, they can find vast troves of personal information on citizens “with just the click of a button.”¹⁶⁸

Finally, *Carpenter*’s concern with the “vast store of sensitive information” in a cell phone’s “immense storage capacity” applies to ICE’s use of big data tracking since in this case, the “store of sensitive information” is the internet itself.¹⁶⁹ In the realm of big data, no public record on the internet is safe from the government, and it is difficult to imagine something that has a more “immense storage capacity” than the internet itself.¹⁷⁰

In the near future, these searches may be deemed even more intrusive and harmful with the addition of artificial intelligence (AI) algorithms. In July 2017, ICE announced the “Extreme Vetting Initiative,” a project that aimed to evangelize “determinations via automation.”¹⁷¹ It referenced “partners whose algorithms could scan social media and other publicly available information” to “assess whether an immigrant was likely to become a ‘positively contributing member of society’—or whether he or she intended ‘to commit criminal or terrorist attacks.’”¹⁷² This new initiative is particularly concerning given the reality that AI cannot currently be designed without biases. In the words of *MIT Technology Review*’s Karen Hao, “bias can creep in at many stages of the deep-learning process, and the standard practices in computer science aren’t designed to detect it.”¹⁷³ In November 2017, Google Translate announced that its AI algorithms were sexist; in October 2018, Amazon scrapped its attempts at using AI to screen potential applicants in light of the software’s aversion to resumes that contained the word “women.”¹⁷⁴ Not only, then, do ICE officers now practice the same “tireless and absolute surveillance” dreaded by *Carpenter*; they could soon also use this information to profile and discriminate against immigrants—who already

167. *Id.*

168. *See Carpenter*, 138 S. Ct. at 2218.

169. *See id.* at 2214.

170. *See id.*

171. Funk, *supra* note 158.

172. *Id.*

173. Karen Hao, *This is How AI Bias Really Happens—and Why It’s So Hard to Fix*, MIT TECH. REV. (Feb. 4, 2019), <https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix>.

174. Michael Li, *Addressing the Biases Plaguing Algorithms*, HARV. BUS. REV. (May 13, 2019), <https://hbr.org/2019/05/addressing-the-biases-plaguing-algorithms>.

suffer from a host of discrimination-related challenges—in frighteningly prejudiced ways.¹⁷⁵

Suffice it to say, then, that government officials now can and *have* easily pulled together vast troves of previously decentralized data to create alarmingly thorough profiles of every citizen. Indeed, the same *New York Times* report on ICE techniques mentioned that “public records make clear that . . . other federal agencies” also engage in ICE-like practices.¹⁷⁶ As discussed above, these practices both implicate severe *Carpenter* concerns and may lead to discriminatory practices that harm vulnerable populations. As Justice Brandeis emphasized in *Olmstead v. United States*, courts should be wary as these “[s]ubtler and more far-reaching means of invading privacy have become available.”¹⁷⁷ Judges should work to “ensure that the ‘progress of science’ does not erode Fourth Amendment protections” by readily extending *Carpenter* to big data surveillance cases.¹⁷⁸

VII. CONCLUSION

In the year since the allegedly groundbreaking *Carpenter* decision, lower courts have been reluctant to extend the holding to forms of digital data not explicitly mentioned in the opinion. Judges have largely declined to apply the *Carpenter* exception to cases involving pole camera footage, IP address information, and home device data. The only type of data these courts have clearly extended *Carpenter* to include is GPS data, which the *Carpenter* Court referenced as being even more intrusive than CSLI.

However, these findings appear to be more a reflection of the specific facts at issue than an indication that courts will never extend *Carpenter* beyond retroactive CSLI. These courts were in fact correct to construe *Carpenter* narrowly given the less invasive methods utilized in the last year’s cases. New surveillance techniques will inevitably proliferate in the coming years, though, so courts should continue pressing *Carpenter*’s key inquiries and seek to extend *Carpenter* in cases involving “seismic shifts in digital technology” that intrude upon “the privacies of life.”¹⁷⁹

Hope is not lost even if lower courts maintain their narrow *Carpenter* applications in the future; state legislatures may still forge ahead themselves and adopt privacy bills broadening *Carpenter*’s logic. Indeed, earlier this year,

175. See *Carpenter*, 138 S. Ct. at 2218.

176. Funk, *supra* note 158.

177. See *Carpenter*, 138 S. Ct. at 2223 (quoting *Olmstead v. United States*, 277 U.S. 438, 473–74 (1928)).

178. See *id.*

179. See *id.* at 2214, 2219.

Utah took a step in this direction with its Electronic Information or Data Privacy Act, which, with certain limited exceptions, requires law enforcement to obtain a warrant to access information shared with third parties.¹⁸⁰ This new bill may inspire other state legislatures to follow suit. Since Utah's law was voted into action, Maine and Illinois have enacted their own *Carpenter*-expanding privacy statutes.¹⁸¹ Even if the judiciary fails to properly extend *Carpenter*, then, the legislature may still intervene and help cement *Carpenter* as “the opinion most privacy law scholars and privacy advocates have been awaiting for decades.”¹⁸²

Finally, even if both courts and state legislatures fail to broaden *Carpenter* to other invasive technologies, corporate actors may themselves push back on privacy-violating law enforcement requests. Google's Sensorvault records, for example, have been protected by Google's demands of law enforcement. In the last few years, law enforcement agents have accessed Sensorvault—an enormous database of detailed location data from iOS devices with Google Maps installed and most Android devices—to investigate, arrest, and convict citizens.¹⁸³ Notably, though, officers have issued warrants each time they review Sensorvault data—Google, then, is holding officers to its own standard and is itself defining Sensorvault access as a Fourth Amendment search that requires a warrant.¹⁸⁴ While investigators have told the *New York Times* that they do not request this kind of information from companies besides Google, the Sensorvault database sets an interesting—even promising—example of the potential corporate-side interventions to come.¹⁸⁵ Hopefully, future companies will follow in Google's footsteps and similarly require warrants for law enforcement officials seeking access to the “privacies of life.”¹⁸⁶

It is important to note, however, that *Carpenter*'s broadening may not be enough to protect citizens from all unreasonable privacy invasions by the

180. Cynthia Cole, Brooke Chatterton & Sarah Phillips, *Utah Blazes Trail with Law Shielding Data from Gov't Search*, LAW360 (Sept. 17, 2019 11:50AM), <https://www.law360.com/articles/1198954/utah-blazes-trail-with-law-shielding-data-from-gov-t-search>.

181. *Id.*

182. Ohm, *supra* note 46.

183. Jennifer Valentino-DeVries, *Tracking Phones, Google is a Dragnet for the Police*, N.Y. TIMES (Apr. 13, 2019), <https://www.nytimes.com/interactive/2019/04/13/us/google-location-tracking-police.html>.

184. Jennifer Lynch, *Google's Sensorvault Can Tell Police Where You've Been*, ELEC. FRONTIER FOUND. (Apr. 18, 2019), <https://www.eff.org/deeplinks/2019/04/googles-sensorvault-can-tell-police-where-youve-been>.

185. Valentino-DeVries, *supra* note 183.

186. *See* *Carpenter v. United States*, 138 S. Ct. 2206, 2218 (2018).

police. The cases discussed in this Note did not contemplate what happens *after* police obtain a warrant for a Fourth Amendment search. What if there are data types that the police should not access under any circumstances, even with a warrant? Consider, for example, a Florida judge's recent approval of a warrant to search the database of GEDmatch, a consumer DNA site with nearly one million users.¹⁸⁷ These searches would affect "huge swaths of the population" outside of just site users; armed with this new forensic method, law enforcement would be able to identify individuals "even through distant family relationships."¹⁸⁸ With these warrants, police will thus have a free license to access millions upon millions of powerless, innocent citizens' DNA profiles, even if they have never used or heard of DNA consumer sites or have never spoken to the distant relative who did use the site.¹⁸⁹ As such, beyond *Carpenter*, there may still be searches so intrusive that law enforcement should never be able to conduct them, even with a warrant.

Carpenter may well find new life, then, in many different arenas—in courts, state legislatures, and even data-harvesting companies. The ruling could feasibly help protect citizens from "seismic shifts in digital technology" for years to come.¹⁹⁰ Extending *Carpenter* alone, though, may not be sufficient to protect citizens from invasive surveillance. As police techniques increase in sophistication over the coming years, judges and legislators should consider whether certain types of data should *never* be accessed by law enforcement, with or without a warrant.

187. Kashmir Hill & Heather Murphy, *Your DNA Profile is Private? A Florida Judge Just Said Otherwise*, N.Y. TIMES, <https://www.nytimes.com/2019/11/05/business/dna-database-search-warrant.html> (last updated Dec. 30, 2019).

188. *Id.*

189. *Id.*

190. *See Carpenter*, 138 S. Ct. at 2219.

BIPA AND ARTICLE III STANDING: ARE NOTICE AND CONSENT MORE THAN “BARE PROCEDURAL” RIGHTS?

Carmen Sobczak[†]

I. INTRODUCTION

Modern technology has increasingly come to rely on the private sector’s collection and use of biometric data. Many individuals willingly hand over this data to facilitate interactions with electronic devices, happy to scan fingerprints and facial geometries so they can unlock their smartphones without a password. Companies also give customers the option to supply biometrics for enhanced security. For example, several banks authenticate clients with fingerprints, voiceprints, face scans, or iris scans.¹ Some biometric data collection, however, is far less voluntary. Schools, employers, and churches use fingerprints and face scans to track attendance.² Hundreds of retail stores have purchased facial recognition systems to “identify known shoplifters,” and can share data with other businesses without providing any notice to their customers.³ Video doorbells equipped with facial recognition, like Nest,

DOI: <https://doi.org/10.15779/Z38W669904>

© 2020 Carmen Sobczak.

† J.D. 2021, University of California, Berkeley, School of Law. Sincere thanks to Jim Dempsey, Professors Kenneth Bamberger and Talha Syed, my fellow students in the 2019 Law & Technology Writing Workshop at Berkeley Law, and the Berkeley Technology Law Journal editors.

1. *See BofA Merrill Adds Biometrics and Integrated Token to CashPro® Mobile*, BANK OF AM. (July 2, 2018, 9:00 AM), <https://newsroom.bankofamerica.com/press-releases/corporate-and-investment-banking-sales-and-trading-treasury-services/bofa-merrill-18>; Dan Hansen, *Voiceprint: A Security Game-Changer for Banks and Credit Unions of All Sizes*, BIZTECH (Nov. 5, 2018), <https://biztechmagazine.com/article/2018/11/voiceprint-security-game-changer-banks-and-credit-unions-all-sizes>; *BBVA, The First Bank with Access to Its Mobile App Via Iris Scanning, Thanks to Samsung*, BBVA (Nov. 16, 2017), <https://www.bbva.com/en/bbva-first-bank-access-mobile-app-iris-scanning-thanks-samsung>.

2. *See* Nicole Nguyen, *Chokepoint: Regulating US Student Mobility Through Biometrics*, 46 POL. GEOGRAPHY 1, 2 (2015); Selena Larson, *Beyond Passwords: Companies Use Fingerprints and Digital Behavior to ID Employees*, CNN BUS. (Mar. 18, 2018, 3:35 PM), <https://money.cnn.com/2018/03/18/technology/biometrics-workplace/index.html>; *Face Recognition App — The Churchix App*, CHURCHIX, <https://churchix.com/face-recognition-app> (last visited Feb. 3, 2021).

3. *See* Alfred Ng, *With Facial Recognition, Shoplifting May Get You Banned in Places You’ve Never Been*, CNET (Mar. 20, 2019, 8:11 AM), <https://www.cnet.com/news/with-facial-recognition-shoplifting-may-get-you-banned-in-places-youve-never-been>.

empower the homeowner to scan and identify any person who approaches their door.⁴

While biometric information can greatly increase efficiency and security, its collection, use, and storage pose significant threats to individual privacy. Biometric identifiers are distinct from other forms of identification; unlike social security numbers, they are impossible—or, at least, incredibly difficult, expensive, and painful—to change.⁵ This permanence drastically raises the stakes of identity theft and fraud.⁶ Moreover, each novel application of biometric technology introduces new threats, such as the clear risk of racial profiling in software used to identify shoplifters.⁷ Risks like these are exacerbated by the fact that facial recognition technologies can be inaccurate and have been found to disproportionately misidentify people of color.⁸ The rapidly growing use of biometric data in the private sector also fuels fears surrounding corporate mass surveillance, as cameras with identification capacity can significantly chill freedoms of speech, assembly, and association.⁹ And corporate surveillance has begun to look more and more like government surveillance through public-private partnerships that give law enforcement agencies access to video footage from individually-owned cameras.¹⁰

4. See Samuel Gibbs, *Nest Hello Review: Google's Smart Facial-Recognition Video Doorbell*, THE GUARDIAN (Sept. 20, 2018, 4:33 AM), <https://www.theguardian.com/technology/2018/sep/20/nest-hello-review-google-smart-facial-recognition-video-doorbell>.

5. See Adam Schwartz, *The Danger of Corporate Facial Recognition Tech*, ELEC. FRONTIER FOUND. (June 7, 2016), <https://www.eff.org/deeplinks/2016/06/danger-corporate-facial-recognition-tech>.

6. See *id.*

7. See *id.*

8. Jacob Snow, *Amazon's Face Recognition Falsely Matched 28 Members of Congress with Mugshots*, AM. CIVIL LIBERTIES UNION (July 26, 2018, 8:00 AM), <https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28>.

9. See Farhad Manjoo, *San Francisco is Right: Facial Recognition Must be Put on Hold*, N.Y. TIMES (May 16, 2019), <https://www.nytimes.com/2019/05/16/opinion/columnists/facial-recognition-ban-privacy.html>; Sigal Samuel, *Activists Want Congress to Ban Facial Recognition. So They Scanned Lawmakers' Faces.*, VOX (Nov. 15, 2019, 10:10 AM), <https://www.vox.com/future-perfect/2019/11/15/20965325/facial-recognition-ban-congress-activism>.

10. See Drew Harwell, *Doorbell-Camera Firm Ring Has Partnered with 400 Police Forces, Extending Surveillance Concerns*, WASH. POST (Aug. 28, 2019), <https://www.washingtonpost.com/technology/2019/08/28/doorbell-camera-firm-ring-has-partnered-with-police-forces-extending-surveillance-reach>; Rani Molla, *Activists Are Pressuring Lawmakers to Stop Amazon Ring's Police Surveillance Partnerships*, VOX (Oct. 8, 2019, 7:00 AM), <https://www.vox.com/recode/2019/10/8/20903536/amazon-ring-doorbell-civil-rights-police-partnerships>.

In 2008, Illinois became the first state to enact a statute explicitly protecting biometric privacy.¹¹ Following the notable bankruptcy of Pay By Touch, a company that created systems for fingerprint-based purchasing, the General Assembly grew concerned about the fate of residents' biometric data.¹² Representative Kathy Ryg asserted that in light of the sale of the Pay By Touch database to a third party, Illinois was in "serious need of protections for [its] citizens . . . when it [came] to biometric information."¹³ The Illinois Biometric Information Privacy Act (BIPA), which regulates the collection and retention of biometric data and prohibits its sale, passed unanimously.¹⁴

While several states have since approved similar legislation giving individuals control over their biometric data, BIPA is thus far unique in that it empowers Illinois residents to enforce that control through a private right of action.¹⁵ Those who are "aggrieved by a violation of" the statute can sue for liquidated damages of up to \$5,000.¹⁶

Beginning in 2019, courts began to see a "flood" of class actions brought under BIPA, including a lawsuit against Facebook that settled for \$650 million.¹⁷ Many of these cases raised the question of whether plaintiffs can sue for a procedural infringement of BIPA without showing additional harm. The Illinois Supreme Court definitively held in 2019 that they can, establishing that "an individual need not allege some actual injury or adverse effect, beyond violation of his or her rights under the Act," to bring suit in state court.¹⁸ But federal courts split as to whether such infringements satisfy constitutional

11. See 740 ILL. COMP. STAT. 14/1–14/99 (2008); Thomas F. Zych, Steven G. Stransky & Brian Doyle-Wenger, *State Biometric Privacy Legislation: What You Need to Know*, LEXOLOGY (Sept. 5, 2019), <https://www.lexology.com/library/detail.aspx?g=ebc0e01c-45cc-4d50-959e-75434b93b250>.

12. H.R. Deb. Transcript, 95th Gen. Assemb. No. 276 (Ill. 2008) (statement of Rep. Kathy Ryg).

13. *Id.*

14. See S.B. 2400: S. Vote, 3d Reading, 95th Gen. Assemb. (Ill. 2008); S.B. 2400: H.R. Roll Call, 3d Reading, 95th Gen. Assemb. (Ill. 2008).

15. Texas and Washington also enacted legislation specifically protecting biometric privacy, though neither includes a private right of action. Quinn Emanuel Urquhart & Sullivan, LLP, *June 2019: The Rise of Biometric Laws and Litigation*, JDSUPRA (June 28, 2019), <https://www.jdsupra.com/legalnews/june-2019-the-rise-of-biometrics-laws-82168>. Arkansas, California, and New York all amended statutory definitions of personal information to include biometrics. Zych et al., *State Biometric Privacy Legislation*, *supra* note 11.

16. 740 ILL. COMP. STAT. 14/20 (2021).

17. See Tiffany Cheung, Michael Burshteyn & Camille Framroze, *Privacy Litigation 2020 Year in Review: BIPA Litigation*, MORRISON FOERSTER (Jan. 12, 2021), <https://www.mofo.com/resources/insights/210111-bipa-litigation.html>; Taylor Hatmaker, *Facebook Will Pay \$650 Million to Settle Class Action Suit Centered on Illinois Privacy Law*, TECHCRUNCH (Mar. 1, 2021 1:36 PM), <https://techcrunch.com/2021/03/01/facebook-illinois-class-action-bipa>.

18. *Rosenbach v. Six Flags Entm't Corp.*, 2019 IL 123186, ¶ 40 (2019).

standing requirements. In particular, courts disagreed over whether plaintiffs suffer an injury in fact for the purposes of Article III standing when their only alleged harm is the collection of their biometric data without written notice and consent in violation of BIPA section 15(b).¹⁹ In 2017, the Second Circuit answered this question in the negative.²⁰ The Ninth Circuit concluded the opposite two years later, becoming the first federal appellate court to hold that a procedural violation of BIPA can amount to an injury in fact.²¹ And in 2020, the Seventh Circuit sided with the Ninth,²² settling a robust debate in the Illinois district courts.

Making sense of BIPA's private right of action, particularly as it relates to the statute's notice and consent requirements, is critical for understanding the current landscape of biometric privacy protection—and the enforceability of privacy rights in general—in the United States. First, as evidenced by the Facebook lawsuit, BIPA has had a huge impact on global companies, who now face massive liability for violations of one state's statute. Plaintiffs have brought suit against Google, Amazon, Snapchat, Vimeo, Juul, WeWork, Home Depot, Dr. Pepper, and many other companies, both technology-centered and not.²³ The ACLU also relied on BIPA section 15(b) as the basis for a lawsuit

19. See 740 ILL. COMP. STAT. 14/15(b) (2021) (establishing a written notice and consent regime for the collection of biometric data). Federal cases were brought under other BIPA provisions as well, but this Note focuses on section 15(b).

20. *Santana v. Take-Two Interactive Software, Inc.*, 717 F. App'x 12 (2d Cir. 2017) (summary order).

21. *Patel v. Facebook, Inc.*, 932 F.3d 1264 (9th Cir. 2019), *cert. denied*, 140 S. Ct. 937 (2020) (mem.).

22. *Bryant v. Compass Grp. USA, Inc.*, 958 F.3d 617 (7th Cir. 2020).

23. See Jennifer Lynch & Adam Schwartz, *Victory! Illinois Supreme Court Protects Biometric Privacy*, ELEC. FRONTIER FOUND. (Jan. 25, 2019), <https://www.eff.org/deeplinks/2019/01/victory-illinois-supreme-court-protects-biometric-privacy>; Daniel R. Stoller, *Amazon Says Alexa Biometric Privacy Claims Should be Dismissed*, BLOOMBERG LAW (Dec. 10, 2019, 9:29 AM), <https://news.bloomberglaw.com/privacy-and-data-security/amazon-says-alexa-biometric-privacy-claims-should-be-dismissed>; Maryam Casbarro, *Update from LitLand: Vimeo Faces BIPA Lawsuit*, JDSUPRA (Nov. 8, 2019), <https://www.jdsupra.com/legalnews/update-from-litland-vimeo-faces-bipa-40471>; Chris Burt, *Juul and WeWork Sued Under BIPA for Collecting Customer Biometrics*, BIOMETRICUPDATE (Nov. 8, 2019), <https://www.biometricupdate.com/201911/juul-and-wework-sued-under-bipa-for-collecting-customer-biometrics>; Chris Burt, *BIPA Suit Brought Against Home Depot for Loss Prevention Biometrics*, BIOMETRICUPDATE (Sept. 9, 2019), <https://www.biometricupdate.com/201909/bipa-suit-brought-against-home-depot-for-loss-prevention-biometrics>; Daniel R. Stoller, *Dr. Pepper Employee Biometric Privacy Case Moves to Federal Court*, BLOOMBERG LAW (Nov. 1, 2019, 11:07 AM), <https://news.bloomberglaw.com/privacy-and-data-security/dr-pepper-employee-biometric-privacy-case-moves-to-federal-court>.

against controversial facial recognition company Clearview AI.²⁴ Second, Illinois's experience with BIPA lawsuits will likely influence other states' decisions regarding their own biometric legislation.²⁵ Absent a federal statute,²⁶ states are acting on their own, and in different ways, to grant residents control over biometric data. BIPA is currently serving as an experiment on whether private enforcement of biometric privacy statutes is desirable, both from legal and policy standpoints.²⁷ Finally, the debate over standing to sue for violations of section 15(b)'s notice and consent requirements could inform other spheres of privacy regulation, as notice and consent regimes are considered "[t]he dominant legal and regulatory approach to protecting information privacy."²⁸

This Note explores the question of standing in BIPA litigation involving allegations that companies have collected biometric data without notice and consent in violation of section 15(b). Ultimately, it argues that section 15(b) protects a concrete privacy interest in controlling information about oneself, an interest that courts have long recognized and the Illinois legislature intended to safeguard. This satisfies the Supreme Court's Article III standing requirements for procedural violations as set forth in *Spokeo, Inc. v. Robins*.²⁹ As

24. Complaint at 31–32, *Am. Civil Liberties Union v. Clearview AI, Inc.*, No. 2020-CH-04353 (Ill. Cir. Ct., Cook Cty., Ch. Div. May 28, 2020), https://www.aclu.org/sites/default/files/field_document/2020.05.28_aclu-clearview_complaint_file_stamped.pdf.

25. In January 2021, New York legislators introduced a bill, the Biometric Privacy Act (A.B. 27), that has been called a “carbon copy” of BIPA. Lydia de la Torre, *A New York BIPA in the Making?*, NAT'L L. REV. (Jan. 28, 2021), <https://www.natlawreview.com/article/new-york-bipa-making>.

26. In 2020, a bill introduced in the U.S. Senate proposed a National Biometric Information Privacy Act, modeled after BIPA and containing a private right of action. Joseph J. Lazzarotti, *National Biometric Information Privacy Act, Proposed by Sens. Jeff Merkley and Bernie Sanders*, NAT'L L. REV. (Aug. 5, 2020), <https://www.natlawreview.com/article/national-biometric-information-privacy-act-proposed-sens-jeff-merkley-and-bernie>.

27. The Illinois legislature has itself considered amending BIPA several times in response to the numerous lawsuits brought under the statute. In 2019, a bill was introduced and ultimately abandoned that would have removed BIPA's private right of action. Meghan C. O'Connor, Gary R. Clark & Sarah A. Erdmann, *Illinois Introduces Bills to Amend BIPA Taking Away Private Right of Action and Adding ECGs*, QUARLES & BRADY LLP (Apr. 25, 2019), <https://www.quarles.com/publications/illinois-introduces-bills-to-amend-bipa-taking-away-private-right-of-action-and-adding-ecgs>. And in March 2021, the Illinois House Judiciary Committee introduced a bill that would (among other amendments) allow electronic consent in place of written consent and create a thirty-day “notice and cure” period before a lawsuit could be brought. Gordon Rees Scully Mansukhani, *Is Illinois Moving Away from its Strict BIPA Law?*, LEXOLOGY (Mar. 15, 2021), <https://www.lexology.com/library/detail.aspx?g=27375046-9836-47d3-95de-5f1dadf50144>. This bill was motivated by the concern that massive class actions brought under BIPA could harm small businesses in Illinois. *Id.*

28. Daniel Susser, *Notice After Notice-and-Consent: Why Privacy Disclosures Are Valuable Even If Consent Frameworks Aren't*, 9 J. INFO. POL'Y 148, 148 (2019).

29. 136 S. Ct. 1540 (2016).

such, plaintiffs alleging section 15(b) violations should not be denied standing in federal court for failure to assert a concrete injury.

Part II of this Note describes BIPA's provisions. Part III explains the current standing landscape of section 15(b) cases in state and federal court. Part IV discusses the privacy interests protected by the statute and then contends that violations of BIPA's notice and consent requirements should amount to concrete injuries for the purposes of Article III standing. Finally, this Note concludes by raising questions about the broader implications of the BIPA standing debate.

II. UNDERSTANDING BIPA'S PROVISIONS

BIPA regulates the collection, use, and disclosure of biometric information. The statute defines a “[b]iometric identifier” as “a retina or iris scan, fingerprint, voiceprint, or scan of hand or face geometry.”³⁰ Biological materials and other medical data, such as x-rays and MRIs, do not fall under the statute.³¹ Although BIPA explicitly states that photographs are not biometric identifiers,³² several courts have either assumed or explicitly held that scans of facial geometry captured from photographs do constitute biometric data.³³ The question of whether information amounts to biometric data under the statute does not depend on how the data is captured.³⁴ Finally, BIPA only applies to private entities, not government agencies or judicial employees.³⁵

BIPA contains five restrictions. First, section 15(b) prohibits companies from collecting data without notice and consent. To “collect, capture, purchase, receive through trade, or otherwise obtain” a data subject’s biometric data, a private entity must first (1) give the data subject a written notice explaining that the data is being collected, describing the purpose for collection, and placing a limit on retention; and (2) obtain the subject’s written consent.³⁶ Second, after collection, biometric data must be stored, transmitted, and secured in a commercially reasonable manner.³⁷ These data security

30. 740 ILL. COMP. STAT. 14/10 (2021).

31. *Id.*

32. *Id.*

33. *See, e.g.,* Patel v. Facebook, Inc., 932 F.3d 1264, 1276 (9th Cir. 2019), *cert. denied*, 140 S. Ct. 937 (2020) (mem.); Rivera v. Google, Inc., 366 F. Supp. 3d 998, 1001 (N.D. Ill. 2018), *abrogated on other grounds by* Bryant v. Compass Grp. USA, Inc., 958 F.3d 617 (7th Cir. 2020); Monroy v. Shutterfly, Inc., No. 16 C 10984, 2017 WL 4099846, at *5 (N.D. Ill. Sept. 15, 2017); Norberg v. Shutterfly, Inc., 152 F. Supp. 3d 1103, 1106 (N.D. Ill. 2015).

34. 740 ILL. COMP. STAT. 14/10 (2021).

35. *Id.*

36. *Id.* 14/15(b).

37. *Id.* 14/15(e).

safeguards must be at least as strong as those that the company employs to protect other types of sensitive personal information.³⁸ Third, a company must destroy biometric data once the purpose for collection is satisfied or once three years have passed since the data subject last interacted with the company, whichever comes first.³⁹ The company must also maintain a publicly available, written notice describing its retention and destruction policy.⁴⁰ Fourth, a company cannot disclose biometric data to a third party without the data subject's consent, unless the disclosure "completes a financial transaction" authorized by the subject or is required by law.⁴¹ And finally, BIPA prohibits companies from profiting off of an individual's biometric identifiers, through sale, lease, or otherwise.⁴²

BIPA contains a private right of action, which permits individuals "aggrieved by a violation of [the] Act" to sue a company for violating the statute's provisions.⁴³ Successful plaintiffs can collect the greater of actual damages or liquidated damages of \$1,000 for each negligent violation, and they can collect the greater of actual damages or liquidated damages of \$5,000 for each intentional or reckless violation.⁴⁴ They can also recover reasonable attorneys' fees and obtain appropriate injunctive relief.⁴⁵

III. THE CURRENT STANDING LANDSCAPE FOR BIPA SECTION 15(B) CASES

Although BIPA provides a private right of action for "aggrieved" Illinois residents, courts have grappled with whether violations of the notice and consent requirements in section 15(b) are sufficient to grant plaintiffs standing without further injury.⁴⁶ The debate over standing in Illinois state court, which rested on the statutory interpretation of "aggrieved," was resolved in January 2019, when the state supreme court held that any violation of a right granted

38. *Id.*

39. *Id.* 14/15(a).

40. *Id.*

41. *Id.* 14/15(d).

42. *Id.* 14/15(c).

43. *Id.* 14/20 ("Any person aggrieved by a violation of this Act shall have a right of action in a State circuit court or as a supplemental claim in federal district court against an offending party.").

44. *Id.*

45. *Id.*

46. Compare *Patel v. Facebook, Inc.*, 932 F.3d 1264, 1267–68 (9th Cir. 2019), *cert. denied*, 140 S. Ct. 937 (2020) (mem.), with *Rivera v. Google, Inc.*, 366 F. Supp. 3d 998, 1014 (N.D. Ill. 2018), *abrogated on other grounds by* *Bryant v. Compass Grp. USA, Inc.*, 958 F.3d 617 (7th Cir. 2020)..

under BIPA could confer standing.⁴⁷ But the approach to answering the standing question is different in federal courts, which are limited in their jurisdiction by Article III of the Constitution and can only hear cases brought by plaintiffs who have suffered a concrete “injury in fact.”⁴⁸

This Part describes how state and federal courts have addressed the standing question in section 15(b) lawsuits thus far.

A. STATUTORY STANDING IN ILLINOIS STATE COURT

In 2019, the Illinois Supreme Court decided *Rosenbach v. Six Flags*, which established that plaintiffs can sue in state court for a violation of BIPA’s provisions even if they suffer no additional harm.⁴⁹ The court held, “[A]n individual need not allege some actual injury or adverse effect, beyond violation of his or her rights under [BIPA], in order to qualify as an ‘aggrieved’ person and be entitled to seek liquidated damages and injunctive relief pursuant to the Act.”⁵⁰

Rosenbach involved Six Flags’ practice of collecting and storing the fingerprints of visitors with season passes in order to authenticate their identities and grant them access to the amusement park.⁵¹ The plaintiff, a fourteen-year-old named Alexander Rosenbach, went on a school field trip to Six Flags.⁵² Because his mother had purchased a season pass for him prior to the outing, he was asked to scan his thumbprint upon arrival at the park.⁵³ Neither Rosenbach nor his mother received written notice informing them of the fact that fingerprints would be collected, describing the purpose for doing so, or indicating how long the data would be retained.⁵⁴ Additionally, they had no opportunity to grant or deny consent in writing.⁵⁵ Rosenbach sued under BIPA section 15(b), seeking statutory damages and an injunction requiring Six Flags to adhere to the statute’s requirements.⁵⁶ Six Flags filed a motion to dismiss, arguing that Rosenbach lacked standing to sue because he had not suffered any real or threatened harm.⁵⁷

47. See *Rosenbach v. Six Flags Entm’t Corp.*, 2019 IL 123186 (2019).

48. See *Lujan v. Defs. of Wildlife*, 504 U.S. 555, 560 (1992).

49. 2019 IL 123186, ¶ 40.

50. *Id.*

51. *Id.* ¶ 4.

52. *Id.* ¶ 5.

53. *Id.* ¶¶ 5–6.

54. *Id.* ¶¶ 8–9.

55. *Id.* ¶ 8.

56. *Id.* ¶ 11.

57. *Id.* ¶ 12.

Relying first on principles of statutory construction, the court held that the plain meaning of the statute evinced the legislature’s intent to allow individuals to sue without showing additional injury.⁵⁸ Section 20 states that anyone “aggrieved by a violation” of BIPA’s provisions has a right of action.⁵⁹ Illinois jurisprudence had previously interpreted “aggrieved” to refer to a person who has suffered an invasion of a legal right,⁶⁰ and dictionary definitions were consistent with that understanding.⁶¹ Therefore, the court found that the legislature would have intended “aggrieved” to have this meaning.⁶² And because BIPA did grant individuals a legal right—namely, “a right to privacy in and control over their biometric identifiers,” safeguarded “by requiring notice before collection and giving them the power to say no by withholding consent”—the legislature must have meant to permit lawsuits for invasions of this right without requiring additional harm.⁶³

Second, the court found that procedural violations of BIPA’s right to biometric privacy and control are not “merely ‘technical’ in nature.”⁶⁴ Rather, it wrote, “When a private entity fails to adhere to the statutory procedures, . . . ‘the right of the individual to maintain [his or] her biometric privacy vanishes into thin air. The precise harm the Illinois legislature sought to prevent is then realized.’”⁶⁵ Such a harm “is no mere ‘technicality’” but a “real and significant” injury.⁶⁶ Thus, the court held, plaintiffs suing under BIPA in Illinois state court are not required to allege actual injury beyond a violation of the statute’s provisions.⁶⁷

B. ARTICLE III STANDING IN FEDERAL COURT

While plaintiffs have a clear avenue to bring BIPA lawsuits in Illinois state court, federal courts have split on the question of whether violations of section 15(b) confer Article III standing.

1. *The Article III Standing Requirement for Intangible Harms*

The modern federal standing doctrine has its origins in Article III of the U.S. Constitution, which limits federal courts’ jurisdiction to cases and

58. *See id.* ¶¶ 24–25.

59. 740 ILL. COMP. STAT. 14/20 (2021).

60. *Rosenbach*, 2019 IL 123186, ¶¶ 30–31.

61. *Id.* ¶ 32.

62. *Id.*

63. *Id.* ¶¶ 33, 34 (citing *Patel v. Facebook Inc.*, 290 F. Supp. 3d 948, 953 (N.D. Cal. 2018)).

64. *Id.* ¶ 34.

65. *Id.* (quoting *Patel*, 290 F. Supp. 3d at 954).

66. *Id.*

67. *Id.* ¶ 40.

controversies.⁶⁸ Although the Constitution is silent on the question of who may bring suit in Article III courts, the Supreme Court has held that the Case or Controversy Clause requires plaintiffs to meet an “irreducible constitutional minimum of standing.”⁶⁹ Three elements compose this minimum. First, plaintiffs must be able to show that they “suffered an injury in fact—an invasion of a legally protected interest which is (a) concrete and particularized, and (b) actual or imminent.”⁷⁰ Second, the defendant’s actions must have caused the injury.⁷¹ And third, it must be likely that the injury can be remedied by a favorable judicial outcome.⁷²

Spokeo, Inc. v. Robins made it clear that even intangible injuries can meet Article III’s requirement of concreteness.⁷³ When intangible harm results from the violation of procedural statutory rights, courts must inquire into the legitimacy of the interests that those rights were designed to protect. *Spokeo* involved a violation of the Fair Credit Reporting Act (FCRA), which, among other things, requires that credit reporting agencies follow certain procedures to ensure that the information they put forth about consumers is as accurate as possible.⁷⁴ The Court faced the question of whether the violation of that requirement—an intangible injury—could amount to a concrete harm for the purposes of the Article III analysis. Ultimately, although somewhat opaquely, the Court held that it could. Justice Alito wrote for the majority, “[I]he violation of a procedural right granted by statute can be sufficient in some circumstances to constitute injury in fact. In other words, a plaintiff in such a case need not allege any *additional* harm beyond the one Congress has identified.”⁷⁵ But, “[d]eprivation of a procedural right without some concrete interest that is affected by the deprivation . . . is insufficient to create Article

68. U.S. CONST. art. III, § 2.

69. *Lujan v. Defs. of Wildlife*, 504 U.S. 555, 560 (1992).

70. *Id.* at 560 (internal quotations, footnote, and citations omitted).

71. *Id.*

72. *Id.* at 561.

73. 136 S. Ct. 1540 (2016).

74. *Id.* at 1545. The defendant, Spokeo, was a reporting agency that sold information about consumers, aggregated from across the internet, to prospective employers, romantic partners, and others wishing to learn about a particular individual. *Id.* at 1546. Robins became aware that his profile on Spokeo’s website contained inaccurate information regarding his age, wealth, level of education, and employment status. *Id.* He sued Spokeo for failing to comply with FCRA’s accuracy requirement. *Id.* On appeal from the district court, where Robins’s claim had been dismissed for lack of standing, the Ninth Circuit held that Robins met the injury in fact requirement because his particular statutory rights had been violated. *Id.* at 1544–45. The Supreme Court remanded the case to the Ninth Circuit to consider whether Robins’s injury was concrete as well as particularized. *Id.* at 1545.

75. *Id.* at 1549.

III standing.”⁷⁶ Put simply, the procedural right infringed upon cannot be *merely* procedural; it must protect a concrete interest.⁷⁷

The Court identified two sources that would help determine whether an intangible statutory violation is sufficiently concrete. First, it directed lower courts to consider common law history. An intangible harm that “has a close relationship to a harm that has traditionally been regarded as providing a basis for a lawsuit in English or American courts” is more likely to constitute an injury in fact.⁷⁸ Second, the legislature’s judgment plays an important role, because lawmakers are well-positioned to identify which harms meet the Article III test.⁷⁹ The Court wrote, “Congress may ‘elevat[e] to the status of legally cognizable injuries concrete, *de facto* injuries that were previously inadequate in law.’”⁸⁰ But, the Court also stated, “Congress cannot erase Article III’s standing requirements by statutorily granting the right to sue to a plaintiff who would not otherwise have standing.”⁸¹

Scholars have lamented the muddled analysis in *Spokeo*. Lauren E. Willis, a professor, wrote, “The majority opinion in *Spokeo* reads like a bad law student exam First, it sets forth superficial and facially contradictory statements . . . with no resolution of those conflicts. Second, it never discusses how those rules apply to the facts of this case.”⁸² Some commentators question whether *Spokeo* changed the legal landscape at all,⁸³ and judges disagree about whether the case creates a useful framework to guide the standing analysis.⁸⁴ The Supreme Court, however, has declined to revisit the question, denying

76. *Id.* (quoting *Summers v. Earth Island Inst.*, 555 U.S. 488, 496 (2009)).

77. The majority gave the example of an incorrect zip code as a “bare” procedural violation. *Id.* at 1550. The dissent contrasted this with the harm that Robins experienced, arguing that incorrect information about his family, employment, and financial status “could affect his fortune in the job market” and thus amounted to a sufficiently concrete harm. *Id.* at 1556 (Ginsburg, J., dissenting).

78. *Id.* at 1549 (majority opinion).

79. *Id.*

80. *Id.* (quoting *Lujan v. Defs. of Wildlife*, 504 U.S. 555, 578 (1992)).

81. *Id.* at 1547–48 (quoting *Raines v. Byrd*, 521 U.S. 811, 820 n.3 (1997)).

82. Lauren E. Willis, *Spokeo Misspeaks*, 50 LOY. L. A. L. REV. 233, 238 (2017).

83. *Compare* *Thomas v. FTS USA, LLC*, 193 F. Supp. 3d 623, 629 (E.D. Va. 2016) (“*Spokeo* did not change the basic requirements of standing.”), *with* Felix T. Wu, *How Privacy Distorted Standing Law*, 66 DEPAUL L. REV. 439, 452, 457 (2017) (explaining that *Spokeo* “requires the courts to assess the nature and cognizability of harms in a way that the Supreme Court had not been doing before” and labeling its holding a “doctrinal shift”).

84. See Matthew S. DeLuca, Note, *The Hunt for Privacy Harms After Spokeo*, 86 FORDHAM L. REV. 2439, 2456 (2018).

certiorari in several cases involving the lower courts' varying interpretations of *Spokeo*.⁸⁵

Despite *Spokeo*'s lack of clarity, its language and reasoning suggest a series of questions that courts should ask when assessing whether an intangible injury stemming from a statutory violation is sufficiently concrete to establish standing. First, is the statutory provision at issue procedural or substantive? *Spokeo*'s analysis applies only to procedural requirements; the credit reporting agency in the case had to follow certain steps (i.e., a procedure) to assure reasonable accuracy of information.⁸⁶ But some statutory provisions establish substantive requirements. For example, in *Eichenberger v. ESPN, Inc.*, the Ninth Circuit wrote that a section of the Video Privacy Protection Act outlawing disclosure of a person's video-viewing data "does not describe a procedure that video service providers must follow. Rather, it protects generally a consumer's substantive privacy interest in his or her video-viewing history."⁸⁷ While infringements of both procedural and substantive provisions can result in intangible harms, courts will have an easier time finding a concrete injury when the violated provision is itself substantive.⁸⁸

If the provision is procedural, a second distinction arises: is it a "bare" procedural requirement or one whose violation results in concrete harm?⁸⁹ In other words, does the provision protect a concrete interest or not? This is the question the court grappled with in *Spokeo*, and it is where common law history and legislative intent become relevant. In addition to these factors, *Spokeo* also suggests that sometimes the answer depends on factual context. A violation of the same clause might give rise to Article III standing in some cases but not others. For example, if *Spokeo*'s failure to follow reasonable procedures to ensure accuracy resulted in the posting of an incorrect zip code, Robins would have suffered a "bare procedural violation" and would not have had standing.⁹⁰

85. See Lee J. Plave & John W. Edson, *First Steps in Data Privacy Cases: Article III Standing*, 37 FRANCHISE L.J. 485, 505 (2018).

86. See *Spokeo, Inc. v. Robins*, 136 S. Ct. 1540, 1545 (2016).

87. 876 F.3d 979, 983 (9th Cir. 2017).

88. See *id.* ("Accordingly, every disclosure of an individual's 'personally identifiable information' and video-viewing history offends the interests that the statute protects.").

89. See *Spokeo*, 136 S. Ct. at 1549 ("[D]eprivation of a procedural right without some concrete interest that is affected by the deprivation . . . is insufficient to create Article III standing." (quoting *Summers v. Earth Island Inst.*, 555 U.S. 488, 496 (2009))). But, "the violation of a procedural right granted by statute can be sufficient in some circumstances to constitute injury in fact." *Id.*

90. *Id.* at 1550. The Court also gives an example of a "bare procedural violation" of FCRA's notice requirement, writing, "[E]ven if a consumer reporting agency fails to provide the required notice to a user of the agency's consumer information, that information regardless

But, as the dissent noted and the Ninth Circuit held on remand, failure to verify the accuracy of familial, educational, and financial information implicated Robins's concrete interests through potential effects on his employment prospects, and he therefore had standing to sue.⁹¹

Although the Supreme Court has not directly stated that *Spokeo* applies to violations of state statutes as well as federal ones, several courts have suggested that it does. Two pre-*Spokeo* cases, one in the Seventh Circuit and one in the Ninth, held that a violation of state law can amount to an injury in fact that grants plaintiffs Article III standing.⁹² In 2016, a district court relied on one of those cases in finding that, since “*Spokeo* said nothing about the ability of state legislatures to create rights sufficient to confer Article III standing,” precedent recognizing standing for injuries to state legal rights was controlling.⁹³ And in two cases discussed below, when assessing whether plaintiffs had standing to sue for BIPA violations, the Seventh and Ninth Circuits both unquestioningly applied *Spokeo*'s analysis to the Illinois statute, considering the General Assembly's intent in lieu of Congress's.⁹⁴ The Supreme Court had a chance to review the Ninth Circuit's holding, but denied certiorari.⁹⁵ Of course, the injury claimed as the result of a state statutory violation must still meet the *Spokeo* requirement of concreteness; a bare procedural infringement of a state statute will never be enough to confer standing.⁹⁶

may be entirely accurate” and thus the statutory violation would “result in no harm.” *Id.* Note that some have challenged the assertion that the publication of an incorrect zip code causes no harm. *See, e.g.,* Leading Case, *Class Action Standing: Spokeo, Inc. v. Robins*, 130 HARV. L. REV. 437, 444–45 (2016) (“[E]ven incorrect zip codes assuredly cause harm of *some* degree, since insurance and marketing companies often segment by zip code, and individuals are prone to make generalizations about race, religion, or ethnicity based on where somebody lives.”).

91. *See Spokeo*, 136 S. Ct. at 1556 (Ginsburg, J., dissenting); *Robins v. Spokeo, Inc.*, 867 F.3d 1108, 1116 (9th Cir. 2017) (“[E]ven if Congress determined that inaccurate credit reporting generally causes real harm to consumers, it cannot be the case that every trivial or meaningless inaccuracy does so.”).

92. *See FMC Corp. v. Boesky*, 852 F.2d 981, 993 (7th Cir. 1988) (“Properly pleaded violations of state-created legal rights, therefore, must suffice to satisfy Article III's injury requirement.”); *Cantrell v. City of Long Beach*, 241 F.3d 674, 684 (9th Cir. 2001) (“[S]tate law can create interests that support standing in federal courts.”).

93. *See Matera v. Google Inc.*, No. 15-CV-04062-LHK, 2016 WL 5339806, at *14 (N.D. Cal. Sept. 23, 2016).

94. *See Bryant v. Compass Grp. USA, Inc.*, 958 F.3d 617 (7th Cir. 2020); *Patel v. Facebook, Inc.*, 932 F.3d 1264, 1273–74 (9th Cir. 2019), *cert. denied*, 140 S. Ct. 937 (2020) (mem.).

95. *Facebook, Inc. v. Patel*, 140 S. Ct. 937 (2020) (mem.).

96. *See Khan v. Children's Nat'l Health Sys.*, 188 F. Supp. 3d 524, 534 (D. Md. 2016) (finding “no authority for the proposition that a state legislature . . . through a state statute or cause of action, can manufacture Article III standing for a litigant who has not suffered a concrete injury”).

2. *Concrete Injuries in Section 15(b) Litigation*

Federal courts have held that some, but not all, violations of BIPA's provisions amount to concrete harms for the purposes of Article III standing. For example, several courts have found that unconsented-to disclosure of biometric data in contravention of section 15(d) constitutes an injury in fact.⁹⁷ Some courts have also suggested that failure to comply with BIPA's data security and retention requirements might give rise to standing.⁹⁸ But section 15(b) has proven the most contentious, with district courts and circuit courts alike disagreeing about whether plaintiffs can bring suit for violations of the statute's notice and consent requirements in various circumstances. This Section describes cases from the three circuit courts that have addressed the question—the Second, Seventh, and Ninth⁹⁹—as well as some cases from the lower courts that further illuminate how federal judges have reasoned about standing in section 15(b) litigation.

a) Face Scans in the Second Circuit

In 2017, the Second Circuit held in *Santana v. Take-Two Interactive Software, Inc.*—a nonprecedential opinion—that plaintiffs lacked standing to sue a company for collecting their biometrics in violation of section 15(b).¹⁰⁰ In this case, plaintiffs sued Take-Two, the makers of a basketball-oriented video game that allowed players to create avatars based on their own facial geometries.¹⁰¹ Players were first required to agree to terms and conditions, which read, “Your

97. See, e.g., *Dixon v. Wash. & Jane Smith Cmty.—Beverly*, No. 17 C 8033, 2018 WL 2445292, at *8–9 (N.D. Ill. May 31, 2018) (finding that an employee had standing to sue her employer for storing her biometric data with a third-party vendor without her consent); *Howe v. Speedway LLC*, No. 17-CV-07303, 2018 WL 2445541, at *4 (N.D. Ill. May 31, 2018) (holding that plaintiff did not have standing because there was no indication that defendant “has released, or allowed anyone to disseminate, any of the plaintiff’s personal information in the company’s possession” (quoting *Gubala v. Time Warner Cable, Inc.*, 846 F.3d 909, 912 (7th Cir. 2017))); *McGinnis v. U.S. Cold Storage, Inc.*, 382 F. Supp. 3d 813, 818 (N.D. Ill. 2019) (“To be sure, disclosing a biometric identifier to a third-party might very well constitute a concrete injury to an individual’s privacy.”).

98. See, e.g., *Howe*, 2018 WL 2445541, at *5 (reasoning that BIPA’s data security requirement is a “substantive provision[]” given the purposes of the statute); *Santana v. Take-Two Interactive Software, Inc.*, 717 F. App’x 12, 15–16 (2d Cir. 2017) (summary order) (noting that violations of BIPA’s data security requirement could “raise[] a material risk that [plaintiffs’] biometric data will be improperly accessed by third parties”); *Miller v. Sw. Airlines Co.*, 926 F.3d 898, 902 (7th Cir. 2019) (“The longer data are retained . . . the greater the risk of disclosure . . .”). But see *Bryant*, 958 F.3d at 626 (holding that a company’s failure to publicly post a biometric data retention schedule in violation of section 15(a) did not confer standing).

99. See *Santana*, 717 F. App’x at 12; *Miller*, 926 F.3d at 898; *Bryant*, 958 F.3d at 617; *Patel*, 932 F.3d at 1264.

100. 717 F. App’x at 17.

101. *Id.* at 13.

face scan will be visible to you and others you play with and may be recorded or screen captured during gameplay. By proceeding you agree and consent to such uses and other uses pursuant to the End User License Agreement.”¹⁰² Then, players had to spend fifteen minutes rotating their heads in front of a camera to generate their avatar.¹⁰³ The plaintiffs alleged that Take-Two violated five of BIPA’s provisions, including section 15(b), by collecting biometrics “without their informed consent.”¹⁰⁴

The Second Circuit did not conduct a thorough *Spokeo* analysis to assess whether section 15(b)’s procedural notice and consent requirements protected a concrete interest. The court concluded that this question was not at issue because the plaintiffs had conceded that “BIPA [was] implicated only if their biometric data [was] collected or disseminated without their authorization or if a procedural violation create[d] a material risk of such an outcome.”¹⁰⁵ The court ultimately rejected the plaintiffs’ argument that Take-Two had collected their biometrics without authorization.¹⁰⁶ Instead, the court found that the software had clearly informed the plaintiffs that the game required face scans,¹⁰⁷ and that any reasonable person would have known that the cameras were conducting such a scan.¹⁰⁸ Ultimately, plaintiffs were not harmed by the lack of opportunity to give written consent. Moreover, Take-Two’s failure to provide notice describing how long the company would store biometric data did not “present[] a material risk that [plaintiffs’] biometric data [would] be misused or disclosed.”¹⁰⁹ As such, the plaintiffs lacked Article III standing.¹¹⁰

b) Biometric Photograph Tagging in the Ninth Circuit and the Northern District of Illinois

In 2019, the Ninth Circuit became the first federal appellate court to hold that plaintiffs had standing to sue a company for failure to comply with section 15(b)’s notice and consent requirements.¹¹¹ *Patel v. Facebook, Inc.* involved Facebook’s Tag Suggestions feature: when users upload photos that contain

102. *Id.* at 13–14.

103. *Id.*

104. *Id.* at 14.

105. *Id.* at 15.

106. *Id.*

107. *Id.* at 15 (holding that the phrasing of the terms and conditions was “sufficient to meet BIPA’s mandates under the circumstances here. . . . [T]o the extent that Take-Two departed from BIPA’s requirements, it only did so insofar as it omitted the term, ‘geometry.’”).

108. *Id.* at 15–16.

109. *Id.* at 16.

110. *Id.* at 17.

111. *See Patel v. Facebook, Inc.*, 932 F.3d 1264, 1275 (9th Cir. 2019), *cert. denied*, 140 S. Ct. 937 (2020) (mem.).

individuals' faces, the application analyzes various geometric points and compares them to a database of face templates.¹¹² If it finds a match, the application asks the user if they would like to tag that person.¹¹³ Plaintiffs in this case alleged that Facebook violated BIPA's section 15(b) requirements by failing to obtain a written release for the collection of their biometric identifiers, and Facebook sought dismissal for lack of standing.¹¹⁴

The Ninth Circuit's analysis closely followed the framework laid out in *Spokeo*. The court first observed, "Privacy rights have long been regarded 'as providing a basis for a lawsuit in English or American courts.'" ¹¹⁵ It discussed the myriad ways in which the law has recognized privacy, from Samuel Warren and Louis Brandeis's pioneering law review article to William Prosser's privacy torts to First and Fourth Amendment jurisprudence.¹¹⁶ The court concluded that, like these historical rights, BIPA protected the interest in controlling information about oneself.¹¹⁷ Second, the court found that the Illinois legislature meant for the statute's procedural rights to protect concrete interests, basing much of its reasoning on the Illinois Supreme Court's decision in *Rosenbach*.¹¹⁸ Because Facebook's Tag Suggestions feature infringed upon the right of individuals to control their biometric information—a right that was not merely procedural and which the Illinois General Assembly intended to protect—Facebook's actions resulted in a concrete injury that gave plaintiffs standing to sue.¹¹⁹

But in *Rivera v. Google, Inc.*, a case similar to *Patel* that also involved the tagging of photographs, the Northern District of Illinois held that plaintiffs lacked standing to sue Google for scanning their facial geometries without notice or consent.¹²⁰ Although *Rivera*'s conclusion as to standing for section 15(b) violations was abrogated in part by the Seventh Circuit's decision in

112. *Id.* at 1268.

113. *Id.*

114. *Id.* at 1268–69, 1274.

115. *Id.* at 1271 (quoting *Spokeo, Inc. v. Robins*, 136 S. Ct. 1540, 1549 (2016)).

116. *Id.* at 1271–72.

117. *Id.* at 1273.

118. *See id.* at 1273–74.

119. *Id.* at 1274.

120. *Rivera v. Google, Inc.*, 366 F. Supp. 3d 998, 1014 (N.D. Ill. 2018), *abrogated on other grounds by* *Bryant v. Compass Grp. USA, Inc.*, 958 F.3d 617 (7th Cir. 2020). Rather than creating tag suggestions, Google Photos creates "face groups," which allow users to group photos of certain individuals together. *Id.* at 1001–02. One of the named plaintiffs in this case was a Google user who had uploaded photos of himself; those photos were then grouped together based on analysis of his facial geometry and associated with his user profile. *Id.* at 1002. The second named plaintiff was *not* a Google user; instead, her friend uploaded photos of her to the application and labeled the resulting group with her name. *Id.*

Bryant v. Compass Group USA, Inc., discussed below, it presents a useful comparison to *Patel* and illustrates how different courts have applied *Spokeo* to questions of standing under section 15(b).

Under *Spokeo*'s history prong, the *Rivera* court examined two possible common law analogues: the privacy torts of intrusion upon seclusion and appropriation of likeness.¹²¹ It held that a violation of section 15(b) was not sufficiently similar to either.¹²² The alleged injury could not compare to intrusion upon seclusion because the information involved was not private; people expose their faces to others constantly.¹²³ And since Google did not use plaintiffs' face scans for a commercial purpose, the second tort, appropriation of likeness, was also a poor fit.¹²⁴

The court also found that the legislature's intent did not support a finding of injury in fact. Although the General Assembly emphasized that the permanency of biometric information heightened the risk of identity theft, the court did not believe that every case would present a sufficient risk of disclosure for this potential harm to amount to a concrete injury.¹²⁵ The court noted, "[I]here is no legislative finding that explains why the absence of consent gives rise to an injury that is *independent* of the risk of identity theft."¹²⁶ While recognizing that the case "presented close legal questions," the district court in *Rivera* ultimately concluded that the violation of section 15(b) did not constitute a concrete injury for the purposes of Article III standing.¹²⁷

c) Fingerprint Scans in the Seventh Circuit and the Northern District of Illinois

The Seventh Circuit has decided two section 15(b) standing cases. Both involved the collection of fingerprints, but the holdings are distinct. While *Miller v. Southwest Airlines Co.*¹²⁸ is of limited applicability due to unique facts, *Bryant v. Compass Group USA, Inc.*,¹²⁹ opened the federal courts to a much wider range of BIPA cases. *Bryant* also overturned a number of Illinois district court cases involving fingerprint collection and section 15(b) standing.

121. *Id.* at 1011–14.

122. *Id.* at 1013–14. The court acknowledged that *Spokeo* does not require a harm to "square on all fours with a common law privacy tort," but found that the relationship must be sufficiently close. *Id.* at 1011.

123. *Id.* at 1012. The court was not persuaded by plaintiffs' argument that their facial biometrics were private, even if their faces themselves were not. *Id.*

124. *Id.* at 1013–14.

125. *Id.* at 1010.

126. *Id.*

127. *Id.* at 1014.

128. 926 F.3d 898 (7th Cir. 2019).

129. 958 F.3d 617 (7th Cir. 2020).

In *Miller*, the Seventh Circuit held that a class of union members had standing to sue their employers for collecting biometrics without written notice and consent.¹³⁰ The plaintiffs worked for airlines that required them to scan their fingerprints when clocking in and out of shifts.¹³¹ As union members, the plaintiffs had the right to bargain over “a material change in [their] terms and conditions of employment,” and “there [could] be no doubt that how workers clock in and out is a proper subject of negotiation between unions and employers.”¹³² In other words, because they were not given the opportunity to consent to biometric data collection, the plaintiffs were denied their right to negotiate an agreement surrounding the collection policy (e.g., their unions could have refused to agree to the policy or pressured employers to raise wages in exchange for its implementation).¹³³ The Seventh Circuit held that denial of the opportunity to negotiate a benefit was a concrete injury.¹³⁴

The Seventh Circuit also found that a non-union plaintiff had standing to sue for a violation of section 15(b) in *Bryant v. Compass Group USA, Inc.*¹³⁵ In this case, the plaintiff’s employer, Compass, had installed vending machines whose products could only be purchased by fingerprint.¹³⁶ During orientation, Compass instructed the plaintiff and her colleagues to scan their fingerprints into the machine and link a form of payment to create an account.¹³⁷ In contravention of section 15(b), Compass failed to (1) give the employees written notice explaining that their biometrics were being collected, describing the purpose for collection, and detailing the length of storage; and (2) obtain their written permission.¹³⁸ The plaintiff argued that these violations “denied her the ability to give informed written consent” and resulted “in the loss of the right to control [her] biometric identifiers and information.”¹³⁹

Like the Ninth Circuit in *Patel*, the Seventh Circuit used *Spokeo* to guide its standing analysis.¹⁴⁰ Interestingly, however, the court relied primarily on a concurrence by Justice Thomas, who reasoned that the injury in fact analysis should distinguish between private rights (e.g., trespass) and public rights (e.g.,

130. *Miller*, 926 F.3d at 901, 905.

131. *Id.* at 901.

132. *Id.* at 902–03.

133. *Id.*

134. *Id.*

135. 958 F.3d 617, 619 (7th Cir. 2020).

136. *Id.*

137. *Id.*

138. *Id.*

139. *Id.* at 620.

140. *See id.* at 623 (“Our starting point is *Spokeo* itself, which provides substantial guidance about cases alleging the kind of intangible harm to personal interests that Bryant asserts.”); *Patel v. Facebook, Inc.*, 932 F.3d 1264 (9th Cir. 2019), *cert. denied*, 140 S. Ct. 937 (2020) (mem.).

failure to comply with regulatory laws).¹⁴¹ In *Bryant*, the Seventh Circuit concluded that the unlawful collection of Bryant's fingerprints was a violation of a private right and was "enough to show injury-in-fact without further tangible consequences."¹⁴² Thus, the court wrote, "This was no bare procedural violation; it was an invasion of her private domain, much like an act of trespass would be."¹⁴³

The court also analyzed the plaintiff's case through the lens of its precedent governing informational injuries—cases where parties fail to comply with statutes that require the disclosure of certain information (e.g., information about public candidates, agency activities, consumer reports, etc.) to enable the recipient to make decisions based on that information.¹⁴⁴ In the Seventh Circuit, "[t]he injury inflicted by nondisclosure is concrete [for the purposes of Article III standing] if the plaintiff establishes that the withholding impaired her ability to use the information in a way the statute envisioned."¹⁴⁵ Here, Compass's failure to disclose information regarding the collection, purpose, and storage of the plaintiff's fingerprints "deprived her of the ability to give the *informed* consent section 15(b) mandates."¹⁴⁶ Because "the informed-consent regime laid out in section 15(b) is the heart of BIPA," the plaintiff's injury was concrete.¹⁴⁷

Bryant abrogated several cases from the Northern District of Illinois that had denied plaintiffs standing to sue private parties for scanning their fingerprints without notice and consent.¹⁴⁸ For example, in *Howe v. Speedway LLC*, the district court held that a plaintiff could not sue his employer for such a violation of section 15(b).¹⁴⁹ The court found that the goal of BIPA was not

141. *Id.* at 624; *Spokeo, Inc. v. Robins*, 136 S. Ct. 1540, 1550–51 (2016) (Thomas, J., concurring) ("Common-law courts more readily entertained suits from private plaintiffs who alleged a violation of their own rights, in contrast to private plaintiffs who asserted claims vindicating public rights.").

142. *Bryant*, 958 F.3d at 624.

143. *Id.* Note that the plaintiff also sued Compass for failing to make publicly available a written data retention schedule, as required by section 15(a); the Seventh Circuit held that this violation did not result in a concrete injury for Article III standing purposes. *Id.* at 626.

144. *Id.*

145. *Id.*

146. *Id.* at 626.

147. *Id.* ("[BIPA's] purpose is to ensure that consumers understand, before providing their biometric data, how that information will be used, who will have access to it, and for how long it will be retained.").

148. *See, e.g., Howe v. Speedway LLC*, No. 17-CV-07303, 2018 WL 2445541 (N.D. Ill. May 31, 2018); *McGinnis v. U.S. Cold Storage, Inc.*, 382 F. Supp. 3d 813 (N.D. Ill. 2019); *McCullough v. Smarte Carte, Inc.*, No. 16 C 03777, 2016 WL 4077108 (N.D. Ill. Aug. 1, 2016); *Colon v. Dynacast, LLC*, No. 19-CV-4561, 2019 WL 5536834, at *4 (N.D. Ill. Oct. 17, 2019).

149. *Howe*, 2018 WL 2445541, at *7.

to grant individuals a right to informed consent, but to protect and secure their biometric data.¹⁵⁰ It based this conclusion on legislative findings that the permanency of biometrics heightened the risk of identity theft, and that regulation would promote “public welfare, security, and safety.”¹⁵¹ Consequently, the court concluded, BIPA’s substantive provisions were those mandating reasonably secure data storage and prohibiting the sale or unauthorized disclosure of biometric information.¹⁵² The notice and consent requirements “operate[d] in support of the data protection goal of the statute” but did not safeguard a concrete right in and of themselves¹⁵³—at least in contexts like fingerprint scanning where plaintiffs had some awareness that their biometrics were being collected.¹⁵⁴

In *Colon v. Dynacast, LLC*, the Northern District of Illinois noted that it had never found standing for a violation of section 15(b) in cases involving the scanning of fingerprints.¹⁵⁵ The reasoning in this case was similar to the Second Circuit’s reasoning in *Santana*. The court wrote that in fingerprint collection cases, “any reasonable person would have known that the respective defendants were collecting, storing, and using biometric data.”¹⁵⁶ Like these earlier cases, “the only purported ‘violation of privacy’ ” the plaintiff in *Colon* could assert “was the failure to explain *in writing* that biometric data was being collected—something that would have been obvious to any employee subject to a fingerprint or hand-scan.”¹⁵⁷ This obviousness led the court to find that any harm resulting from such collection was negligible and failed to amount to an Article III injury in fact.¹⁵⁸

While *Howe* and *Colon* provide interesting arguments against the concreteness of section 15(b) injuries, *Bryant* opens the door for plaintiffs in fingerprint collection cases—and cases involving the collection of biometrics in other contexts—to bring suit in the Seventh Circuit and Illinois district courts for violations of BIPA’s notice and consent requirements.

150. *Id.* at *5.

151. *Id.* (quoting 740 ILL. COMP. STAT. 14/5(g) (2020)).

152. *Id.*

153. *Id.* (quoting *Vigil v. Take-Two Interactive Software, Inc.*, 235 F. Supp. 3d 499, 513 (S.D.N.Y.), *vacated in part sub nom. Santana v. Take-Two Interactive Software, Inc.*, 717 F. App’x 12 (2d Cir. 2017)).

154. *Id.* (distinguishing this case from those involving the nonobvious collection of biometrics through the scanning of photographs posted online, including *Patel* and *Rivera*).

155. No. 19-CV-4561, 2019 WL 5536834, at *4 (N.D. Ill. Oct. 17, 2019).

156. *Id.*

157. *Id.* at *5.

158. *Id.*

IV. THE CONCRETENESS OF NOTICE AND CONSENT

This Part presents a theoretical assessment of whether BIPA's notice and consent requirements protect concrete privacy interests, the invasion of which should give rise to Article III standing. It focuses on section 15(b), as opposed to other provisions, for two reasons. First, as just discussed, much of the controversy surrounding BIPA's private right of action has stemmed from disagreements over whether data collection without notice and consent is a sufficiently concrete injury. Second, informed consent provisions are a core element of modern privacy protection, both in the United States and abroad.¹⁵⁹ Resolving the question of whether BIPA's notice and consent requirements are merely procedural, or whether they protect concrete rights in certain contexts, informs the broader conversation surrounding Article III standing in privacy cases.

This Part raises but does not resolve two important questions. First, is notice and consent a good regime for protecting privacy? Since the Organisation for Economic Co-operation and Development first articulated its Privacy Guidelines in the 1980s (also known as Fair Information Practice Principles, or FIPPs), numerous privacy laws around the globe have employed notice and consent as key safeguards.¹⁶⁰ But some scholars argue that such provisions do little to protect privacy, as they are often hidden in lengthy privacy policies, written in a manner that is difficult to understand, and offer a choice only between consenting to data collection or not using a service at all.¹⁶¹ While the question of notice and consent's effectiveness is hugely important, this Note assumes that BIPA section 15(b) is an important mechanism for securing biometric privacy.

Second, should notice and consent be viewed as one provision with two separate requirements, or two distinct provisions whose violation must be assessed differently within the standing analysis? This Note primarily treats them as one, although it does discuss how violations of each requirement

159. Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1880 (2013).

160. See ORGANISATION FOR ECON. CO-OPERATION & DEV., THE OECD PRIVACY FRAMEWORK 75 (2013); Memorandum regarding The Fair Information Practice Principles from Hugo Teufel III, Chief Privacy Officer, Dep't of Homeland Sec. (Dec. 29, 2008); Notice of Privacy Practices for Protected Health Information, 45 C.F.R. § 164.520 (2019); Privacy of Consumer Financial Information, 16 C.F.R. § 313 (2020); Council Regulation (EU) 2016/679, of the European Parliament and the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1 at art. 7, recital 32.

161. See, e.g., Susser, *supra* note 28, at 43–47.

might independently undermine the right to control one's personal information. But the two could easily be seen as protecting distinct interests. Daniel Susser writes that there can be no meaningful consent without notice, because informed consent requires an understanding of what one is agreeing to; however, he argues that notice does more than offer procedural support for the substantive goals of consent.¹⁶² Rather, notice serves normatively important objectives on its own, such as raising user awareness regarding data collection and encouraging companies to develop articulable privacy policies.¹⁶³ Susser claims that critiques about the value of notice and consent (e.g., the impossibility of truly informed consent and lack of meaningful choice) are more properly directed at consent alone.¹⁶⁴ The severability of the two provisions is, again, a deeply interesting question that lies mostly outside the scope of this Note.

This Part asserts that plaintiffs suffer a concrete harm when private entities collect biometric data without notice and consent. Section IV.A describes how BIPA reflects a widely recognized conceptualization of privacy—control theory—which acknowledges that notice and consent protect substantive rights. Section IV.B argues that alleged violations of section 15(b) can quite easily be found concrete under the *Spokeo* analysis. Finally, Section IV.C explains that a finding of concrete harm for these violations serves the standing doctrine's underlying goal of preserving the separation of powers.

A. CONCEPTUALIZING BIOMETRIC PRIVACY THROUGH THE LENS OF CONTROL THEORY

Because privacy is a nebulous term involving a multitude of legal interests, a clearer understanding of the interests that BIPA's notice and consent requirements protect is necessary to determine whether violating section 15(b) infringes upon a concrete right.

Philosopher Judith Jarvis Thomson has remarked, "Perhaps the most striking thing about the right to privacy is that nobody seems to have any very clear idea what it is."¹⁶⁵ A diverse collection of relatively unrelated laws invoke the concept of privacy, ranging from the Fourth Amendment right to be secure from government intrusion in one's own home¹⁶⁶ to the parental right to control online collection of a child's personal information.¹⁶⁷ Indeed, many academics agree that privacy's indeterminacy stems from the fact that it "seems

162. *See id.* at 52–56.

163. *Id.*

164. *Id.* at 43–47.

165. Judith Jarvis Thomson, *The Right to Privacy*, 4 PHIL. & PUB. AFF. 295, 295 (1975).

166. U.S. CONST. amend. IV.

167. Children's Online Privacy Protection Act, 15 U.S.C. §§ 6501–6506 (2018).

to be about everything, and therefore it appears to be nothing.”¹⁶⁸ Recognizing that such an amorphous right is difficult to litigate and regulate effectively,¹⁶⁹ scholars have endeavored for decades to understand privacy’s myriad meanings in the law.¹⁷⁰ Efforts to classify privacy harms have produced several prominent conceptualizations—i.e., ideas about what privacy actually is—which justify, though do not perfectly mirror, the rights to privacy afforded at law.¹⁷¹ This Section asserts that BIPA’s notice and consent provision reflects the widely accepted “control theory” of privacy. It discusses both the theory’s conceptualization of why privacy is important as well as the legal recognitions of privacy reflecting that conceptualization.¹⁷²

1. *Foundations of Control Theory*

Early control theorist Alan Westin wrote, “Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others.”¹⁷³ While some scholars suggest that privacy-as-control reflects a liberal focus on the individual as an entity entitled to make his or her own choices,¹⁷⁴ others argue that it can only be understood in the context of societal relations.¹⁷⁵ A person’s ability to control the information known about her by others protects her ability to create bonds characterized by essential human values like trust, respect, and love.¹⁷⁶ That is to say, control theory is not about secrecy; it does not presume that people have privacy only in that which they choose not to convey to others at all.¹⁷⁷ Instead, it recognizes that people may wish to share

168. Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 479 (2006).

169. *Id.* at 480.

170. Yvonne F. Lindgren, *Personal Autonomy: Towards a New Taxonomy for Privacy Law*, 31 WOMEN’S RTS. L. REP. 447, 450 (2010).

171. Hyman Gross, *The Concept of Privacy*, 42 N.Y.U. L. REV. 34, 36 (1967) (“The law does not determine what privacy is, but only what situations of privacy will be afforded legal protection Privacy, no less than good reputation or physical safety, is a creature of life in a human community and not the contrivance of a legal system concerned with its protection.”).

172. BIPA arguably reflects other theories of privacy as well as control (e.g., personhood). But the discussion in this Note is limited to control theory, which strongly supports the argument that BIPA protects concrete interests and which several courts clearly draw upon in their analyses. *See, e.g.*, *Rosenbach v. Six Flags Entm’t Corp.*, 2019 IL 123186; *Patel v. Facebook, Inc.*, 932 F.3d 1264 (9th Cir. 2019), *cert. denied*, 140 S. Ct. 937 (2020) (mem.).

173. ALAN F. WESTIN, *PRIVACY AND FREEDOM* 7 (1967).

174. *See* Anita L. Allen, *Privacy-as-Data Control: Conceptual, Practical, and Moral Limits of the Paradigm*, 32 CONN. L. REV. 861, 862 (2000).

175. *See* Charles Fried, *Privacy*, 77 YALE L.J. 475, 482 (1968) (“To refer . . . to the privacy of a lonely man on a desert island would be to engage in irony. The person who enjoys privacy is able to grant or deny access to others.”).

176. *Id.* at 477.

177. *See id.* at 483.

aspects of themselves in some contexts but not others, and contends that the right to privacy safeguards their ability to make those determinations.¹⁷⁸ Under this theory, privacy is violated when an individual's ability to control personal information is taken away.

2. *Control Theory as Reflected in BIPA Section 15(b)*

Several courts have concluded that BIPA protects the right to control one's biometric information.¹⁷⁹ Indeed, section 15(b) clearly reflects the control theory of privacy.

Notice and consent provisions like the one in BIPA are at the heart of control theory, as they give people power to make educated decisions about when, how, and for what purpose others may gather their data. Control theorists recognize that choices about the amount of information to share, and at what level of detail, are highly contextual.¹⁸⁰ BIPA allows people to decide on a case-by-case basis whether a certain company should have access to a certain category of biometrics; consent to one circumstance does not imply consent to others. For example, someone may allow a bank to collect her fingerprint, knowing that it will only be used to grant her access to her financial records. She may feel uncomfortable using the same fingerprint to pay for groceries at a local supermarket, because the store associates all her past purchases with her fingerprint and recommends certain products when she scans it at checkout. Furthermore, she may choose not to consent to any company's collection, use, or storage of her facial geometry. Section 15(b) empowers her to make these contextual decisions.

Choices about whether or not to consent to a certain type of collection are meaningless without knowing to what exactly one is consenting; thus, notice is required if a statute is to guarantee *informed* consent. In this way, as proposed above, notice and consent can be considered separate but inherently intertwined requirements, and a violation of each can infringe upon control. When a company fails to ask for permission before scanning a person's face, then creates a series of data points describing his facial geometry, stores that data on servers for an undetermined amount of time, and uses it for any

178. *Id.* (giving an example of a person who tells a friend he is ill, but does not wish that friend to witness him actually experiencing the symptoms of that illness); see also WESTIN, *supra* note 173, at 7 (“[E]ach individual is continually engaged in a personal adjustment process in which he balances the desire for privacy with the desire for disclosure and communication of himself to others . . .”).

179. See, e.g., *Rosenbach v. Six Flags Entm't Corp.*, 2019 IL 123186; *Patel v. Facebook, Inc.*, 932 F.3d 1264 (9th Cir. 2019), *cert. denied*, 140 S. Ct. 937 (2020) (mem.); *Bryant v. Compass Grp. USA, Inc.*, 958 F.3d 617 (7th Cir. 2020).

180. Fried, *supra* note 175, at 483; WESTIN, *supra* note 173, at 7.

number of purposes, the company quite clearly takes control over that data away from the subject. But even if the company had sought consent for the original face scan, that consent would be hollow if the subject had no understanding of what would happen to his data after it was collected, as that information might change his decision. Section 15(b)'s required notices explaining purpose, retention, and destruction of data are crucial for informed consent and thus play an important role in preserving individual control.

The fact that control of information mandates informed consent through adequate notice casts doubt on the Second Circuit's conclusion in *Santana* and the Northern District of Illinois's in *Colon* that no concrete injury occurs when an individual is aware that a company is collecting their biometrics. If, like in *Colon*, employees are required to scan their fingerprints to clock in and out of shifts without any idea of how long their employer might store the data or whether it will be used for any other purposes, the employees cannot legitimately be said to have given informed consent.¹⁸¹ Despite the fact that employees know their biometrics are being collected, they are denied the ability to make a decision based on a clear understanding of what will happen to their data, and they thus suffer a loss of control. This idea challenges the notion that infringement in the context of knowing collection is a bare procedural violation and suggests that the notice provision protects a concrete right despite the individual's awareness. Indeed, the Seventh Circuit appropriately held in *Bryant* that Compass's failure to tell the plaintiff about its collection and storage procedures deprived her of the right to give informed consent and thus injured her concretely.¹⁸²

3. *When the Loss of Control Requires Legal Protection*

In conjunction with theories about the meaning and import of privacy, some scholars have aimed to classify specific types of privacy harms to better understand which are afforded legal protection. The most prominent example of such an effort is William Prosser's 1960 synthesis of the four privacy torts.¹⁸³ Nearly five decades later, Daniel Solove published a taxonomy of privacy that looked beyond tort law to constitutional doctrines, evidentiary privileges, and federal and state statutory protections, ultimately identifying sixteen privacy harms within four categories.¹⁸⁴ Several of the harms discussed in Solove's taxonomy serve as strong examples of how the loss of control over personal

181. See *Colon v. Dynacast, LLC*, No. 19-CV-4561, 2019 WL 5536834 (N.D. Ill. Oct. 17, 2019).

182. See *Bryant*, 958 F.3d at 626.

183. See William L. Prosser, *Privacy*, 48 CALIF. L. REV. 383 (1960).

184. See Solove, *A Taxonomy of Privacy*, *supra* note 168, at 482–83.

information can cause injury; indeed, some early control theorists wrote about the same specific harms within their conceptualizations.¹⁸⁵ Many of these harms have also been acknowledged in the courts.

This Section describes two such harms—identification (and the correlative interest in anonymity) and aggregation—both of which have received legal recognition, and both of which are implicated in BIPA section 15(b).

a) Identification

Identification is the harm that results from a person’s loss of control regarding information about who they are. The inverse of identification is anonymity, which Westin calls a “basic state[] of individual privacy [that] occurs when the individual is in public places or performing public acts but still seeks, and finds, freedom from identification and surveillance.”¹⁸⁶ At its core, Westin says, the value of anonymity is freedom—it allows a person to move about “open spaces and public arenas” without having to conform to societal roles and rules.¹⁸⁷ Anonymity allows people to control whether information about what they do in public will be linked back to them in a way that could compromise their social relationships.

Anonymity has received a significant amount of legal protection in certain contexts. In *McIntyre v. Ohio Elections Commission*, the Supreme Court determined that privacy encompasses a person’s right to speak and associate without being forced to identify oneself.¹⁸⁸ The Court wrote, “Anonymity is a shield from the tyranny of the majority. . . . [I]t protect[s] unpopular individuals from retaliation . . . at the hand of an intolerant society.”¹⁸⁹ Similarly, the Court observed in *NAACP v. Alabama* that “[i]nviolability of privacy in group association may in many circumstances be indispensable to preservation of freedom of association, particularly where a group espouses dissident beliefs.”¹⁹⁰ The Court also noted the troubling history of identification in the context of religion.¹⁹¹

Without the requirements of notice and consent, many biometric applications could force unwanted identification in spaces that were previously

185. See WESTIN, *supra* note 173, at 31 (referring to anonymity as a “state of privacy”).

186. *Id.*

187. *Id.*

188. 514 U.S. 334 (1995).

189. *Id.* at 357.

190. 357 U.S. 449, 462 (1958).

191. *Id.* (quoting *Am. Comm’ns Ass’n v. Douds*, 339 U.S. 382, 402 (1950) (“A requirement that adherents of particular religious faiths or political parties wear identifying arm-bands, for example, is obviously of this nature,” i.e., obviously interferes with freedom of assembly.)

anonymous, and, depending on the context, could chill freedoms of speech and association.¹⁹² For example, Nest doorbells have the capacity to capture and analyze face scans of people walking along public streets.¹⁹³ If every homeowner in a neighborhood purchased a Nest, people might be deterred from exercising their right to protest in that neighborhood for fear that they would be identified and suffer consequent reprisal. Because of BIPA, Nests sold in Illinois are not equipped with facial recognition (likely because providing the required notice and obtaining consent from every individual who approached a doorbell would be impossible).¹⁹⁴ The provision has thus successfully protected Illinoisans' anonymity. Similarly, as discussed in Part I, some churches use surveillance cameras equipped with facial recognition technology to identify parishioners.¹⁹⁵ Given the right to anonymous association, combined with the great deference religion is given in the United States, obtaining biometrics without consent in this circumstance certainly infringes upon privacy.

Apart from surveillance, Solove argues that de-anonymization can also be harmful in the information privacy context because it “attaches informational baggage to people” and thus removes their control over what others are able to learn about them in a particular circumstance.¹⁹⁶ In this view, biometric identification can result in harm by linking someone's identity to other types of information.¹⁹⁷ Countless biometric applications demonstrate this effect. For example, bank verification can attach an individual's entire financial record to their thumbprint. Employer attendance software connects the same thumbprint with habits of truancy or tardiness. Shoplifter identification systems associate visitors to retail stores with potentially damning information about their past behaviors. And online applications that recognize a person in an uploaded photograph could scour the internet and link that person to a decades-old mugshot. While some of these applications appear more innocuous than others, all can be said to cause harm in that they “inhibit

192. Not all biometric applications threaten anonymity: some are utilized in environments where anonymity is not an option, such as employers' use of thumbprints to track time.

193. Gibbs, *supra* note 4.

194. See Amy Korte, *Privacy Law Prevents Illinoisans from Using Google App's Selfie Art Feature*, ILL. POLICY (Jan. 23, 2018), <https://www.illinoispolicy.org/privacy-law-prevents-illinoisans-from-using-google-apps-selfie-art-feature>.

195. See *Face Recognition App – The Churchix App*, *supra* note 2.

196. Solove, *A Taxonomy of Privacy*, *supra* note 168, at 513 (describing a French case where a transgender person was unable to change her gender on identifying documents).

197. *Id.*

people's ability to change and can prevent their self-development by tying them to a past from which they want to escape."¹⁹⁸

b) Aggregation

The problem of informational baggage is especially pronounced when applications aggregate many different types of data from many different sources. Solove classifies aggregation as a harm distinct from identification, but recognizes that they are often intertwined.¹⁹⁹ Aggregation can generate highly detailed insights that data subjects could not have envisioned when each individual data point was collected, and identification links those insights to a specific person.²⁰⁰

Solove recognizes aggregation as a harm for two reasons. First, it causes a dignitary injury by violating people's expectations that there will be "certain limits on what is known about them and on what others will find out."²⁰¹ The Supreme Court voiced a similar concern in *Carpenter v. United States*, a case involving the collection of location data from a cellphone over a period of 127 days.²⁰² The Court found that the aggregation of many individual location data points—each of which could theoretically be observed by police physically tracking a person in public—could cause harm by violating people's presumptions that "law enforcement agents and others would not . . . secretly monitor and catalogue every single movement of an individual[] . . . for a very long period."²⁰³ Aggregation thus infringed upon the defendant's reasonable expectation of privacy. Second, Solove notes that aggregation can cause a power imbalance when compiled data is used to make crucial decisions about whether someone will receive certain benefits, such as loans and mortgages.²⁰⁴

These harms are certainly possible in the context of biometric data collection. Facebook's Tag Suggestions feature, for example, can associate an individual's face template with any photograph uploaded to the site.²⁰⁵ While someone may expect that he will only be tagged in his friends' photographs, any image uploaded by any stranger that unwittingly captures him in the background can theoretically be linked to his profile. Given that hundreds of

198. *Id.* at 514.

199. *Id.*

200. *Id.* at 507, 514 (explaining that aggregation creates a "digital person" and identification "links the digital person directly to a person in realspace").

201. *Id.* at 508.

202. 138 S. Ct. 2206 (2018).

203. *Id.* at 2217 (quoting *United States v. Jones*, 565 U.S. 400, 430 (2012)).

204. Solove, *A Taxonomy of Privacy*, *supra* note 168, at 508.

205. The Ninth Circuit raises this concern in *Patel*. See *Patel v. Facebook, Inc.*, 932 F.3d 1264, 1273 (9th Cir. 2019), *cert. denied*, 140 S. Ct. 937 (2020) (mem.).

millions of photographs are uploaded each day, there is a potential for data aggregation that provides overly detailed insight into what a person does, who he is with, and (through geotagging) where he goes. This may violate that person's expectations for how the Facebook site functions and represents a significant loss of control over the amount and detail of collected information. But providing adequate notice at the collection stage detailing how the data will be used and whether it will be combined with other information can properly orient an individual's expectations regarding the extent of data use. And requiring Facebook to obtain consent gives the individual the power to choose whether the company can create a "digital person"²⁰⁶ based on the aggregation of the individual's data.

B. NOTICE AND CONSENT UNDER *SPOKEO*

A clearer understanding of how section 15(b) implicates the privacy-as-control theory makes it easier to analyze whether a violation of that provision constitutes a concrete injury under *Spokeo*. The examples discussed above already suggest that failure to provide adequate notice and obtain consent can cause real harms, such as loss of control over highly personal information, infringement of the right to speak and associate anonymously (as well as the potential chilling effect on the exercise of those freedoms), inability to disassociate oneself from past actions or contextually unrelated information, and violation of expectations governing the amount of information known about oneself. So, while section 15(b) may be procedural, it protects undeniably concrete interests, and violations merit a finding of injury in fact. This Section demonstrates how the two prongs of the *Spokeo* analysis support that conclusion.

1. History

Spokeo directs courts to "consider whether an alleged intangible harm has a close relationship" to one historically recognized at common law.²⁰⁷

As discussed above, courts throughout history have acknowledged several of the privacy harms that section 15(b) implicates as deserving of legal redress. Some interests are more modern, but others, like that of anonymous speech and association, are First Amendment rights that go back to the country's origins. Moreover, although scholars only began to articulate control theory in the 1960s, they did so in an attempt to extract insight from much older

206. Solove, *A Taxonomy of Privacy*, *supra* note 168, at 508.

207. *Spokeo, Inc. v. Robins*, 136 S. Ct. 1540, 1549 (2016).

common law cases that established legal protections for privacy rights.²⁰⁸ So, it can be said that many of the harms examined in Section IV.A, are sufficiently analogous to historically protected harms.

Crucially, *Spokeo*'s "close relationship" between alleged and common law harm does not have to be an exact match; several courts have found general comparisons to be satisfactory.²⁰⁹ Although some scholars assert that comparing privacy harms to historical analogs is likely to severely diminish standing for privacy cases, since traditional conceptions of privacy do not encompass modern problems,²¹⁰ the broad interpretation of the history prong gives courts sufficient leeway to identify old injuries that are sufficiently similar to new ones.

The Ninth Circuit accurately described some of the specific historical interests that BIPA protects in its analysis of common law privacy harms in *Patel*. The court began by broadly acknowledging that the common law protected "a general right to privacy."²¹¹ It discussed this right in the context of Warren and Brandeis's ideas about privacy, Prosser's four privacy torts, and the intertwined constitutional and common law notions regarding "zones of privacy."²¹² The court also noted recent Fourth Amendment cases, like *United States v. Jones*,²¹³ *Carpenter v. United States*,²¹⁴ and *Riley v. California*,²¹⁵ which found that new technologies introduced significant risks to privacy.²¹⁶ While this initial analysis was somewhat vague, the court eventually clarified how the specific interests BIPA protects compared to traditionally recognized harms. First, it wrote, "[B]oth the common law and the literal understandings of privacy encompass the individual's control of information concerning his or her person."²¹⁷ In other words, the statute reflects control theory's conceptualization of privacy. Second, the court discussed how Facebook's Tag

208. WESTIN, *supra* note 173, at 330–64 (describing how privacy rights in American law evolved in the years between 1790 and the mid-twentieth century).

209. DeLuca, *supra* note 84, at 2463–64.

210. See, e.g., Solove, *A Taxonomy of Privacy*, *supra* note 168, at 564 ("[S]ome of the privacy problems we face today are different in nature, and do not track traditional conceptions of privacy.").

211. *Patel v. Facebook, Inc.*, 932 F.3d 1264, 1271 (9th Cir. 2019), *cert. denied*, 140 S. Ct. 937 (2020) (mem.) (quoting Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 198 (1890)).

212. *Id.* at 1271–72.

213. 565 U.S. 400 (2012).

214. 138 S. Ct. 2206 (2018).

215. 573 U.S. 373 (2014).

216. *Patel*, 932 F.3d at 1272–73.

217. *Id.* at 1273 (quoting U.S. Dep't of Justice v. Reporters Comm. for Freedom of the Press, 489 U.S. 749, 763 (1989)).

Suggestions feature could obtain “encyclopedic” information and associate a user with “hundreds of millions of photos.”²¹⁸ This point reflects concerns about data aggregation. The court then noted that Facebook’s face scans could be used to identify an individual “from a surveillance photo taken on the streets or in an office building.”²¹⁹ Here, we see anxiety about the harms of identification in conjunction with surveillance and how it can lead to the loss of anonymity. The court concluded that these harms “invade[] an individual’s private affairs,”²²⁰ suggesting that the concrete interest in privacy that BIPA protects is sufficiently similar to the common law’s general right to privacy.

On the other hand, the Northern District of Illinois’s overly restrictive analysis in *Rivera* failed to acknowledge any possible connection between common law privacy rights and the theory of privacy-as-control underlying BIPA’s provisions. In *Rivera*, the court examined whether a violation of section 15(b) bore a sufficiently close relationship to one of two specific privacy torts, intrusion upon seclusion or appropriation of likeness.²²¹ It concluded that neither was an apt analogue, because several key elements of the torts were not met in the BIPA case.²²² Although the court recognized that *Spokeo* did not require “[an] alleged injury . . . [to] square on all fours with a common law privacy tort,” it ultimately held that the differences between the statute and the torts were too great.²²³ The *Rivera* court’s analysis was thus flawed, in part because it construed *Spokeo*’s “close relationship” guidance as requiring an excessively strict standard. But beyond that, the court assessed how a violation of notice and consent squared with the *elements* of traditional privacy harms, when it should have compared the *interests* that each protects. After all, the overarching purpose of the *Spokeo* analysis is not to find an adequate historical match for a modern statute involving modern technology—it is to determine whether that modern statute protects interests that can rightly be judged as concrete. Because *Rivera*’s factual context was so similar to *Patel*’s, the Northern District of Illinois should have identified all the harms discussed by the Ninth Circuit and found standing as a result.

2. *Legislative Intent*

Spokeo also held that courts should consider Congress’s judgment and assess whether it meant for the statute to protect a concrete interest. The state

218. *Id.*

219. *Id.*

220. *Id.*

221. *Rivera v. Google, Inc.*, 366 F. Supp. 3d 998, 1011 (N.D. Ill. 2018), *abrogated on other grounds by* *Bryant v. Compass Grp. USA, Inc.*, 958 F.3d 617 (7th Cir. 2020).

222. *Id.* at 1012–14.

223. *Id.* at 1011.

legislature's intent is equally instructive when analyzing whether plaintiffs have standing under a state statute.²²⁴

Federal courts have disagreed over whether the Illinois General Assembly intended for the procedural requirements in section 15(b) to protect an underlying concrete interest. The Northern District of Illinois concluded in *Howe* that the purpose of BIPA is “the protection and security of biometric data”; according to this court, the notice and consent requirements merely “support” those goals and do not in themselves safeguard a concrete interest.²²⁵ The same court came to a similar conclusion in *Rivera*, noting that none of the legislature's findings addressed a harm separate from identity theft.²²⁶ In contrast, the Ninth Circuit, relying heavily on the Illinois Supreme Court's decision in *Rosenbach*, interpreted BIPA as granting a right to control one's personal information and found that notice and consent are integral to preserving true authority over one's data.²²⁷ And in *Bryant*, the Seventh Circuit found that the purpose of the statute was “to ensure that consumers understand, before providing their biometric data, how that information will be used, who will have access to it, and for how long it will be retained.”²²⁸ Informed consent and the corresponding right to withhold consent—which the court called “[a] key part of the right to control biometric information”²²⁹—were at “the heart of BIPA.”²³⁰

The legislature's focus on *proactively* preventing the harms associated with identity theft, as well as its interest in safeguarding the public welfare by regulating data collection, strongly indicates that it intended section 15(b) to protect concrete interests, the violation of which would suffice for standing. After noting the legislative findings regarding identity theft, the Illinois Supreme Court in *Rosenbach* found that BIPA was designed to “head off such problems before they occur,” in part by “imposing safeguards to insure that

224. See *supra* text accompanying notes 92–96.

225. *Howe v. Speedway LLC*, No. 17-CV-07303, 2018 WL 2445541, at *5 (N.D. Ill. May 31, 2018). The court did postulate that, in different factual contexts where plaintiffs were completely unaware that companies were collecting their biometrics, violations of section 15(b) might infringe upon an underlying privacy interest. *Id.* However, the same court seemed to reject that reasoning several months later in *Rivera*, a case where plaintiffs did not know that Google was scanning their photographs to collect their facial geometries. See *Rivera v. Google, Inc.*, 366 F. Supp. 3d 998, 1001–02 (N.D. Ill. 2018), *abrogated on other grounds by Bryant v. Compass Grp. USA, Inc.*, 958 F.3d 617 (7th Cir. 2020).

226. See *Rivera*, 366 F. Supp. 3d at 1010–11.

227. See *Rosenbach v. Six Flags Entm't Corp.*, 2019 IL 123186, ¶ 34 (2019); *Patel v. Facebook, Inc.*, 932 F.3d 1264, 1273 (9th Cir. 2019), *cert. denied*, 140 S. Ct. 937 (2020) (mem.).

228. *Bryant v. Compass Grp. USA, Inc.*, 958 F.3d 617, 626 (7th Cir. 2020).

229. *Id.* at 621.

230. *Id.* at 626.

individuals' . . . rights in their biometric identifiers . . . are properly honored and protected to begin with, before they are or can be compromised."²³¹ *Patel* cited this language to come to the same conclusion.²³² A failure to find standing in section 15(b) cases would allow plaintiffs to bring suit only in situations where their biometric information was already in a heightened state of risk, because a company either failed to store it securely or disclosed it to a third party without consent. It certainly seems that the legislature was interested not only in preventing identity theft itself but also in preventing (unwanted) increased risk as well. As the Seventh Circuit found in *Bryant*, notice and consent enable informed decisions about whether one wishes to assume such a risk.²³³ Thus, when a company violates section 15(b), "the right of the individual to maintain . . . biometric privacy vanishes into thin air [and] [t]he precise harm the Illinois legislature sought to prevent is then realized."²³⁴

Furthermore, the *Howe* court's judgment that section 15(b)'s notice and consent requirements "support" BIPA's goals does not inevitably lead to the conclusion that those provisions do not protect a concrete interest.²³⁵ Notice and consent supply individuals with control in part by enabling them to exercise other rights encompassed by other provisions. For example, without adequate notice that a company is collecting and storing biometric information, how can anyone sue for the improper storage or disclosure of that data? Absent this knowledge, they may not learn that their information is being held insecurely, or held by a third party, until after it is compromised.

In sum, the Seventh Circuit and Ninth Circuit correctly concluded that the Illinois General Assembly intended BIPA section 15(b) to grant a right to control the collection of one's own biometric data through informed consent, and that that control safeguards concrete interests. Because the provision also prevents harms historically recognized in the Constitution and at common law, the *Spokeo* analysis supports a finding that violations of section 15(b) amount to concrete harms for the purposes of the injury in fact analysis.

231. *Rosenbach*, 2019 IL 123186, ¶ 36.

232. *Patel*, 932 F.3d at 1273.

233. *Bryant*, 958 F.3d at 626 ("The judgment of Illinois's General Assembly is that the sensitivity of biometric information and the risk of identity theft or other privacy or economic harm that may result from its dissemination, necessitates that people be given the opportunity to make informed choices about to whom and for what purpose they will relinquish control of that information.").

234. *Rosenbach*, 2019 IL 123186, ¶ 34 (quoting *Patel v. Facebook Inc.*, 290 F. Supp. 3d 948, 954 (N.D. Cal. 2018)).

235. *See Howe v. Speedway LLC*, No. 17-CV-07303, 2018 WL 2445541, at *5 (N.D. Ill. May 31, 2018).

C. BIPA AND THE GOALS OF STANDING

Finally, interpreting section 15(b) as protecting concrete interests in accordance with the legislature's intent serves the standing doctrine's underlying goal of preserving the separation of powers.

The majority in *Spokeo* wrote that the standing doctrine “developed in our case law to ensure that federal courts do not exceed their authority.”²³⁶ Indeed, preserving the separation of powers is one of the most commonly-offered justifications of the doctrine.²³⁷ Justice Antonin Scalia argued in a highly influential article that standing, by requiring plaintiffs to have suffered a concrete and particularized injury, restricts the judiciary to its traditional role of protecting individuals, rather than “prescribing how the other two branches should function.”²³⁸ In Justice Scalia's view, this limit is a good thing, both because courts are inherently undemocratic and because judges are “governed by a body of knowledge that values abstract principle above concrete result.”²³⁹

While standing does place important limits on the courts' ability to answer questions better suited for Congress, stringent enforcement of standing requirements can actually undermine the power of the legislature to define and enforce rights.²⁴⁰ In *The Structure of Standing*, Judge William Fletcher asserts that the balance of powers is likely to tip too far towards the courts when judges are interpreting whether a statutory violation amounts to an injury in fact. He argues, “[T]o limit . . . the power of Congress to create standing . . . is to limit the power of Congress to define and protect against certain kinds of injury that the Court thinks it improper to protect against.”²⁴¹

The increasing number of privacy statutes represents a clear effort by federal and state legislatures to define new injuries arising as a result of technological progress. As previously discussed, some of these injuries share characteristics with the privacy torts and older common law claims. But some represent entirely new harms. In Judge Fletcher's view, a restrictive standing requirement that prevents plaintiffs from vindicating rights the legislature has granted violates the separation of powers. Felix Wu makes a similar argument regarding *Spokeo*'s effect on the standing analysis in privacy cases. He writes:

236. *Spokeo, Inc. v. Robins*, 136 S. Ct. 1540, 1547 (2016).

237. Along with promoting separation of powers, standing also serves the goals of heightened judicial efficiency, improved judicial decision-making, and increased fairness. ERWIN CHERMERINSKY, *FEDERAL JURISDICTION* 56–57 (7th ed. 2016).

238. Antonin Scalia, *The Doctrine of Standing as an Essential Element of the Separation of Powers*, 17 SUFFOLK U. L. REV. 881, 894 (1983).

239. *Id.* at 896.

240. William A. Fletcher, *The Structure of Standing*, 98 YALE L.J. 221, 233 (1988).

241. *Id.*

When courts deny standing in [privacy] cases on the basis of the injuries being insufficiently concrete, they are not deciding whether the cases are ones that concern individual rights, but rather deciding the substantive content of those rights. Far from supporting an appropriate separation of powers, this move amounts to a usurpation of legislative power by the federal judiciary.²⁴²

In enacting BIPA, the Illinois legislature determined that the unauthorized collection of biometric data by a corporate entity represents a significant harm. Because plaintiffs are suing companies for illegally capturing their unique data, rather than for biometric practices more generally, questions brought to the court will often involve individual rights.²⁴³ Federal courts thus are well within their purview to answer these questions. Indeed, refusing to do so based on an overly strict standing requirement would infringe upon legislative power.

V. CONCLUSION

A clear analysis of the privacy theory underlying BIPA's notice and consent requirements, as well as the harms BIPA prevents, supports the conclusion that section 15(b) protects a concrete right. As the federal courts see more and more BIPA litigation involving violations of section 15(b), they should hold that infringements upon the interests it safeguards meet the requirements of injury in fact for the purposes of Article III standing.

This analysis may also have implications for other privacy statutes that mandate proper notice and informed consent for the collection and use of data. As discussed above, many statutes in the United States and elsewhere rely upon these provisions as key elements of privacy protection. A finding of concreteness for violations of section 15(b) does not necessarily suggest the same conclusion for other privacy statutes; perhaps biometric data should be seen as unique, given the heightened identity theft risks resulting from its permanency, as well as other distinctive characteristics.²⁴⁴ Regardless, in analyzing the concreteness of similar privacy harms, courts should make a greater effort to clarify the underlying conceptualizations of privacy that a given statute protects. Doing so will preserve the power of federal and state

242. Wu, *supra* note 83, at 458.

243. *See id.* ("The vast majority of privacy and security cases . . . are indeed ones involving individual rights, not merely broad questions of public interest. Almost invariably, privacy plaintiffs are specific individuals who claim that their own personal information has been mishandled in some way.")

244. Here, the personhood theory of privacy, introduced *supra* note 172, could support an argument that BIPA is unique because the biometric data it protects is inextricably tied to the body and the self, while other types of information like addresses and credit card numbers are less intrinsically personal.

legislatures to enact laws like BIPA to ensure that new technologies do not dilute or eliminate important, long-recognized rights to privacy.

“ILLINOIS BRICK-BREAKER” FOR SALE IN THE APP STORE: *APPLE V. PEPPER* AND THE NEED FOR A NEW ANTITRUST STANDING DOCTRINE

Haley Johnson[†]

I. INTRODUCTION

As the tech economy has boomed through the first two decades of the twenty-first century, a select few firms have come to dominate the space.¹ In response, heightened antitrust scrutiny of big tech companies and major digital platforms is slowly but surely on the rise and will likely continue ramping up in the foreseeable future.² However, existing antitrust doctrine “incompletely capture[s]” the economic structure of these platforms, presenting courts with a unique challenge.³

The inadequacy of existing precedent was on clear display in May of 2019, when the Supreme Court decided *Apple Inc. v. Pepper*.⁴ The case involved Apple’s allegedly anticompetitive behavior in the administration of its App Store, a digital marketplace that is the only place where Apple consumers can purchase apps for their Apple devices. The plaintiffs, a group of iPhone

DOI: <https://doi.org/10.15779/Z38KK94D0N>

© 2020 Haley Johnson.

† J.D., 2021, University of California, Berkeley, School of Law. Appreciation to Professor Talha Syed, Professor Kenneth Bamberger, and Megan McKnelly for their thoughtful comments and suggestions.

1. See Farhad Manjoo, *Tech’s ‘Frightful 5’ Will Dominate Digital Life for Foreseeable Future*, N.Y. TIMES (Jan. 20, 2016), <https://www.nytimes.com/2016/01/21/technology/techs-frightful-5-will-dominate-digital-life-for-foreseeable-future.html>; see also Lina M. Khan, *Amazon’s Antitrust Paradox*, 126 YALE L.J. 710, 712–13 (2016) (noting that as of 2013, Amazon “sold more than its next twelve online competitors combined”).

2. See Brent Kennedy, *Justice Department to Open Broad, New Antitrust Review of Big Tech Companies*, WALL ST. J. (July 23, 2019, 5:34 PM), <https://www.wsj.com/articles/justice-department-to-open-broad-new-antitrust-review-of-big-tech-companies-11563914235>; see also Tony Romm, *Amazon could face heightened antitrust scrutiny under a new agreement between U.S. regulators*, WASH. POST (June 1, 2019), <https://www.washingtonpost.com/technology/2019/06/02/amazon-could-face-heightened-antitrust-scrutiny-under-new-agreement-between-us-regulators/> (reporting that the FTC and the Department of Justice were entering into the type of agreement to divvy up competition oversight of large tech companies that “typically presages more serious antitrust scrutiny”); Cecilia Kang, David Streitfeld & Annie Karni, *Antitrust Troubles Snowball for Tech Giants as Lawmakers Join In*, N.Y. TIMES (June 3, 2019), <https://www.nytimes.com/2019/06/03/technology/facebook-ftc-antitrust.html>.

3. Geoffrey A. Manne & Kristian Stout, *The Evolution of Antitrust Doctrine After Ohio v. Amex and the Apple v. Pepper Decision That Should Have Been*, 98 NEB. L. REV. 425, 440 (2019).

4. 139 S. Ct. 1514 (2019).

owners, alleged that Apple abused its monopoly power over the iPhone apps market in the App Store by charging app developers a thirty-percent commission on each app they sell.⁵ This in turn, the plaintiffs alleged, forced consumers to pay higher prices than they would have in a competitive market.⁶

The App Store is an example of a two-sided marketplace. It acts as an intermediary selling “different products to different groups of consumers.”⁷ To the app developers, it provides a marketplace to sell their apps to iPhone users. On the other side, the App Store sells the apps to the iPhone users. Two-sided markets are by no means a new phenomenon: credit card networks that cater to both merchants on one side and cardholding purchasers on the other are a prime example of a market in which consumers and sellers meet on a platform.⁸ However, two-sided markets are increasingly common in the modern economy, particularly in the tech sector with the rise of digital platforms.

The crucial feature of two-sided marketplaces is that effects on one side of the market impact users on the other side.⁹ Indeed, the demands of platform participants are interdependent.¹⁰ Participants on one side of the marketplace rely on the participation of the users on the other side.¹¹ Take Uber and Lyft as examples. Riders’ demand for the platforms depends on active participation by drivers on the other side, and vice versa.¹² This phenomenon, in which consumers on one side of a platform benefit from increased use by the consumers on the other side, is called an “indirect network effect.”¹³ The rideshare platforms’ attractiveness to riders depends on the riders’ ability to request rides from their specific location on short notice, which requires widespread availability of drivers.¹⁴ Riders will only value the service if they can trust that their trip will be accepted by a driver in a reasonably short time frame,

5. Apple, Inc. v. Pepper, 139 S. Ct. 1514, 1519 (2019).

6. *Id.*

7. ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT, RETHINKING ANTITRUST TOOLS FOR MULTI-SIDED PLATFORMS 10 (2018), available at www.oecd.org/competition/rethinking-antitrust-tools-for-multi-sided-platforms.htm.

8. *See id.*; Michael Katz & Jonathan Sallet, *Multisided Platforms and Antitrust Enforcement*, 127 YALE L.J. 2142, 2143 (2018); Erik Hovenkamp, *Platform Antitrust*, J. CORP. L. 713, 714 (2019).

9. David S. Evans & Richard Schmalensee, *Markets with Two-Sided Platforms*, in 1 ISSUES IN COMPETITION L. & POL’Y 667, 667 (2008).

10. Hovenkamp, *supra* note 8, at 720.

11. *Id.*

12. *See id.*

13. Evans & Schmalensee, *supra* note 9, at 671.

14. *See* Hovenkamp, *supra* note 8, at 720.

and drivers will only find offering their services worthwhile if there is a steady stream of riders taking trips.¹⁵

In the antitrust context, the interdependence of two-sided marketplaces adds a degree of complexity to conventional economic analyses of pricing and competition.¹⁶ Substantive antitrust doctrines have evolved to account for these types of phenomena; for example, by developing the “rule of reason” analysis.¹⁷ However, more procedural doctrines, such as the standing analysis, have been much slower to evolve, leaving courts with anachronistic tools for evaluating modern business structures and practices. This tension came to a head in *Apple v. Pepper*.

In light of the increasing ubiquity of digital platforms and the heightened antitrust scrutiny they continue to attract, it is imperative that courts have the appropriate tools to analyze these types of marketplaces. In this Note, I argue that the *Illinois Brick* indirect purchaser rule, applied in *Apple v. Pepper*, is a poor tool for analyzing antitrust standing in two-sided marketplaces and that the Court missed a critical opportunity to dispense with it and instead adopt a test better suited to the nuances of two-sided marketplaces.¹⁸ I then consider an alternative test that I argue would reach the same correct outcome, but based on reasoning that better suits two-sided marketplaces like the App Store.

In Part II, I provide the necessary background information for considering *Apple v. Pepper*. I first provide an overview of the core goals of antitrust law and the relevant statutes implicated in the decision. I then summarize *Hanover Shoe* and *Illinois Brick*, two antitrust cases that lay the foundation for the indirect purchaser rule at issue in *Apple v. Pepper*. In Part III, I outline the procedural history of *Apple v. Pepper* and summarize the majority and dissenting opinions. In Part IV, I discuss why *Illinois Brick* does not serve two-sided marketplaces. Finally, I evaluate an alternative test courts might use to analyze standing issues and consider how *Apple v. Pepper* might have been decided using this test.

15. *See id.*

16. *Id.* at 719.

17. In the rule of reason analysis, courts evaluate countervailing pro- and anticompetitive effects incrementally through a multi-stage burden-shifting framework. For an overview of the rule of reason analysis and its application in practice, see Hovenkamp, *supra* note 8, at 744.

18. I express no opinion about whether *Illinois Brick*'s indirect purchaser rule remains a useful tool for analyzing other types of marketplaces, such as those with vertical distribution chains.

II. BACKGROUND

A. THE FEDERAL ANTITRUST LAWS & PRIVATE ENFORCEMENT

The federal antitrust laws aim to “maximize consumer welfare” by encouraging competitive behavior between firms while allowing them to reap benefits that may come from “internal or jointly created production efficiencies.”¹⁹ In short, they encourage competition and collaboration and forbid collusion.

The Sherman Act of 1890 and the Clayton Act of 1914 are the primary vehicles for promoting these goals. The Sherman Act sets forth “general prohibitions” on “trade restraints and monopolization.”²⁰ Over time, the Sherman Act has been “interpreted to condemn only *unreasonable* restraints of trade.”²¹ A few decades later, Congress passed the Clayton Act to specifically prohibit certain categories of conduct, including price discrimination, tying, exclusive dealing arrangements, and mergers by acquisition, which would substantially lessen competition in the industry or “tend to create a monopoly.”²²

The Sherman Act and the Clayton Act comprise a two-pronged enforcement regime to promote competition, with private enforcement supplementing “inevitably selective” public enforcement.²³ Indeed, the Clayton Act empowers “any person” injured in their business or property by “anything forbidden in the antitrust laws” to sue and recover “threefold [] damages.”²⁴ The treble damages provision generates a “powerful financial incentive” to both detect violations of the antitrust law and prosecute violators in the court system.²⁵ Without the possibility of increased damages, injured parties may not find it worthwhile to proceed through the complex litigation process.²⁶ Similarly, if damages were fixed at the amount of the overcharge, it would often be profitable for defendant firms to continually violate the

19. Phillip E. Areeda & Herbert Hovenkamp, *Antitrust Law: An Analysis of Antitrust Principles and Their Application* 100a (4th ed. 2013–2018).

20. *Id.* at 301b3.

21. *Id.* (emphasis added); *see also* *Ass’n. Gen. Contractors of Cal., Inc. v. Cal. State Council of Carpenters*, 459 U.S. 519, 531 (1983) (citing *Nat’l Society of Professional Engineers v. U.S.*, 435 U.S. 679, 687–88 (1978)) (noting that “restraint is the very essence of every contract” and if the Sherman Act were read literally, it would “outlaw the entire body of private contract law”).

22. Areeda & Hovenkamp, *supra* note 19, at 301b1.

23. *Id.* at 330b.

24. 15 U.S.C. § 15(a) (1914).

25. Areeda & Hovenkamp, *supra* note 19, at 330b.

26. *See id.*

antitrust laws since, at worst, they would merely have to pay back their ill-gotten gains if they were ever caught.²⁷

Just as courts have developed the substantive aspects of federal antitrust law over time, they have similarly grappled with the procedural issue of who has standing to privately enforce the antitrust laws.

B. CASE LAW: *HANOVER SHOE & ILLINOIS BRICK*

Two particular cases lay the groundwork for the current federal antitrust standing doctrine, which is at the heart of the dispute in *Apple v. Pepper*. In the case, the Court was faced with the question of whether the plaintiffs were “direct purchasers,” and thus had standing to sue under the 1977 Supreme Court decision *Illinois Brick Co. v. Illinois*.²⁸ *Illinois Brick*, in turn, interpreted the applicability of a rule decided in a previous case, *Hanover Shoe, Inc. v. United Shoe Machinery Corp.*²⁹

1. *Hanover Shoe*

In *Hanover Shoe*, the defendant, United, produced shoe-making machinery, which it leased to shoe manufacturers, including the plaintiff Hanover Shoe.³⁰ Hanover Shoe brought suit against United, alleging that United’s monopoly over the shoe-making equipment allowed it to lease out the equipment at inflated prices.³¹ Hanover Shoe sought to recover the amount of the overcharge, or the difference between what it paid United and what it would have paid in a competitive market where United could not have inflated the prices.³²

In its defense, United argued that Hanover suffered “no legally cognizable injury” because it “passed on” the overcharge to its consumers by increasing the price of its shoes.³³ The Supreme Court rejected this defense, finding that Hanover Shoe made out its prima facie case of injury and damage by showing that it had paid a price that was “illegally high” and demonstrating “the amount of the overcharge.”³⁴ Hanover remained “equally entitled to damages” regardless of whether it absorbed the overcharge or opted to pass the overcharge on to its own customers in the form of higher shoe prices.³⁵

27. *Id.*

28. *See* 431 U.S. 720 (1977).

29. 392 U.S. 481 (1968).

30. *Id.* at 481.

31. *Id.* at 483.

32. *Id.* at 484.

33. *Id.* at 487, 494.

34. *Id.* at 489.

35. *Hanover Shoe, Inc. v. United Shoe Machinery Corp.*, 392 U.S. 481, 489 (1968).

The Court justified its decision by alluding to the administrability concerns that would befall courts if defendants were able to rely on a “pass-on” defense.³⁶ In particular, the Court noted that a successful pass-on theory of defense would require “a convincing showing of . . . virtually unascertainable figures,” such as whether Hanover Shoe would have raised prices even absent the overcharge.³⁷ Moreover, if such a defense were available, the Court feared that defendants would “frequently” invoke it to avoid potential treble-damage liability and that this would lead to overly drawn out and complex proceedings involving “massive evidence and complicated theories.”³⁸

2. Illinois Brick

Illinois Brick broadened the restriction on the pass-on theory, holding that it should apply equally to plaintiffs and defendants.³⁹ In *Illinois Brick*, the State of Illinois attempted to sue a group of concrete brick manufacturing companies for allegedly engaging in a conspiracy to fix the prices of concrete blocks.⁴⁰ However, Illinois did not purchase the blocks directly from Illinois Brick or the other manufacturing companies. Instead, the manufacturers sold the bricks to masonry contractors who used them to build masonry structures.⁴¹ The masonry contractors then sold the structures to general contractors, who incorporated them into entire buildings.⁴² The State of Illinois, as the end-consumer in the chain, then bought those buildings from the general contractors.⁴³ The concrete blocks had passed through “two separate levels” of the distribution chain before reaching the State of Illinois.⁴⁴ Nevertheless, Illinois alleged that it overpaid for the buildings by more than \$3 million because of Illinois Brick’s price-fixing conspiracy at the top of the chain.⁴⁵

In deciding the case, the Supreme Court extended the logic of *Hanover Shoe*, holding that if defendants could not use a “pass-on” theory of defense, Illinois as a plaintiff could not seek to recover an anticompetitive overcharge that was “passed on” to it by an intermediary in the supply chain.⁴⁶ This meant that only *direct* purchasers—those who actually purchase the goods directly from

36. *Id.* at 493.

37. *Id.*

38. *Id.*

39. *Ill. Brick Co. v. Ill.*, 431 U.S. 720, 728 (1977).

40. *Id.* at 726–27.

41. *Ill. Brick Co. v. Ill.*, 431 U.S. 720, 727 (1977).

42. *Id.*

43. *Id.* at 726.

44. *Id.*

45. *Id.* at 727.

46. *Id.* at 728.

the alleged antitrust violator—can seek treble damages for antitrust violations, while *indirect* purchasers—those more than one step removed from the antitrust violator in the supply chain—cannot bring such an action.

Three particular policy rationales motivated the *Illinois Brick* Court to disallow all indirect purchaser suits. First, the Court feared that allowing indirect purchaser suits would expose firms to “serious risk of multiple liability.”⁴⁷ Allowing an offensive pass-on theory of recovery presumes that the direct purchaser can seek full recovery for the overcharge. However, *Hanover Shoe* would prevent the defendant from “using that presumption against the other plaintiff.”⁴⁸ Such a result would mean that “overlapping recoveries are certain to result,” since multiple plaintiffs could each seek the full amount of the overcharge, even if they had passed the overcharge directly on to their own consumers.⁴⁹ This could potentially be devastating for a firm given that each antitrust plaintiff would be empowered to seek treble damages.⁵⁰

Second, the Court reasoned that a regime that only allowed the direct purchaser to pursue treble damages would better incentivize vigorous private enforcement of the antitrust laws, as opposed to a rule which allowed all injured parties to sue.⁵¹ Not only would this regime permit direct purchasers to “recover the full amount of the overcharge,” even if they had passed it on to their consumers, it would also spare direct purchasers “the burden of litigating the intricacies of pass-on,”⁵² thus potentially making them more likely to come forward with their claims.

Third, the Court echoed *Hanover Shoe*’s administrability concern that pass-on theories of recovery would “injec[t] extremely complex issues into the case.”⁵³ The Court considered the complexity to be a problem unto itself but also feared it might create uncertainty as to how to apportion overcharges at different levels of a distribution chain, which would “further reduce the incentive to sue.”⁵⁴ In foreclosing the possibility of complicated damages calculations by forbidding indirect purchaser suits, the Court hoped to promote efficient administration of the antitrust laws. For these reasons, *Illinois Brick* created a *per se* bar against indirect purchaser suits.

A simple illustration of the state of the law under *Illinois Brick* may be helpful at this point. In a hypothetical marketplace, Manufacturer A has a

47. *Ill. Brick Co. v. Ill.*, 431 U.S. 720, 730 (1977).

48. *Id.*

49. *Id.*

50. *Id.*

51. *Id.* at 745–46.

52. *Id.*

53. *Ill. Brick Co. v. Ill.*, 431 U.S. 720, 745 (1977).

54. *Id.*

monopoly on the production of shoelaces. It produces shoelaces and sells them to Distributor B at a price of \$15, though in a competitive marketplace it would only charge \$10. Manufacturer A is thus collecting a monopoly rent of \$5. Distributor B incorporates the shoelaces into pairs of shoes, which it sells to Consumer C. In a competitive marketplace, Distributor B would charge \$25 for the shoes and would be able to sell fifty pairs, but to account for the \$5 overcharge on shoelaces imposed by Manufacturer A, Distributor B must now charge \$30. At this price, it can only sell forty pairs, reducing its total revenue from \$1,250 (in a competitive market, charging \$25 for fifty pairs of shoes) to \$1,200. In this scenario, Consumer C is overpaying for shoes by \$5 because of Manufacturer A's anticompetitive conduct, but Consumer C cannot sue Manufacturer A to recover the overcharge because it is not a direct purchaser. Nor can Consumer C sue Distributor B, because Distributor B is not a monopolist. Distributor B, the direct purchaser, has passed on the overcharge to Consumer C. However, Distributor B can sue Manufacturer A to recover the lost profits it suffers as a result of Manufacturer A's prices. Moreover, the Clayton Act empowers Distributor B to seek treble damages for its injury, meaning it can seek \$150 in damages from Manufacturer A.

In sum, classification as an indirect purchaser is fatal for plaintiffs seeking to recover an overcharge passed on to them from an intermediary in a standard vertical distribution chain.

III. *APPLE V. PEPPER*

A. CASE BACKGROUND AND PROCEDURAL HISTORY

Apple v. Pepper revisits the direct versus indirect purchaser distinction, in light of a twenty-first century twist: the Apple App Store. The App Store is the only place where iPhone users can legally purchase apps for their Apple devices.⁵⁵ Independent app developers, and not Apple, create the vast majority of apps available on the App Store.⁵⁶ In *Apple v. Pepper*, consumers who purchased iPhone apps from the App Store sued Apple, claiming that it abused its monopoly over the sale of iPhone apps to overcharge consumers.⁵⁷

The plaintiffs' suit may first seem to be directed at the wrong defendant, since technically it is the app developers and not Apple that set the prices of the apps. However, Apple requires that the price of all apps end in ".99."⁵⁸ This requirement has led to a phenomenon in which most app prices "cluster"

55. *Apple, Inc. v. Pepper*, 139 S. Ct. 1514, 1519 (2019).

56. *Id.*

57. *Id.*

58. *Id.* at 1528 (Gorsuch, J., dissenting).

at \$0.99.⁵⁹ The requirement also restricts app developers' ability to increase the price of a \$0.99 app, since the only option would be to effectively double the price to \$1.99. In addition to the price requirement, Apple requires that developers pay a \$99 annual fee to list their apps for sale on the App Store.⁶⁰

Consumers purchase the apps from Apple, which takes a thirty-percent commission and passes the sale proceeds on to the app developer.⁶¹ The lawsuit centered on this thirty-percent commission, which the plaintiffs alleged was an abuse of Apple's monopoly power because it forced consumers to pay prices higher than what they would have otherwise paid in a competitive market for the same apps.⁶²

In its defense, Apple alleged that iPhone users did not have standing to bring the suit because the app *developers*, and not Apple, set the price of the apps.⁶³ Even if the consumers technically gave their money to Apple, Apple contended that it acted only as an intermediary in the transaction, and consumers were not directly purchasing the apps from Apple.⁶⁴ To the extent that the consumers were paying any overcharge, Apple argued that it was the app developers who had chosen to pass the overcharge onto the consumers, and thus the consumers should not have standing to sue Apple.⁶⁵

The trial court granted Apple's motion to dismiss on the grounds that any injury the plaintiffs suffered was an "indirect effect resulting from the software developers' own costs."⁶⁶ The Ninth Circuit reversed, emphasizing that even if app developers were responsible for setting the price, consumers did in fact purchase their apps directly from Apple, which passed on the proceeds of the sale to the developers, less the thirty-percent commission.⁶⁷

B. CASE SUMMARY

Justice Brett Kavanaugh, writing for a narrow 5-4 majority, affirmed the Ninth Circuit's interpretation of *Illinois Brick*—a consumer cannot sue an alleged monopolist who is "two or more steps removed from the consumer in

59. *Id.*

60. *Id.* at 1519.

61. *Id.*

62. *Apple, Inc. v. Pepper*, 139 S. Ct. 1514, 1519 (2019).

63. *Id.*

64. *Id.*

65. *Id.* at 1521–22.

66. *In re Apple iPhone Antitrust Litig.*, 2013 WL 6253147, at *6 (N.D. Cal., Dec. 2, 2013), *rev'd*, 846 F.3d 313, 325 (9th Cir. 2017), *aff'd sub nom.* *Apple Inc. v. Pepper*, 139 S. Ct. 1514, 1525 (2019).

67. *In re Apple iPhone Antitrust Litig.*, 846 F.3d at 324–25, *aff'd sub nom.* *Apple v. Pepper*, 139 S. Ct. at 1525.

a vertical distribution chain.”⁶⁸ However, applying that rule to this case would not bar the consumer plaintiffs’ suit, because they purchased apps directly from Apple.⁶⁹ Moreover, the Court was persuaded by the broad language of the Clayton Act, which empowers “any person who shall be injured in his business or property” by a violation of the antitrust laws to sue the defendant “and [] recover threefold the damages” suffered.⁷⁰

The majority also criticized Apple’s proposed alternative construction of the direct versus indirect purchaser distinction: instead of considering from whom the consumers purchased the apps, Apple would have the inquiry focus on who sets the price.⁷¹ Apple argued that even if the consumers purchased the apps from Apple, it merely acted as an intermediary to facilitate the transaction between the developers and the consumers. Since the developers are responsible for the price of the app, Apple argued that the developers should be the ones that the consumers sue for anticompetitive pricing behavior.

The Court dismissed Apple’s theory as “not persuasive economically or legally” and laid out two examples of different pricing models to illustrate this point.⁷² In one, a traditional markup pricing model, an allegedly monopolistic retailer could purchase a product from a manufacturer for \$6, sell it for \$10, and make a \$4 profit as a result. In a commission-pricing model, a retailer can agree to sell a manufacturer’s product on its behalf at \$10 and take a forty-percent commission, returning \$6 to the manufacturer. Both scenarios produce economically identical results: the retailer makes \$4, the manufacturer makes \$6, and the consumer pays \$10.⁷³ And yet, Apple’s “who set the price” theory would allow the consumer to sue the monopolistic retailer in the former markup-pricing model but preclude a suit in the latter commission-pricing model, since there the manufacturer set the price.⁷⁴

68. *Apple v. Pepper*, 139 S. Ct. at 1519–20.

69. *Id.* at 1521.

70. *Id.* at 1520; 15 U.S.C. § 15(a).

71. *Apple v. Pepper*, 139 S. Ct. at 1522.

72. *Id.*

73. Justice Kavanaugh’s reasoning has been criticized on this point, as it “erroneously suggests that ad valorem royalties [such as the thirty-percent commission here] are economically equivalent to linear prices.” Bruce H. Kobayashi & Joshua D. Wright, *What’s Next in Apple Inc. v. Pepper? The Indirect-Purchaser Rule and the Economics of Pass-Through*, CATO SUPREME COURT REV. 249, 262 (2019). In fact, when faced with an ad valorem royalty, the app developer would lower gross prices in order to maximize total revenue, leading to lower prices, higher output, higher consumer welfare, and higher joint profits for the app developers and Apple. For more, see *id.* at 261–67.

74. *Apple v. Pepper*, 139 S. Ct. at 1522.

Justice Kavanaugh dismissed Apple's theory as serving only to "gerrymander Apple out of this and similar lawsuits."⁷⁵ He further observed that, if accepted, the theory would "provide a roadmap" for monopolistic retailers to adopt pricing models so as to "evade antitrust claims by consumers and [] thwart effective antitrust enforcement."⁷⁶ An inquiry focused on the pricing model would elevate form over substance in the majority's view, whereas the current indirect purchaser rule properly allows consumers to sue when a "retailer's unlawful monopolistic conduct" caused them to pay "higher-than-competitive" prices.⁷⁷

Apple further raised concerns that allowing iPhone users to sue would burden the courts with overly complex damages calculations.⁷⁸ Justice Kavanaugh dismissed this concern as well, noting that "*Illinois Brick* [was] not a get-out-of-court-free card for monopolistic retailers to play" whenever faced with the possibility of complicated damages calculations.⁷⁹ Antitrust cases often require expert testimony and deal with difficult-to-compute damages, but *Illinois Brick* "surely did not wipe out consumer antitrust suits against monopolistic retailers."⁸⁰

Finally, the majority responded to Apple's concerns that the Court's ruling could leave it vulnerable to lawsuits from both the "downstream" consumers and "upstream" app developers, a scenario that the *Illinois Brick* Court explicitly tried to prevent, according to Apple.⁸¹ The majority disagreed with Apple that this scenario would result in "conflicting claims to a common fund" because the consumers and app developers would be seeking damages for different types of anticompetitive behavior.⁸² The consumers would seek to recover the full amount of the unlawful overcharge they paid directly to Apple.⁸³ Since there is no distribution chain in this scenario, there would be no need to trace the overcharge back to the original purchaser; the consumers are the initial and only purchaser of the good, and thus bear the entire amount of the alleged overcharge.⁸⁴

The developers, on the other hand, would seek to recover the lost profits they suffered because they increased the price of the app to account for the

75. *Id.* at 1522–23.

76. *Apple v. Pepper*, 139 S. Ct. at 1523.

77. *Id.*

78. *Id.* at 1524.

79. *Id.*

80. *See id.*

81. *Id.* at 1523.

82. *Apple v. Pepper*, 139 S. Ct. at 1525.

83. *Id.*

84. *See id.*

overcharge (in other words, they “passed on” the overcharge).⁸⁵ In a competitive market, the argument goes, the price of their product would be lower, and they would be able to sell more apps for greater total profit. The majority pointed out that it would hardly be unusual for a retailer in a similar case to face claims from multiple classes of plaintiffs in cases where, as here, an intermediary is an alleged monopolist to consumers on one end and an alleged monopsonist⁸⁶ to manufacturers on the other. Though both the consumers and app developers could sue, their suits “would rely on fundamentally different theories of harm,” and thus the rationale from *Illinois Brick* would not bar either.⁸⁷

In dissent, Justice Neil Gorsuch, joined by Chief Justice John Roberts and Justices Clarence Thomas and Samuel Alito, argued that because the thirty-percent commission Apple charged initially fell on the developers, the commission can only cause injury to the plaintiffs if the developers “are able and choose to pass on the overcharge” by increasing the price of the apps.⁸⁸ In the dissent’s view, this made the case “just the sort of pass-on theory that *Illinois Brick* forbids,” meaning that iPhone users were plainly barred from bringing suit.⁸⁹ However, Justice Gorsuch expressed doubt that the developers were even capable of passing the overcharge on to the consumers given Apple’s \$0.99 price requirement.⁹⁰ This requirement restrains the app developers from increasing the price just enough to recoup the thirty-percent commission, since a developer who set the price of her app at \$0.99 would have no choice but to increase it to \$1.99 at a minimum (which would also “doubl[e] the commission in the process”⁹¹).

Though I agree with the outcome the majority reaches, I argue that *Illinois Brick*’s indirect purchaser rule does not support the majority’s conclusion. Indeed, Justice Gorsuch makes a more persuasive argument that the proper application of *Illinois Brick* would preclude iPhone users from bringing suit for the simple reason that the thirty-percent commission initially falls on app

85. *See id.*

86. A “monopolist” is generally understood to have a monopoly over the market in which it sells, while a “monopsonist” has a monopoly “in the market in which it buys.” AREEDA & HOVENKAMP, *supra* note 19, at 1478c, n. 24. Applying both terms here, Apple is a monopolist toward its consumers, since they cannot purchase or otherwise obtain apps for their iPhones from any firm besides Apple. On the other hand, Apple acts as a monopsonist toward app developers, since it is the only consumer of their iPhone app developing services.

87. *See* Apple v. Pepper, 139 S. Ct. at 1525.

88. *Id.* at 1528 (Gorsuch, J., dissenting).

89. *See id.* at 1527.

90. *Id.* at 1528.

91. *Id.*

developers.⁹² iPhone users are only injured if the developers pass the overcharge on to them.⁹³

Despite my view that the majority used a faulty approach, I agree with the majority's conclusion that iPhone users *should* have standing to sue Apple because, as the end-consumers, they bear the full brunt of the alleged overcharge.⁹⁴ However, in order to reach this outcome, the majority distorted *Illinois Brick*, even if nominally affirming it. First, by stating that any ambiguity in the precedent should be resolved in the direction of the statutory text, Justice Kavanaugh severely undermined the *Illinois Brick* test, which purposefully restricted the broad language of the Clayton Act.⁹⁵ Moreover, Justice Kavanaugh cast doubt on the *Illinois Brick* policy rationale of allowing only the direct purchaser to recover treble damages in order to incentivize private enforcement.⁹⁶ He also gave little credence to administrability concerns about the difficulty of apportioning damages among multiple parties.⁹⁷

As a result, the majority opinion in *Apple v. Pepper* effectively lays the groundwork “for a subsequent case overturning *Illinois Brick*,” despite superficially hewing to its precedent.⁹⁸ This raises the question: would courts be better off without *Illinois Brick*? If it were overturned, what tools might they use in its place?

IV. UPDATING ANTITRUST STANDING ANALYSIS FOR TWO-SIDED MARKETPLACES

A. WHO NEEDS *ILLINOIS BRICK* ANYWAY?

The viability of *Illinois Brick* forty years after its passage is increasingly being called into question. Indeed, thirty-one states filed an amicus brief in *Apple v. Pepper* calling on the Court to explicitly overturn *Illinois Brick* since the

92. *Apple v. Pepper*, 139 S. Ct. at 1528.

93. *Id.* at 1525 (majority opinion).

94. I express no opinion whether Apple's thirty-percent commission does in fact cause the plaintiffs to pay a higher-than-competitive price for apps. I argue only that iPhone users should have standing to bring the suit.

95. Matthew Perlman, *DOJ Antitrust Deputy Blasts Supreme Court's Apple Ruling*, LAW360 (Sept. 26, 2019, 6:35 PM), <https://www.law360.com/articles/1203203/doj-antitrust-deputy-blasts-supreme-court-s-apple-ruling>.

96. *See Apple v. Pepper*, 139 S. Ct. at 1525.

97. *See id.*

98. Perlman, *supra* note 95; *see also* Jeffrey L. Harrison, *After Forty Years of Antitrust Revision and Apple v. Pepper, What Now Illinois Brick?* 1, 5 (July 8, 2019) (manuscript), *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3413197 (observing that “both the majority and the dissent make arguments that seem to undermine the holding in *Illinois Brick*”).

“predictions and policy concerns” undergirding the case “have been undermined by subsequent experience and events.”⁹⁹ The Court declined to do so, and it instead “nominally [affirmed] *Illinois Brick* in this case.¹⁰⁰ I argue that this was a mistake and that the Court should have explicitly overruled *Illinois Brick*’s application to two-sided marketplaces. In this Part, I first argue that *Illinois Brick*’s indirect purchaser rule has outlived any usefulness it had with respect to two-sided marketplaces. Then, I proceed to consider an alternative test courts should adopt to analyze antitrust standing in cases dealing with two-sided marketplaces.

1. *Illinois Brick’s Indirect Purchaser Rule Does Not Fit Two-Sided Marketplaces*

Illinois Brick was decided based on a traditional resale model in which a straight line can be drawn from the manufacturer, through successive intermediary purchasers, to the final “indirect purchasing consumer.”¹⁰¹ Each entity in the vertical distribution chain purchased the goods from the entity above it, incorporated those goods into its own product, and resold the new product to the entity below it in the chain. To determine the direct purchaser in such a distribution chain, one need only look at the entity one production level up.

The lower courts have struggled to apply *Illinois Brick*’s bright-line rule to modern tech platforms where the identity of the direct purchaser is not so simple.¹⁰² For example, in its App Store, Apple essentially acts as an intermediary licensing agent. By listing an app developer’s app for sale in the App Store, Apple gives iPhone users the license to install the app on their devices.¹⁰³ Consumers in this model are technically the first to pay money to another party (here, Apple), thus laying the groundwork for their argument that they are direct purchasers from Apple. However, Apple does not directly impose the thirty-percent commission at the point of purchase. Rather, the commission falls on app developers and is subtracted from any profits they generate on each app sale.¹⁰⁴

99. Brief for Texas, Iowa, and 29 Other States as Amici Curiae in Support of Respondents, *Apple v. Pepper*, 139 S. Ct. 1514 (2019) (No. 17-204) 2018 WL 4808836, at *2.

100. See Perlman, *supra* note 95.

101. Brief for Texas, *supra* note 99, at *26; AREEDA & HOVENKAMP, *supra* note 19, at 346j.

102. See Seth E. Miller, *Seeing Over the Brick Wall: Limiting the Illinois Brick Indirect Purchaser Rule and Looking at Antitrust Standing in Campos v. Ticketmaster Corp. Through a New Lens*, 32 FLA. ST. U.L. REV. 197, 208–16 (2004).

103. Brief for Texas, *supra* note 99, at *26.

104. *Apple, Inc. v. Pepper*, 139 S. Ct. 1514, 1519 (2019).

In a case with a similar economic business model, *Campos v. Ticketmaster Corp.*, ticket-purchaser plaintiffs sued an event ticket distributor for charging service fees that they alleged were higher than would have been sustainable in a competitive market.¹⁰⁵ According to the plaintiffs, Ticketmaster entered into exclusive contracts with “almost every promoter of concerts in the United States.”¹⁰⁶ This gave it “ironclad control” over ticket distribution services for concerts at major venues in America, which allowed it to impose service fees that were supracompetitive—higher than what would be expected in a competitive marketplace—on the consumer plaintiffs.¹⁰⁷

The Eighth Circuit held that the plaintiffs were indirect purchasers under *Illinois Brick* and thus lacked standing to bring suit.¹⁰⁸ The court found the fact that the plaintiffs could not obtain ticket delivery services in a competitive market was “simply the consequence of the antecedent inability of venues to do so.”¹⁰⁹ Essentially, the court held that the consumers were not direct purchasers because, in an “antecedent transaction,” the venues purchased ticket distribution services from Ticketmaster first.¹¹⁰ Under that logic, any injury suffered by plaintiffs was derivative of the venues’ purchase.¹¹¹

The dissent sharply criticized the majority’s test.¹¹² It first pointed out that “antecedent transaction” was not a term from any cited antitrust scholarship, but rather, was created by the majority.¹¹³ Moreover, it noted that a “mere ‘antecedent transaction’ will not turn all purchasers of a monopolized product into indirect purchasers” without the passing on of monopoly costs from the direct purchaser to the indirect purchaser.¹¹⁴ The dissent argued this condition was not satisfied since Ticketmaster provided ticket distribution services directly to concertgoers, as opposed to a model in which the venues purchased a product that they would then sell to concertgoers.¹¹⁵

105. *Campos v. Ticketmaster Corp.*, 140 F.3d 1166, 1168 (8th Cir. 1998).

106. *Id.* at 1169.

107. *See id.*

108. *Id.* at 1171. Curiously, neither the majority nor dissent in *Apple v. Pepper* mentioned the *Ticketmaster* case, despite the similarities between the two cases and the fact that *Apple v. Pepper* reaches the opposite conclusion—effectively overruling *Ticketmaster*. *See* Herbert Hovenkamp, *Apple v. Pepper: Rationalizing Antitrust’s Indirect Purchaser Rule*, 120 COLUM. L. REV. F. 14, 18 (2020).

109. *Ticketmaster*, 140 F.3d at 1171.

110. *See id.*

111. *See id.*

112. *Id.* at 1174 (Arnold, J., dissenting).

113. *Id.*

114. *Id.*

115. *Id.*

Despite bearing the entirety of any monopoly overcharge, the *Ticketmaster* concertgoers were found to be indirect purchasers and their suit was dismissed for lack of standing.¹¹⁶ Interestingly, when the Supreme Court confronted a strikingly similar economic business model in *Apple v. Pepper*, it reached the opposite conclusion.

Indeed, the difficulty of applying *Illinois Brick*'s indirect purchaser rule came to a head with the core disagreements between the majority and dissent in *Apple v. Pepper*. In his reasoning, Justice Kavanaugh argued that Apple's proposed rule would "elevate form . . . over substance" by focusing on the "precise arrangement between manufacturers or suppliers and retailers" instead of whether or not the consumer is "paying a higher price because of the monopolistic retailer's actions."¹¹⁷ However, Justice Gorsuch leveled a similar criticism at the majority by arguing that allowing iPhone users to sue in this case distorts *Illinois Brick*, taking it from a "rule of proximate cause and economic reality" to a "formalistic rule of contractual privity."¹¹⁸ Additionally, both opinions express concern that the other's rule creates simple loopholes for potential antitrust violators to exploit; for example, by simply restructuring their pricing models.¹¹⁹

This tension between the two opinions only serves to underscore what a poor tool *Illinois Brick* is for deciding who should have standing in two-sided markets, which are not simple vertical distribution chains. *Illinois Brick* assumes verticality, where only one purchaser can be said to be the "direct" purchaser from the antitrust violator. But the App Store is not a vertical distribution chain, since both the app developers, who are themselves a consumer of Apple on one end of the marketplace, and Apple itself contribute to the prices that the consumers ultimately pay.¹²⁰

Thus, while Apple is the intermediary from which the consumers technically purchase the product directly, the allegedly anticompetitive commission is not immediately borne by the consumers. Plaintiffs can "be injured *only* if the developers . . . pass on the overcharge to them."¹²¹ Since

115. *Id.* at 1171 (majority opinion).

117. *Apple, Inc. v. Pepper*, 139 S. Ct. at 1523.

118. *See id.* at 1526 (Gorsuch, J., dissenting).

119. *See id.* at 1523 (majority opinion) ("Apple's theory would provide a roadmap for monopolistic retailers to structure transactions with manufacturers or suppliers so as to evade antitrust claims by consumers and thereby thwart effective antitrust enforcement. . . . [R]estructuring would allow a monopolistic retailer to insulate itself from antitrust suits by consumers"); *see also id.* at 1530 (Gorsuch, J., dissenting) ("To evade the Court's test, all Apple must do is amend its contracts.").

120. *Id.* at 1521 (majority opinion).

121. *Id.* at 1528 (Gorsuch, J., dissenting).

the commission first falls on the app developers, who could then pass it on to the consumers, the case presents a classic “pass-on” theory of recovery that is strictly disallowed by *Illinois Brick*’s bright-line rule.¹²²

As such, Justice Gorsuch made a more persuasive argument that a straightforward application of *Illinois Brick* would have found that the plaintiffs here lacked standing to sue. The *Illinois Brick* Court designed the indirect purchaser rule to operate in a traditional, “familiar” business model with “strong temporal and course-of-commerce separation” between the different entities in the chain: the producer at the top, the distributor in the middle, and the consumer at the bottom.¹²³ “[T]oday’s marketing arrangements were likely unforeseeable” to the *Illinois Brick* Court.¹²⁴ The decision failed to anticipate a marketplace like the App Store in which the party that sets the price, and controls whether or not an overcharge is passed on to its subsequent consumers, is not the party that directly sells the product to the purchaser. This makes the *Illinois Brick* indirect purchaser rule very difficult to apply to contemporary two-sided platforms and thus, a poor tool for assessing whether a plaintiff has standing to sue for an antitrust violation in these types of marketplaces.

2. *The Illinois Brick Administrability Concerns No Longer Apply*

Moreover, the administrability concerns that led the *Illinois Brick* Court to its decision are no longer valid. First, it is worth noting that *Illinois Brick* was highly controversial at the time of its deciding and continues to be so.¹²⁵ The

122. Justice Gorsuch points out that Apple’s sole pricing requirement—that the price of an app ends in “.99”—may actually make it extremely difficult for an app developer to “pass on” the overcharge of Apple’s thirty-percent commission fee. *See* *Apple v. Pepper*, 139 S. Ct. at 1528 (“[A] developer charging \$0.99 for its app can’t raise its price by just enough to recover the 30-cent commission.”). For example, an app developer who wanted to increase the price of an app currently priced at \$0.99 in order to recoup some or all of the thirty-percent commission would have no choice but to double the price to \$1.99 in order to keep it listed in the App Store. But if all competitor apps remained priced at \$0.99, the app developer risks pricing herself out of the market by doing so. However, this scenario assumes that the price ending in \$0.99 does not already capture (if not over capture) the commission. For example, absent Apple’s pricing requirement, if an app developer would otherwise price an app at \$0.76 and wanted to pass on the full thirty-percent commission to its purchasers, it would raise the price to \$0.99. Thus, any app that would be priced at less than \$0.76 if not for Apple’s \$0.99 pricing requirement is more than recovering the thirty-percent commission by being forced to price its app at \$0.99.

123. Manne & Stout, *supra* note 3, at 439.

124. Harrison, *supra* note 98, at 3.

125. *See* Edward D. Cavanagh, *Illinois Brick: A Look Back and a Look Ahead*, 17 LOY. CONSUMER L. REV. 1, 3 n.8 (2004) (noting that opinions about *Illinois Brick* are so unyielding that “panelists speaking at the ABA Antitrust Section’s annual Antitrust Remedies Forum in

Court imposed a limitation on recovery of damages that was plainly inconsistent with the broad language of the antitrust damages statute.¹²⁶ Furthermore, critics of the indirect purchase rule have argued that it perpetuates an “anti-consumer” antitrust system in which offending firms are under-deterred from violating the law.¹²⁷

In *Illinois Brick*'s wake, dozens of states enacted “*Illinois Brick* repealer statutes,” explicitly permitting indirect purchaser suits.¹²⁸ As a result of these statutes, federal courts have developed “decades” of experience analyzing indirect purchaser suits.¹²⁹ These cases can be seen as an experimental test run to determine if the administrability concerns that preoccupied the *Illinois Brick* Court ever came to pass. In short, they did not.

First, in barring indirect purchasers from recovering in antitrust suits, the Court intended to mitigate the “serious risk of multiple liability for defendants” that could arise if both indirect *and* direct purchasers were able to recover for all or part of an overcharge passed on.¹³⁰ However, antitrust commentators have noted that after decades of indirect purchaser suits under state law regimes, the risk of duplicative recovery remains a “theoretical problem” only. Indeed, in testimony the Antitrust Modernization Committee heard, as it prepared its 2007 Final Report and Recommendations, none of the witnesses to testify could “identify a single instance of ‘unfair or multiple recovery.’”¹³¹

Additionally, allowing these suits in federal court may actually further diminish the risk of multiple recovery under federal law, since “related actions in various districts can be consolidated for class certification” under the Class Action Fairness Act.¹³² For example, one potential case of duplicative recovery was foreclosed when private litigation followed the FTC’s case against Mylan Labs.¹³³ There, the defendants successfully argued that the settlement that the

spring 2003 in Washington, D.C. agreed not to discuss” the decision); *see also* Hovenkamp, *infra* note 108, at 14.

126. The Clayton Act provides that “any person who shall be injured in his business or property by reason of anything forbidden in the antitrust laws may sue.” 15 U.S.C. § 15(a).

127. *See* Andrew I. Gavil, *Thinking Outside the Illinois Brick Box: A Proposal for Reform*, 76 ANTITRUST L.J. 167, 171–72 (2009).

128. Cavanagh, *supra* note 125, at 2 n.4.

129. *See id.* at 2; *see also* Brief for Texas, *supra* note 99, at *4; *California v. ARC America Corp.*, 490 U.S. 93, 100 (1989) (holding that *Illinois Brick* does not preempt these state statutes that authorize indirect purchaser suits, or bar federal courts from hearing state law indirect purchaser claims).

130. *Ill. Brick v. Ill.*, 431 U.S. 720, 730 (1977).

131. Gavil, *supra* note 127, at 192 n.76.

132. Brief for Texas, *supra* note 99, at *20.

133. Gavil, *supra* note 127, at 192 n.76.

FTC obtained “precluded later certification of a class of direct purchasers on the ground that duplicative recovery would result.”¹³⁴

It is worth noting on this point that the existing *Illinois Brick* rule arguably facilitates duplicative recovery when actions are simultaneously brought under federal antitrust laws and state antitrust laws.¹³⁵ In this scenario, defendants can be forced to compensate direct purchasers for the full overcharge amount as federal law damages, in addition to making a “duplicate payment” to indirect purchasers for their recovery under state law.¹³⁶

Second, the *Illinois Brick* Court was concerned that authorizing indirect purchaser suits would “transform treble damages actions into massive efforts to apportion the recovery” among all injured plaintiffs that might have absorbed some portion of the overcharge.¹³⁷ Any attempt to appropriately allocate the overcharge would “add whole new dimensions of complexity” to the calculation of damages in antitrust suits, threatening to “seriously undermine their effectiveness” as a tool of deterrence.¹³⁸ Putting aside the fact that federal courts already face the difficulty of apportioning damages when state law indirect purchaser suits find their way into federal courts, advancements in economic theory and technology have made apportioning damages among multiple parties significantly easier.¹³⁹

Further, courts have come to rely on relatively simple econometric methods to calculate damages when an overcharge exists. Indeed, neither of the two most common approaches, the “before and after” method and the “yardstick” method, actually “compute [any] pass-on at all.”¹⁴⁰ The yardstick method simply compares the price that plaintiffs paid after the “monopolization or price-fixing activity” to that paid by “similarly situated persons” prior to the activity.¹⁴¹ This method aims to capture a plaintiff’s actual economic condition by comparing it to “what its condition would have been ‘but for’ the unlawful behavior of the defendant[].”¹⁴² The latter “yardstick” method aims to identify a firm that is comparable in all important respects to the plaintiff, ideally a “clone or an identical twin” to the plaintiff.¹⁴³ The

134. *Id.*

135. *See* AREEDA & HOVENKAMP, *supra* note 19, at 346a.

136. *See id.*

137. *Ill. Brick v. Ill.*, 431 U.S. 720, 737 (1977).

138. *Id.*

139. Gavil, *supra* note 127, at 190.

140. *Id.*

141. *Id.*; Brief of Antitrust Scholars as Amici Curiae in Support of Respondents, *Apple v. Pepper*, 139 S. Ct. 1514 (2019) (No. 17-204) 2018 WL 4773103, at *27.

142. AREEDA & HOVENKAMP, *supra* note 19, at 392e.

143. *Id.* at 392f.

comparable firm's economic performance in the market is used to estimate what the plaintiff's performance would have been "but for" the violation.¹⁴⁴

A "broad consensus in the literature" supports the notion that these methods are "fully capable" of apportioning damages between multiple parties, despite not calculating the pass-on in a case.¹⁴⁵ This suggests that calculating damages for indirect purchasers "is not as difficult as the Supreme Court believed" when it decided *Illinois Brick*.¹⁴⁶

In sum, the administrability concerns that the *Illinois Brick* Court foresaw in allowing indirect purchaser suits have not come to pass. These concerns are no longer viable reasons for continuing to adhere to *Illinois Brick*'s bright-line rule.

3. *Illinois Brick Does Not Promote Effective Private Enforcement of the Antitrust Laws*

Illinois Brick's policy of vesting the entirety of the recoverable treble damages in just the direct purchaser no longer serves its own goals of promoting vigorous private enforcement of the antitrust laws. The *Illinois Brick* Court acknowledged that this policy could not guarantee full private enforcement.¹⁴⁷ Still, it suggested that "on balance" the aims of antitrust law were "better served by holding direct purchasers to be injured to the full extent of the overcharge" they paid rather than trying to "apportion the overcharge among all that may have absorbed a part of it."¹⁴⁸ However, decades of decisions since *Illinois Brick* demonstrate that a bar on indirect purchaser suits hinders rather than helps to promote antitrust's core goals of deterring potential violators and compensating victims.

First, barring indirect purchaser suits limits deterrence, since the party that is the direct purchaser will often lack an incentive to sue. For example, under Justice Gorsuch's application of *Illinois Brick*, the app developers are "the parties who are directly injured" by the allegedly monopolistic commission that Apple collects.¹⁴⁹ However, suppliers, such as the app developers, will often forego bringing suit over anticompetitive behavior perpetrated by a monopolistic retailer either because they are " beholden to the monopolist to distribute their product[s]" or because "they may [stand to] benefit from the

144. *Id.*

145. Brief of Antitrust Scholars, *supra* note 141, at *28; AREEDA & HOVENKAMP, *supra* note 19, at 346a.

146. *Id.*

147. *Ill. Brick Co. v. Ill.*, 431 U.S. 720, 746 (1977).

148. *Id.*

149. *Apple v. Pepper*, 139 S. Ct. at 1528.

monopoly.”¹⁵⁰ Indeed, the app developers may share in Apple’s monopoly rent profits, thus benefitting directly from its anticompetitive conduct.¹⁵¹ For example, Apple’s requirement that app prices end in “.99” may actually allow app developers to charge higher prices than they would in a competitive market.¹⁵² Thus, even if the commission constitutes a monopoly rent that Apple collects, app developers would be unlikely to sue for this since they benefit from Apple’s anticompetitive conduct in other ways.

Second, *Illinois Brick*’s indirect purchaser rule created a scheme that inevitably undercompensates actual victims. The Court solely vested damages in direct purchasers, even though direct purchasers will often pass on any overcharge that suppliers impose on them and thus suffer no actual injury. This rule elevates direct purchasers “to a preferred position as private attorneys general” at the expense of the “indirect purchasers who may have been actually injured by antitrust violations.”¹⁵³ Nevertheless, the *Illinois Brick* rule denies the indirect purchasers recovery.¹⁵⁴ The Court reasoned that a rule which allowed indirect purchasers to pursue damages claims would “simply deplet[e] the overall recovery in litigation over pass-on issues,” since most indirect purchasers would have such small damages in a suit that “only a small fraction” would “come forward to collect” them.¹⁵⁵

In other words, to prevent direct purchasers from receiving fewer damages, the Court barred the actually injured end-consumers from pursuing damages because they would be unlikely to collect them anyway. Meanwhile, the direct purchasers may seek the full amount of the overcharge “even if this exceed[s] the actual harm suffered by that purchaser.”¹⁵⁶ This necessarily

150. See Brief for the American Antitrust Institute as Amicus Curiae in Support of Respondents, *Apple v. Pepper*, 139 S. Ct. 1514 (2019) (No 17-204) 2013 WL 4846924, at *2; see also Robert G. Harris & Lawrence A. Sullivan, *Passing on the Monopoly Overcharge: A Comprehensive Policy Analysis*, 128 U. PA. L. REV. 269, 351–52 (1979) (arguing that the risk of being cut off entirely from a supplier might deter a direct purchaser from suing even where it absorbs a significant part of an overcharge).

151. See Miller, *supra* note 102, at 223 (noting that the analogous concert venues in *Ticketmaster* “receive part of the monopoly overcharge” that Ticketmaster extracted, “in exchange for granting [] Ticketmaster the exclusive right to distribute” event tickets).

152. Transcript of Oral Argument at 58, *Apple v. Pepper*, 139 S. Ct. 1514 (2019) (No. 17-204).

153. See *Illinois Brick*, 431 U.S. at 746.

154. See *id.* at 747.

155. *Id.*

156. Roger D. Blair & Jeffrey L. Harrison, *Reexamining the Role of Illinois Brick in Modern Antitrust Standing Analysis*, 68 GEO. WASH. L. REV. 1, 1 (1999).

creates a scheme of liability which “greatly overcompensates intermediaries and greatly undercompensate[s] consumers.”¹⁵⁷

Further, court decisions following *Illinois Brick* have dramatically limited direct purchasers’ ability to recover, further reducing the likelihood that victims are efficiently compensated for their injuries. Broadly, scholars have observed the general erosion of “per se rules” in antitrust law.¹⁵⁸ Instead, there has been a broad push for more antitrust cases to be assessed under the more defense-friendly “rule of reason.”¹⁵⁹ Under the rule of reason, the plaintiff has the initial burden of “demonstrating that the defendant’s action was anticompetitive.”¹⁶⁰ If the plaintiff meets its initial burden, the burden shifts to the defendant to provide sufficient evidence that the action had a “procompetitive effect[]”; the plaintiff can then demonstrate that the defendant could have achieved the same pro-competitive effects through alternative, “less restrictive methods.”¹⁶¹ At that point, the court balances the pro- and anticompetitive effects to determine which side prevails.¹⁶² The balancing inquiry has resulted in courts dismissing the vast majority of rule of reason cases at the motion to dismiss phase.¹⁶³ Indeed, in a study conducted of all rule of reason cases decided between 1977 and 2009, almost ninety-seven percent were dismissed after the pleadings alone.¹⁶⁴

Several recent cases in particular have dampened an antitrust plaintiff’s likelihood of success in court. *Bell Atlantic v. Twombly*¹⁶⁵ imposed a higher “plausibility” pleading burden on antitrust plaintiffs. Later, in *American Express v. Italian Colors Restaurant*,¹⁶⁶ the Court held that potential antitrust defendants could use arbitration clauses in standard-form contracts to ban antitrust class actions.¹⁶⁷ This requires plaintiffs to individually arbitrate their antitrust disputes, which rarely makes economic sense because the dollars at stake in each individual case are unlikely to outweigh the substantial costs of pursuing the claim in arbitration. Moreover, this “deprive[s] overcharged direct

157. AREEDA & HOVENKAMP, *supra* note 19, at 346k.

158. *See* Harrison, *supra* note 98, at 27.

159. *Id.* at 29–32.

160. *Id.* at 32.

161. *Id.*

162. *Id.*

163. *Id.*

164. Michael A. Carrier, *The Rule of Reason: An Empirical Update for the 21st Century*, 16 GEO. MASON L. REV. 827, 829 (2009).

165. *Bell Atlantic Corp. v. Twombly*, 550 U.S. 544, 556 (2007).

166. *American Exp. Co. v. Italian Colors Restaurant*, 570 U.S. 228, 239 (2013).

167. Mark A. Lemley & Christopher R. Leslie, *Antitrust Arbitration and Illinois Brick*, 100 IOWA L. REV. 2115, 2116 (2015).

purchaser[s] of the tools [of] antitrust law,” making private enforcement less effective and less likely.¹⁶⁸

Additionally, in its term prior to *Apple v. Pepper*, the Supreme Court heard a significant antitrust case that might have even further decreased an antitrust plaintiff’s likelihood of success: *Ohio v. American Express* (“*Amex*”).¹⁶⁹ *Amex* held that anticompetitive conduct in a two-sided marketplace, that is analyzed as one market, is not forbidden where its pro-competitive effects outweigh its anticompetitive effects.¹⁷⁰ The impact of *Amex* is merely speculative at this point, since lower courts are in the early stages of interpreting the rule. However, critics have characterized the decision as a “huge blow” to antitrust enforcement that will make it easier for dominant tech firms to “abuse their market power with impunity.”¹⁷¹ The general trend set into motion by *Amex* and its predecessor cases suggests the odds are increasingly stacked against plaintiffs in antitrust suits.

The policy rationales that motivated the *Illinois Brick* decision no longer justify adherence to its harsh and arbitrary restriction against recovery for indirect purchasers, and the Court should have explicitly overruled it in *Apple v. Pepper*. First, *Illinois Brick* makes a distinction between direct and indirect purchasers that assumes a vertical distribution chain that makes it a bad fit for application to two-sided marketplaces. Second, the administrability concerns of allowing indirect purchaser suits never came to pass. Finally, *Illinois Brick*’s policy of vesting treble damages solely in the direct purchaser arguably leads to an under-enforcement of the antitrust laws. This is even further complicated by the recent trend toward more defendant-friendly standards of analysis in antitrust cases.

B. ONE POSSIBLE WAY FORWARD WITHOUT *ILLINOIS BRICK*

If the Court were to overturn *Illinois Brick*, what should the antitrust standing analysis look like? Any “optimal” solution to the rule would be one

168. *See id.*

169. *Ohio v. American Express*, 138 S. Ct. 2274, 2280 (2018). The case has no connection to *American Express v. Italian Colors*.

170. *Id.* at 2280.

171. Lina Khan, *The Supreme Court just quietly gutted antitrust law*, VOX (July 3, 2018, 9:40 AM), <https://www.vox.com/the-big-idea/2018/7/3/17530320/antitrust-american-express-amazon-uber-tech-monopoly-monopsony>; *see also* Manne & Stout, *supra* note 3, at 455 (“[W]hereas alleging a two-sided market may make it easier for plaintiffs to demonstrate standing, *Amex*’s requirement that net harm be demonstrated across interrelated sets of users makes it more difficult for plaintiffs to make out a prima facie case.”); Hovenkamp, *supra* note 8, at 752 (“The *AmEx III* decision acts to increase the plaintiff’s burden of production in making a prima facie case There are no potential defendants who would not benefit from this.”).

that sidesteps the unnecessary difficulty of computing complicated pass-on damages, avoids duplicative recoveries, and “satisfies the statutory language” of the Clayton Act, which provides for damages to be recovered broadly by any person injured by something forbidden in the antitrust laws.¹⁷² A test which meets these criteria would allow the so-called direct purchaser (here, the app developers, since the commission falls on them first) to seek damages for any lost profits resulting from the higher price they were forced to charge.¹⁷³ The test would also allow “end-use consumers,” who are not in a position to pass on any overcharges, to pursue recovery of the full overcharge they paid.¹⁷⁴

Importantly, overturning *Illinois Brick* would not have to upend decades of antitrust standing doctrine. Any test which replaced *Illinois Brick* could rely on the same basic policy principles that undergirded that decision but could offer a “more sophisticated” approach that could “better comport with more complex and nuanced economic analysis.”¹⁷⁵

1. AGC Remoteness Test

As different approaches continue to develop in the scholarly antitrust literature, the “remoteness” test for antitrust standing articulated in *Associated General Contractors of California, Inc. v. California State Council of Carpenters* (“AGC”) has emerged as a strong candidate to replace *Illinois Brick*’s indirect purchaser rule.¹⁷⁶

a) Case Summary

In AGC, the Court considered a dispute between a group of construction worker unions and the Associated General Contractors of California (“Associated”), a membership corporation of building and construction contractors.¹⁷⁷ The unions brought suit against Associated for conspiring to “weaken the collective-bargaining relationship” the unions struck with the signatory employers through coercion, which allegedly amounted to an unlawful restraint of trade under the Sherman Act.¹⁷⁸

To evaluate whether the unions had standing to bring these claims, the Court in AGC articulated a two-step test. The first step requires a plaintiff to demonstrate an antitrust injury, meaning that the plaintiff must demonstrate

172. AREEDA & HOVENKAMP, *supra* note 19, at 346a; 15 U.S.C. § 15(a).

173. See Herbert Hovenkamp, Apple v. Pepper: *Rationalizing Antitrust’s Indirect Purchaser Rule*, 120 COLUM. L. REV. F. 14, 18 (2020).

174. AREEDA & HOVENKAMP, *supra* note 19, at 346a; see also *id.*

175. Manne & Stout, *supra* note 3, at 435.

176. 459 U.S. 519, 537–38 (1983).

177. *Id.* at 521.

178. AGC, 459 U.S. at 522 (1983).

they suffered a harm “that the antitrust statute was intended to forestall.”¹⁷⁹ If the plaintiff convincingly pleads an antitrust injury, the court then balances several factors against one another. Principally, the court considers the “directness or indirectness” of the injury,¹⁸⁰ the existence of plaintiffs whose “self-interest would normally motivate them” to sue for antitrust violations,¹⁸¹ and whether the complexity of the trial would exceed “judicially manageable limits.”¹⁸²

In the case of *AGC*, the Court first determined that plaintiffs did not meet the initial “antitrust injury” threshold.¹⁸³ It found that the alleged anticompetitive conduct that the contractors engaged in did not threaten the “economic freedom of [the market] participants” that the Sherman Act was enacted to protect.¹⁸⁴ The Court found that since the unions aimed to enhance the earnings of its membership, they might in fact be harmed by enhanced competition in the market, which may encourage employers to reduce costs to compete with rival firms.¹⁸⁵ Moreover, the Court noted that a distinct body of federal labor law had developed to “protect and encourage” the various activities of labor unions.¹⁸⁶ This would tend to shift similar disputes from the scope of the Sherman Act toward the purview of the relevant labor laws, constituting another reason that the Sherman Act was not designed to protect against the unions’ alleged injury.¹⁸⁷

Despite the plaintiffs’ failure to meet this threshold inquiry, the Court proceeded to analyze the other factors. It first determined that the indirectness of the unions’ injury was a factor weighing against a finding of standing.¹⁸⁸ The Court identified several “vaguely defined links” in the chain of causation between the unions’ alleged injury and the alleged restraint of trade in the construction subcontracting market.¹⁸⁹ The unions primarily argued that the defendants coerced landowners to divert business away from the unions’ employees and instead toward nonunion contractors.¹⁹⁰ The Court determined that any injury this caused the unions was an “indirect result” of the harm

179. *Id.* at 540.

180. *Id.*

181. *Id.* at 542.

182. *Id.* at 543.

183. *Id.* at 538.

184. *Id.*

185. *AGC*, 459 U.S. at 538 (1983).

186. *Id.* at 540.

187. *Id.*

188. *Id.*

189. *Id.*

190. *See id.* at 540–41.

suffered by the “immediate victims” of the defendants’ alleged coercion.¹⁹¹ The Court further reasoned that the more immediate victims would likely have their own interest in pursuing claims against the defendants, which also weighed against the plaintiffs’ argument.¹⁹²

Finally, the Court determined that it would be unduly burdensome to the court system to identify and apportion damages among the “directly victimized” parties and the “indirectly affected employees and union entities.”¹⁹³ It also worried that given the widespread pool of potential victims, the defendants faced a serious risk of duplicative damages.¹⁹⁴ Given the already speculative and indirect nature of the unions’ injuries, the Court held that the complexity of the damages calculation and the potential for duplicative recovery weighed heavily against a finding that the unions had standing.¹⁹⁵

b) Evaluating the *AGC* Test

One clear benefit of the *AGC* Test is that it provides courts with a flexible framework for analyzing standing in antitrust cases. It can just as easily be applied to traditional business arrangements, such as the union dispute in *AGC*,¹⁹⁶ and to more cutting-edge “marketing arrangements,” increasingly common amongst digital tech platforms.¹⁹⁷ Instead of articulating a bright-line rule that is only responsive to a narrow subset of business arrangements, the *AGC* Test allows courts to consider a handful of pertinent factors. This is unlike Illinois Brick’s “blunt-edged test” which, in cases like *Apple v. Pepper*, over-emphasizes formalities like contractual privity.¹⁹⁸ This comes at the expense of the economic reality of the transaction, in which the end-consumer who bears the brunt of an overcharge is unable to pass it on or recover the damage in court.¹⁹⁹

191. *AGC*, 459 U.S. at 541 (1983).

192. *See id.* at 545.

193. *Id.*

194. *Id.*

195. *Id.*

196. *See* Harrison, *supra* note 98, at 3. Notably, the nature of the claim in *AGC* was very different from that of *Illinois Brick* and *Apple v. Pepper*. The facts did not deal with the distribution of a product down a supply chain. The Court did not have to grapple with whether the plaintiffs were direct purchasers, since they did not purchase anything, and the indirect purchaser rule did not apply to bar the plaintiffs’ claims. Still, the Court considered the plaintiff union’s injury to be rather indirect, since they were not the immediate victims of the allegedly coercive conduct employed by Associated. Instead, Associated allegedly coerced third parties into contracting with non-union contractors. *See AGC*, 459 U.S. at 540–41.

197. *See* Harrison, *supra* note 98, at 3.

198. Brief for Texas, *supra* note 99, at *35.

199. Hovenkamp, *supra* note 108, at 18.

In contrast, a balancing inquiry like the AGC Test might better capture the nuances of a two-sided marketplace while still allowing a court to be mindful of the concerns expressed in *Illinois Brick*, to the extent they remain relevant. For example, if a plaintiff alleges an injury with too tenuous of a connection to the alleged anticompetitive conduct, a court would have the flexibility to consider this “indirectness” as simply a factor that cuts against the plaintiff’s claim.²⁰⁰ In a case where a plaintiff has otherwise met its burden of pleading anticompetitive conduct, a court may nevertheless allow the plaintiff’s claim to go forward if it determines that no other party exists that is able or sufficiently motivated to bring the claim.²⁰¹ Alternatively, where a plaintiff’s injury is too indirect and a court is concerned that the facts of a particular case might expose a firm to a high risk of duplicative recovery, the court could dismiss the action for lack of standing.

However, the AGC Test is not without flaws. In particular, the first and second factors seem to consider the same issue from different angles, and they should arguably be collapsed into one if the test were to be widely adopted. Evaluating the indirectness of a potential plaintiff’s injury necessarily requires consideration of whether someone else is better placed to sue, since what makes the plaintiff’s injury indirect is that it derives from a more direct harm suffered by another party.²⁰² Where a potential plaintiff’s injury is found to be indirect, there will always be another potential plaintiff with a more direct injury, since the directness inquiry is inherently relative.²⁰³ If the AGC Test were in fact adopted as a replacement for *Illinois Brick*, courts should articulate a more analytically clear approach for applying the test.

Still, the AGC Test would allow a court to consider the policy rationales undergirding the *Illinois Brick* decision individually on a case-by-case basis, instead of sweeping them into a single *per se* bar against lawsuits brought by certain types of participants in a marketplace. This is particularly beneficial when it becomes clear that certain policy rationales for the bar on indirect purchaser suits are overstated, such as the *Illinois Brick* Court’s concern that the complexity of calculating pass-on damages in indirect purchaser suits will inevitably overwhelm the court system.²⁰⁴ Furthermore, the AGC Test would give courts the flexibility to respond to changes in the economic structure of

200. *See AGC*, 459 U.S. at 540.

201. *See id.* at 545.

202. *See id.* at 541 (finding it “obvious” that the Union’s alleged injuries were “only an indirect result” of harm suffered by other parties).

203. *See id.* (noting that the “immediate victims” of the alleged coercion are the parties who suffered a direct injury, as opposed to the Union).

204. Gavil, *supra* note 127, at 190; *see also* AREEDA & HOVENKAMP, *supra* note 19, at 346a.

firms and marketplaces, which the indirect purchaser rule does not allow courts to do.

Of course, bright-line rules have their advantages. Even if the administrability concerns of Illinois Brick were overstated,²⁰⁵ having a per se bar on indirect purchaser suits is certainly easier for courts to administer than a multi-factor balancing inquiry which tries to account for the full picture of an alleged antitrust violation. Still, rules excluding indirect purchaser suits under federal antitrust law will likely do little to conserve courts' resources because federal courts already face the possibility of hearing indirect purchaser suits under state law.²⁰⁶

If the Court were to adopt the AGC Test for analyzing standing, it would have to clarify the confusing relationship that currently exists between Illinois Brick and AGC. Currently, most courts seem to view Illinois Brick as standing for the proposition that “the step-by-step analysis of AGC can be dispensed with” when the plaintiff is an indirect purchaser and that the Illinois Brick per se bar should apply instead.²⁰⁷ Since the two tests would often reach opposite outcomes on the same set of facts, with Illinois Brick flatly prohibiting indirect purchasers from bringing suit, and AGC merely counting indirect purchaser status as a factor weighing against the plaintiff,²⁰⁸ the Court would likely need to explicitly overrule Illinois Brick's application to two-sided marketplaces in order to adopt the AGC Test.

c) Apple v. Pepper Under the AGC Test

Had the Court instead applied the AGC Test to Apple v. Pepper, it likely would have reached the same outcome. At the first step of the analysis, a court would have considered whether the federal antitrust laws were enacted to protect against the particular violation the plaintiffs alleged.²⁰⁹ The Sherman Act was enacted specifically to “assure customers the benefits of price competition,” which plaintiffs could easily allege are lacking in the App Store given that Apple has no direct competitors who sell iPhone apps.²¹⁰ Indeed, the “monopoly overcharge” which allegedly injured the iPhone users is of the type of competition-stifling harm that the antitrust laws were enacted specifically to prevent.²¹¹ Of course, the antitrust injury threshold question

205. See *supra* Section IV.A.2.

206. See *id.*

207. Blair & Harrison, *supra* note 156, at 17; see also Miller, *supra* note 102, at 220.

208. Blair & Harrison, *supra* note 156, at 26.

209. See *AGC*, 459 U.S. at 538.

210. See *id.*

211. See Miller, *supra* note 102, at 209 n.101.

would not be dispositive of the merits of the actual claim, and Apple could still defeat these allegations at later stages in the litigation.

Moving to the next step of the AGC analysis, the Court would weigh several factors, beginning with the directness of the injury. It is likely that this factor would have cut against the plaintiffs for the reasons that Justice Gorsuch laid out in the dissent.²¹² If the plaintiff iPhone users were harmed by Apple's commission, the app developers must have passed the commission on to them, since the commission fell on the developers first.²¹³ In that way, the iPhone users were one step removed from the overcharge, and the injury would not be considered "direct." However, under the AGC Test, indirectly suffering an injury is not a per se bar against recovery; it is merely a factor that weighs against it.

The remaining factors would likely have cut in the iPhone users' favor. For instance, there were not "more direct victims of the alleged conspiracy" whose own self-interest might have motivated them to sue.²¹⁴ The only potential option was the app developers, who arguably had an incentive not to sue.²¹⁵ Not only would they have faced the risk of their supplier relationship with Apple souring, but they also likely benefitted if Apple engaged in anticompetitive conduct because they shared a portion of the monopoly overcharge.²¹⁶

Finally, the potential complexity of a trial would have also likely cut in the plaintiffs' favor. Bruce Kobayashi argues that the pass-through analysis in *Apple v. Pepper* is "neither complex nor speculative."²¹⁷ This is for two main reasons: first, the costs that the developers incur to make and update the apps are "largely fixed with respect to output," making the marginal cost of producing, distributing, and maintaining another "copy" of the app zero.²¹⁸ Second, Kobayashi argues that the effect of Apple's thirty-percent commission—an "ad valorem royalty" as opposed to a "unit royalt[y]"²¹⁹—on the optimal price of an app is zero.²²⁰ In the absence of any commission collected by Apple, the app developers will set the price of the app to maximize

212. *Apple, Inc. v. Pepper*, 139 S. Ct. 1514, 1528 (2019) (Gorsuch, J., dissenting).

213. *Id.*

214. *See AGC*, 459 U.S. at 542–45.

215. *See supra* Section IV.A.3.

216. *See supra* Section IV.A.3.

217. Kobayashi & Wright, *supra* note 73, at 261.

218. *See id.* at 261–62.

219. An ad valorem royalty is a tax based on a fixed percentage of the value of a product, while a unit royalty is a tax in a fixed amount per product sold. *See id.* at 254.

220. *Id.* at 262.

total revenue.²²¹ However, Kobayashi argues that the app developers, when faced with Apple's thirty-percent commission, will lower the gross price of the app to the same level as in the absence of the commission in order to maximize total revenue by increasing the costs of apps sold.²²²

Though the plaintiffs would certainly introduce contrary economic evidence to counter the Kobayashi argument, the simplicity of Kobayashi's models and the standard nature of the dispute do not suggest that the evidence that either side presents would exceed "judicially manageable limits" beyond what is typical of antitrust cases.²²³ Thus, the complexity factor would likely have favored the plaintiffs. Under this application of the AGC Test factors, it is likely that the Court would have found that the iPhone users, even if they were indirect purchasers, had standing to sue Apple.

V. CONCLUSION

Not only does Illinois Brick's indirect purchaser rule contradict the plain text of the Clayton Act, but the outdated policy rationales motivating the decision no longer justify continued adherence to its arbitrary restriction. *Apple v. Pepper* is a prime example of how Illinois Brick has outlived its usefulness in application to two-sided markets, an increasingly common feature in the modern economy. The AGC Remoteness Test is just one example of a potential alternative to the indirect purchaser rule that would allow courts to analyze standing in light of the unique features of two-sided marketplaces. In light of the heightened antitrust scrutiny of major digital platforms, courts should be prepared with the proper tools for analyzing standing issues in two-sided marketplace antitrust cases.

221. *See id.*

222. *See id.* at 266.

223. *See* *Associated Gen. Contractors v. Cal. State Council of Carpenters*, 459 U.S. 519, 543 (1983).