36:4 BERKELEY TECHNOLOGY LAW JOURNAL

2021

Pages

1341

to

1752

Berkeley Technology Law Journal

Volume 36, 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.

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 36 NUMBER 4: ANNUAL REVIEW 2021

TABLE OF CONTENTS

ARTICLES

SUBSCRIBER INFORMATION

The *Berkeley Technology Law Journal* (ISSN1086-3818), a continuation of the *High Technology Law Journal* effective Volume 11, is edited by the students of the University of California, Berkeley, School of Law 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 36 BERKELEY TECH. L.J. 4 (2021).

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

WHITE & CASE LLP

Benefactors

BAKER BOTTS LLP

COVINGTON & BURLING LLP

FENWICK & WEST LLP

FISH & RICHARDSON P.C.

KIRKLAND & ELLIS LLP

LATHAM & WATKINS LLP

MCDERMOTT WILL & EMERY LLP

MORRISON & FOERSTER LLP

ORRICK HERRINGTON & SUTCLIFFE LLP

PAUL HASTINGS LLP

SIDLEY AUSTIN LLP

WEIL, GOTSHAL & MANGES LLP

WILMER CUTLER PICKERING HALE AND DORR LLP

WILSON SONSINI GOODRICH & ROSATI

WINSTON & STRAWN LLP

Corporate, Government, Individual, and Foundation Sponsors

ATLASSIAN

BOIES SCHILLER & FLEXNER LLP

CISCO SYSTEMS, INC.

THE CITIZEN LAB

COMCAST CABLE

CORNERSTONE RESEARCH

DARTS IP

GEN LAW FIRM

GOODWIN PROCTER LLP

GOOGLE INC.

INTEL CORPORATION

INVENTIONSHARE INC.

JENNER & BLOCK

KILBURN & STRODE

LITINOMICS, INC.

MARKS & CLERK LLP

MICROSOFT CORPORATION

MOZILLA CORPORATION

NOKIA CORPORATION

PALANTIR TECHNOLOGIES

QUALCOMM INCORPORATED

RLM TRIALGRAPHIX

STARZ

TYSON & MENDES

UNIFY CONSULTING

VIA LICENSING CORPORATION

VYNL

WESTERN DIGITAL

Members

BAKER & MCKENZIE LLP

BEIJING EAST IP

DESMARAIS LLP

DURIE TANGRI LLP

GREENBERG TRAURIG LLP

GTC LAW GROUP LLP & AFFILIATES

HAYNES AND BOONE, LLP

IRELL & MANELLA LLP

KEKER VAN NEST & PETERS LLP

KILPATRICK TOWNSEND & STOCKTON LLP

KNOBBE MARTENS LLP

MORGAN LEWIS & BROCKIUS

ROBINS KAPLAN, MILLER & CIRESI LLP

TENSEGRITY LAW GROUP LLP

VAN PELT, YI & JAMES LLP

WANHUIDA INTELLECTUAL PROPERTY

WILLKIE FARR & GALLAGHER LLP

WOMBLE BOND DICKINSON LLP

BOARD OF EDITORS 2020–2021

MEMBERSHIP

Vol. 36 No. 4

BTLJ ADVISORY BOARD

BERKELEY CENTER FOR LAW & TECHNOLOGY 2020–2021

HOW JUDICIAL REVIEW PROMOTES A "THRYV-ING" PATENT SYSTEM

Angela Chen Griggs†

I. INTRODUCTION

Our patent system is a mess. New disruptive technologies, an ever-increasing number of patents, and bad actors who try to exploit the system are just some of the issues that plague the current patent law regime.[1] One of the most recent controversies surrounds a new way to challenge the validity of a patent, a way that could have wide-ranging implications on patent litigation and innovation incentives.[2]

Historically, the United States Patent and Trademark Office (PTO) played a limited role in post-issuance procedures and was not involved in enforcement of patents or determinations of validity after-the-fact.[3] This post-issuance role has expanded over the past few decades, and this expansion is controversial precisely because it departs from the PTO's historically limited role.[4] In 2011, the Leahy-Smith America Invents Act (AIA) revitalized the U.S. patent system.[5] The AIA created several new post-grant patent review proceedings, including inter partes review (IPR), a procedure used by petitioners to challenge the validity of an already-issued patent.[6] These post-grant challenges are heard by the newly re-named Patent Trial and Appeal Board (PTAB or "Board") and are required by statute to be filed within one year of the start of litigation between the parties. But in *Thryv Inc. v. Click-To-Call Technologies*, the Supreme

---

DOI: https://doi.org/10.15779/Z386D5PC21

† J.D., 2022, University of California, Berkeley School of Law.

1. *See* Paul R. Gugliuzza, *Patent Litigation Reform: The Courts, Congress, and the Federal Rules of Civil Procedure*, 95 B.U. L. REV. 279, 279–280 (2015).

2. *See, e.g.*, Brief of J. Kenneth Blackwell et al. as Amici Curiae Supporting Petitioner at 21–25, Oil States Energy Servs., LLC v. Greene's Energy Grp., LLC, 138 S. Ct. 1365 (2018) (No. 16-712), 2017 WL 3888214.

3. *See* McCormick Harvesting Mach. Co. v. C. Aultman & Co., 169 U.S. 606, 608–09 (1898), *superseded by statute*, Patent Act of 1952, Pub. L. No. 82-593, 66 Stat. 792, *as recognized in Oil States*, 138 S. Ct. at 1376.

4. *See, e.g.*, Brief of J. Kenneth Blackwell et al., *supra* note 2, at 21–25.

5. U.S. PAT. & TRADEMARK OFF., STUDY AND REPORT ON THE IMPLEMENTATION OF THE LEAHY-SMITH AMERICA INVENTS ACT at 12 (2015).

6. Leahy-Smith America Invents Act, Pub. L. No. 112-29, 125 Stat. 284 (2011) (codified as amended in scattered sections of 35 U.S.C.).

Court held that decisions by the PTAB that ignore the statutorily defined one-year time bar of IPR proceedings are not appealable.[7] This holding is the latest in a string of cases in which the Supreme Court has gradually given more and more deference to the PTO and PTAB, pushing the boundaries of long-held traditions like judicial review.[8] This could potentially change litigant behavior and the patent litigation landscape for the worse.

Just how did *Thryv*[9] potentially throw open the doors to greater deference to the PTO and PTAB? This Note seeks to answer this question, while arguing that those doors should be closed in order to protect patent owners. Part II gives a primer on the key components of post-grant patent review—the PTO/PTAB, IPR proceedings, and administrative law principles—including a discussion of the history and development of each and how these components have culminated in growing judicial deference to the PTO. Part III analyzes the *Thryv* case opinion, focusing on comparing the majority's and dissent's underlying reasons for their interpretation of § 314(d) and how it does or does not apply to § 315(b)'s time bar provision. Though at first glance this case may seem like "small potatoes,"[10] the implications from *Thryv* have far-reaching effects on both litigation strategy and the bigger question of separation of powers. Part IV argues that the majority's decision in *Thryv* creates a lack of judicial review of PTAB decisions. This leads to (1) improper strategic gaming during litigation; (2) undermining predictability and stability in the patent system, ultimately disincentivizing inventors from participating in the first place; and (3) unacceptable administrative law deference accorded to the PTO, which at its core does not operate as a true administrative agency. A contract law theory demonstrates how the PTO's role is not to regulate the patent system but rather to serve as an agent in the offer and acceptance of a contract. In light of this understanding, this Note proposes that courts should refer back to a *de novo* standard of review when reviewing questions of law and mixed questions of law and fact from the PTO. Because a complete overhaul of the Patent Act is unlikely, a change to judicial review standards is the best avenue for this proposal.

7. Thryv, Inc. v. Click-To-Call Techs., LP, 140 S. Ct. 1367, 1369 (2020).

8. *Id.*; *see* Cuozzo Speed Techs., LLC v. Lee, 136 S. Ct. 2131 (2016).

9. Pronounced like "thrive."

10. This quote is a reference to Chief Justice Roberts's comment during oral arguments, in which he referred to the underlying issue of whether the agency should institute an inter partes review as "small potatoes" because the decision of patent validity is ultimately judicially reviewable. *See* Transcript of Oral Argument at 16, *Thryv*, 140 S. Ct. 1367 (No. 18-916).

Part V summarizes how the current IPR system fails to protect patent owners, as well as the various loopholes that litigants can use to their advantage. It then concludes that judicial review is the best way to restore balance to the system.

II. BACKGROUND

A. THE PTO AND THE PTAB

Created by the Patent Act of 1790, the United States Patent and Trademark Office has existed in some form since the founding of this nation.[11] The PTO currently operates as part of the Department of Commerce and has two basic statutory responsibilities: (1) "granting and issuing of patents and the registration of trademarks" and (2) "disseminating to the public information with respect to patents and trademarks."[12] The PTO Director ("Director") is "responsible for providing policy direction and management supervision for the Office" and is appointed by the President with advice and consent from the Senate.[13]

The PTAB is the PTO's adjudicative body.[14] It is charged with deciding appeals from the decisions of patent examiners and adjudicating the patentability of issued patents challenged by third parties in post-grant proceedings.[15] The Board has existed in some form since the 1800s and was known as the Board of Patent Appeals and Interferences before the passing of the AIA.[16] Currently, the Board consists of both statutory members and administrative patent judges.[17] The statutory members of the Board include the Director, Deputy Director, Commissioner for Patents, and Commissioner for Trademarks.[18] Administrative patent judges are appointed by the Secretary of Commerce and are legally and technically trained.[19] These judges have extensive patent legal experience before their appointment to the Board, whether in private

11. *The Early History of the U.S. Patent Office*, LEXISNEXIS INTELL. PROP. (Oct. 26, 2015), https://www.lexisnexisip.com/knowledge-center/the-early-history-of-the-u-s-patent-office.

12. 35 U.S.C. § 2(a).

13. *Id.* § 3(a)(1), 3(a)(2)(A).

14. *See* Janet Gongola, *The Patent Trial and Appeal Board: Who Are They and What do They o?*, INVENTORS EYE (2019), https://www.uspto.gov/learning-and-resources/newsletter/inventors-eye/patent-trial-and-appeal-board-who-are-they-and-what.

15. *Id.*

16. *Id.*

17. *Id.*

18. *New to PTAB*, U.S. PAT. & TRADEMARK OFF., https://www.uspto.gov/patents-application-process/patent-trial-and-appeal-board/ptab-inventors.

19. *Id.*

practice, government practice, or as in-house counsel.[20] The Patent Act requires that PTAB judges hear proceedings in panels of "at least 3 members," who are chosen by the Director in each case.[21] Because of this power, critics have been fast to point out how the Director's influence may taint the neutrality of the selected panel or otherwise affect the PTAB's adjudicatory decision-making.[22]

B. IPR PROCEEDINGS

In addition to reforming the PTAB, the AIA also created three new fast-track proceedings for challenging issued patents: inter partes review (IPR), post-grant review, and covered business method review.[23] The goal in creating these new post-grant reviews was to reduce litigation costs, provide more certainty, and improve patent quality.[24]

The inter partes review replaces the previous inter partes reexamination procedure and converts the process from one of examination to one of adjudication.[25] Unlike previous post-grant review proceedings, the IPR proceeding offers a much more robust process, as *any* third party can request an IPR of a granted patent.[26] As the name suggests, an IPR allows for the petitioner (i.e., the challenger to patent validity and the outside third party) to also participate in the proceedings.[27] In previous post-grant reviews, only the patentee was allowed to participate.[28] The petitioner starts the proceeding by filing a petition,[29] which can *only* challenge the patentability of claims under 35 U.S.C §§ 102 and 103 *solely* on the basis of patents or printed publications.[30] IPRs are available

20. *Id.*

21. 35 U.S.C. § 6(c).

22. *See generally* John M. Golden, *PTO Panel Stacking: Unblessed by the Federal Circuit and Likely Unlawful*, 104 IOWA L. REV. 2447 (2019) (arguing that allowing the PTO Director to configure a rehearing panel to achieve a predetermined result presents serious constitutional questions under the Due Process Clause and is unauthorized by the Patent Act).

23. Leahy-Smith America Invents Act, Pub. L. No. 112-29, 125 Stat. 284 (2011) (codified as amended in scattered sections of Title 35 of the United States Code). This Note will not address AIA derivation proceedings.

24. Gregory Dolin & Irina D. Manta, *Taking Patents*, 73 WASH. & LEE L. REV. 719, 721–22 (2016).

25. *See* Abbot Lab'ys v. Cordis Corp., 710 F.3d 1318, 1326 (Fed. Cir. 2013) ("The purpose of this reform was to 'convert[ ] inter partes reexamination from an examinational to an adjudicative proceeding . . . .'" (quoting H.R. REP. NO. 112-98, pt. 1, at 46–47 (2011))).

26. 35 U.S.C. § 311(a).

27. *Id.*

28. Eugene T. Perez, *Chart Summarizing Different Post Grant Proceedings*, BSKB POST GRANT PROC., https://postgrantproceedings.com/aia-uspto-links/pgr-chart (last visited Apr. 7, 2022).

29. 35 U.S.C. § 311(a).

30. *Id.* § 311(b).

for all issued patents with a few limits. First, a party is barred from filing a petition if the party previously filed a declaratory judgment action that challenged the validity of a claim of the patent.[31] In other words, if a party has already challenged the validity of a patent in federal courts, they cannot seek additional relief at the PTO. Notably, this bar is not triggered where a party challenges only their alleged infringement of a patent (i.e., the party only claims that their product does not fall within the scope of the patent's claims).[32] Second, a petitioner is limited to filing a petition within a year "after the date on which the petitioner, real party in interest, or privy of the petitioner is served with a complaint alleging infringement of the patent."[33] Seemingly a straightforward rule, this time bar provision has proven quite the opposite, and its controversy and downstream effects are the focus of this Note.

After the proper and timely filing of the petition, the patent owner has three months to file a preliminary response to the petition.[34] The response lays out reasons why the IPR should not be instituted (a decision to institute starts the IPR proceeding) and can include supporting evidence.[35] The PTAB must decide within three months from receipt of the response (or, if no response is received, from expiration of the time to file one) whether to grant the petition and institute an IPR proceeding.[36] The Director of the PTO *may* institute an IPR if "there is a reasonable likelihood that the petitioner would prevail with respect to at least [one] of the claims challenged in the petition."[37] If the PTO institutes IPR, the proceeding is conducted before a panel of three technically trained PTAB administrative patent judges.[38] The parties are allowed to take limited expert discovery and respond to each other's arguments, similar to district court proceedings.[39] Parties also have a right to an oral hearing.[40] The PTAB must issue a final, written decision within one year from the date of

31. *Id.* § 315(a)(1).

32. *See* Ariosa Diagnostics v. ISIS Innovation Ltd., No. IPR2012-00022 (MPT), 2013 WL 2181162, at 7–8 (P.T.A.B. Feb. 12, 2013) (holding that a declaratory judgment action for non-infringement did not bar the filing of an IPR petition, nor did the assertion of an affirmative defense of invalidity in response to the patentee's counterclaim of infringement).

33. 35 U.S.C. § 315(b).

34. 37 C.F.R. § 42.107(b) (2020).

35. *Id.* § 42.107(a), (c).

36. 35 U.S.C. § 314(b).

37. *Id.* § 314(a).

38. *Id.* §§ 6(a)–(c), 311.

39. *Id.* § 316(a)(5).

40. *Id.* § 316(a)(10).

institution.[41] After receiving the final written decision from the PTAB, the parties have the option to appeal to the Federal Circuit.[42]

If instituted, an IPR will typically conclude within 18 months of the filing date.[43] In contrast, the average time from filing to the conclusion of the previous inter partes reexamination procedure ranged from 28.9 to 41.7 months.[44] In sum, an IPR provides a path to receive expert guidance from the PTO under a more-accelerated timeline than the previous inter partes reexamination proceeding.

C. THE ADMINISTRATIVE STATE

Many believe that the PTO operates as an administrative agency and think administrative law principles should be applied.[45] However, this thinking is flawed as there are critical differences between the PTO and other administrative agencies.[46] These differences impact how we understand judicial review of agency decision-making.

1. *Historical Development*

American administrative law has a complex history. Some hail it as a visionary solution to Congress's ever-expanding duties.[47] Others believe that the administrative state has grown to be problematic, both constitutionally and practically, imposing on private rights and freedoms and restricting the relationship between government and the society it seeks to govern.[48]

The late nineteenth century saw the Industrial Revolution transform American society as new industries like railroads quite literally redefined the physical and economic landscapes.[49] These new industries required government oversight and regulation, which led Congress to create administrative

41. *Id.* § 316(a)(11). This one-year period can be extended for good cause. *Id.*

42. *Id.* § 319.

43. *Id.* § 316(a)(11).

44. *See* Changes to Implement Inter Partes Review Proceedings, Post-Grant Review Proceedings, and Transitional Program for Covered Business Method Patents, 77 Fed. Reg. 48680, 48721 (Aug. 14, 2012).

45. *See, e.g.*, Stuart Minor Benjamin & Arti K. Rai, *Who's Afraid of the APA? What the Patent System can Learn from Administrative Law*, 95 GEO. L.J. 269 (2007).

46. This belief can be seen in early writing on administrative law that reflects the idea that administrative law principles should only apply when agencies are delegated rulemaking authority and act in a regulatory capacity. *See, e.g.*, Fed. Trade Comm'n v. Ruberoid Co., 343 U.S. 470, 490–91 (1952) (Jackson, J., dissenting).

47. *See* Orin S. Kerr, *Rethinking Patent Law in the Administrative State*, 42 WM. & MARY L. REV. 127, 157 (2000).

48. Christopher DeMuth, *Can the Administrative State be Tamed?*, 8 J. LEGAL ANALYSIS 121, 122 (2016).

49. *See id.* at 123.

agencies to both monitor and regulate these growing industries.[50] Unlike previous agencies that merely followed the laws set by Congress and the courts (the PTO being one of them), the new regulatory agencies had lawmaking and enforcement power.[51] This effectively allowed the transfer of power from the legislative branch to the executive branch.[52] This transfer was largely justified by the belief that these agencies contained subject matter experts who could act in broad and imaginative ways for the public interest.[53]

With these new responsibilities came new procedures. Administrative law is governed largely by the Administrative Procedure Act of 1946 ("APA"), whose long-reaching effects govern most administrative disputes to this day.[54] In general, the APA created three categories of administrative action: (1) rulemaking, in which the agencies impose regulations; (2) adjudication, in which agencies resolve disputes by factfinding and making conclusions of law; and (3) discretionary decision making.[55] For decisions that affect interests of specific parties (e.g., granting of licenses, imposing service controls, settling disputes among parties), agencies generally follow formal adjudication procedures but with much looser evidentiary standards than judicial trials.[56] Nevertheless, formal rulemaking only takes place when a statute authorizing a new rule requires it, and only then when the statute expressly states that rulemaking must take place "on the record."[57] Adjudications are heard by administrative law judges, and final decisions are subject to judicial review unless otherwise specified by statute.[58] Agencies typically create new rules and modify, amend, or repeal existing rules through the informal rulemaking process, which requires notice to the public and an opportunity to comment on the proposed rule.[59] These rules are subject to judicial review under an "arbitrary and capricious, an abuse of discretion, or otherwise not in accordance with law" standard.[60]

---

50. Kerr, *supra* note 47, at 157–58; *see* Felix Frankfurter, *The Task of Administrative Law*, 75 U. PA. L. REV. 614, 617–18 (1927).

51. Kerr, *supra* note 47, at 159.

52. *Id.*

53. *Id.*

54. Pub. L. No. 79-404, 60 Stat. 237 (codified as amended in scattered sections of 5 U.S.C.).

55. Roni Elias, *The Legislative History of the Administrative Procedure Act*, 27 FORDHAM ENV'T L. REV. 207, 214 (2015).

56. DeMuth, *supra* note 48, at 123–24.

57. *Id.* (quoting United States v. Fla. E. Coast Ry. Co., 410 U.S. 224, 234 (1973)).

58. *Id.*

59. *Formal Rulemaking*, JUSTIA (2018), https://www.justia.com/administrative-law/rulemaking-writing-agency-regulations/formal-rulemaking/#:~:text=The%20federal%20Administrative%20Procedures%20Act,requires%20notice%20to%20the%20public.

60. DeMuth, *supra* note 48, at 124.

This standard is much more deferential than the normal *de novo* judicial review standard used by appellate courts when reviewing lower court decisions.

2. *Judicial Review Standards Under the APA*

The deferential review granted to administrative agencies is unwarranted for the PTO because the PTO does not hold actual rulemaking authority.

Judicial review of an administrative agency's decision for a question of law is reviewed for reasonableness, a standard commonly referred to as "*Chevron* deference."[61] In the Supreme Court's 1984 decision in *Chevron, U.S.A., Inc. v. Natural Resources Defense Council, Inc.*, the Court outlined a two-step framework for determining when a reviewing court should defer to the agency's interpretation of a statute the agency was charged with administering.[62] Under step one, courts should examine "whether Congress has directly spoken to the precise question at issue,"[63] i.e., whether the statute is ambiguous. If so, this is the end of the inquiry and courts must enforce the "unambiguously expressed intent of Congress."[64] If the statute is silent or ambiguous on the issue, courts must proceed to step two, which requires courts to defer to an agency's interpretation of the statutory text so long as it is reasonable, even if the court would have reached a contrary conclusion.[65]

Applying *Chevron* deference essentially declares that the agency is the authoritative interpreter of that area of law.[66] The threshold question in *Chevron* requires a court to determine if the deference accorded should apply at all; the deference is limited to agency interpretation of statutes that the agency itself administers.[67] This analysis is further limited to what is sometimes referred to as *Chevron* step zero, which asks whether Congress has delegated that authority to the agency.[68]

---

61. Chevron, U.S.A., Inc. v. Nat. Res. Def. Council, 467 U.S. 837, 842–43 (1984); Benjamin & Rai, *supra* note 45, at 295.

62. *See Chevron*, 467 U.S. at 842–43.

63. *Id.* at 842.

64. *Id.* at 842–43.

65. *Id.* at 843.

66. *See, e.g.*, Nat'l Cable & Telecomms. Ass'n v. Brand X Internet Servs., 545 U.S. 967, 983 (2005) (noting that when *Chevron* applies, the agency—not the reviewing court—is "the authoritative interpreter (within the limits of reason) of such statutes").

67. Adams Fruit Co. v. Barrett, 494 U.S. 638, 649 (1990) ("A precondition to deference under *Chevron* is a congressional delegation of administrative authority.").

68. Thomas W. Merrill & Kristin E. Hickman, Chevron*'s Domain*, 89 GEO. L.J. 833, 836 (2001).

There is a large debate in various scholarly circles about whether the PTO, and by extension the PTAB, should be accorded *Chevron* deference.[69] Step zero of *Chevron* says that deference only applies to interpretations of text that the agency has been "entrusted to administer."[70] Yet, others argue that the PTO has historically never been entrusted to interpret the text of the Patent Act and thus should not be granted *Chevron* deference.[71] The question of which standard of deference should be applied to the PTO's interpretation of the time bar provision is a crucial determination in both the majority view and minority view in *Thryv* and sets up the bigger conversation of whether the PTO should be granted deference at all.

D. PATENT LAW AND THE APA

1. *Partial Deference: Zurko's Influence*

While *Chevron* has largely been accepted and applied to administrative agencies, the Court of Appeals for the Federal Circuit ("Federal Circuit") has historically declined to apply *Chevron* deference to the PTO.[72] This decision was largely predicated on the fact that despite having the power to adjudicate patent denials, PTAB proceedings did not fully resemble formal trial-like adjudications.[73]

It was not until the Supreme Court's 1999 decision in *Dickinson v. Zurko* that the Federal Circuit started to apply administrative law to patent law.[74] In that decision, the Court held that the lower APA standard of "substantial evidence" applies to review of the PTO for questions of fact, in place of clear error.[75] Notably, the Federal Circuit has declined to extend the APA's deferential review standard to questions of law, arguing that full *Chevron* deference

69. *See generally* John M. Golden, *The USPTO's Soft Power: Who Needs* Chevron *Deference?*, 66 SMU L. REV. 541 (2013) (describing this debate as a question of substantive rulemaking authority and further proposing that the PTO can be effective even without access to *Chevron* deference).

70. *Chevron*, 467 U.S. at 844.

71. Kerr, *supra* note 47, at 165.

72. *See* Melissa F. Wasserman, *The Changing Guard of Patent Law:* Chevron *Deference for the PTO*, 54 WM. & MARY L. REV. 1959, 1975 (2013).

73. *Id.* at 1973.

74. *See* Dickinson v. Zurko, 527 U.S. 150, 152 (1999).

75. *See id.*; *see also In re* Alton, 76 F.3d 1168, 1172 (Fed. Cir. 1996) ("We review questions of fact arising from Board rejections under a clearly erroneous standard. We review questions of law *de novo*." (citation omitted)).

is not warranted because the Patent Act did not give the PTO rulemaking authority on substantive issues of patent law.[76] However, recent Supreme Court decisions have tested that long-held assumption.[77]

2. *A Move Towards Full Deference? Cuozzo and Beyond*

Perhaps the most influential driver for the majority's holding in *Thryv* comes from *Cuozzo Speed Technologies v. Lee*.[78] Four years before the Court addressed the *Thryv* case, the Court heard a case concerning the scope and reviewability of IPR proceedings.[79] Briefly, *Cuozzo* involved an appeal of the Board's decisions to institute review of three claims, which ultimately led to the claims' invalidations.[80] On appeal to the Federal Circuit, Cuozzo claimed that the Board erred in instituting review of two of the claims because they had not been challenged with sufficient "particularity" pursuant to § 312(a)(3).[81] The Federal Circuit cited to § 314(d), which says that the Director's decision whether to institute an IPR is "non-appealable."[82] The Supreme Court agreed, citing to both *Chevron* and *Mead* in determining that the language of the statute, as well as the practicalities of permitting challenges after institution, would have the unintended effect of validating a patent even though the Board had determined that it had been issued in error.[83]

Many federal agencies have statutes that provide for judicial review of some agency decisions but preclude review of others.[84] For over 200 years, the Court has applied a strong presumption of judicial review and has construed preclusion statutes very narrowly.[85] The *Cuozzo* decision signals a potential that the Court is more willing to apply *Chevron* deference to the PTO, at least as far

76. *See In re* Swanson, 540 F.3d 1368, 1374 n.3 (Fed. Cir. 2008) ("[T]he Board's statutory interpretation in a particular case is given no deference . . . .").

77. *See infra* Section II.D.2.

78. *See generally* Cuozzo Speed Techs., LLC v. Lee, 136 S. Ct. 2131 (2016).

79. *Id.*

80. *Id.* at 2138–39.

81. *See id.* at 2139.

82. *Id.* at 2136.

83. Ronald Mann, *Opinion Analysis: Justices Validate PTO Rules for Inter Partes Review*, SCOTUSBLOG (June 20, 2016, 9:52 PM), https://www.scotusblog.com/2016/06/opinion-analysis-justices-validate-pto-rules-for-inter-partes-review. *But see* SAS Inst., Inc. v. Iancu, 138 S. Ct. 1348, 1355 (2018) (claiming, in a majority opinion by Justice Gorsuch, that Congress had intentionally designed IPR as a "party-directed, adversarial process" that mimics civil litigation and therefore does not warrant applying *Chevron* deference).

84. Elias, *supra* note 55, at 221.

85. *See* Thryv, Inc. v. Click-To-Call Techs., LP, 140 S. Ct. 1367, 1378 (2020) (Gorsuch, J., dissenting).

as questions that are "closely tied to the application and interpretation of statutes related to the [PTO]'s decision."[86] Unfortunately, the Court has provided little guidance as to what types of questions are closely tied with the decision to institute.

Some scholars argue that the passing of the AIA and the creation of the PTAB have expanded rulemaking power for the PTO.[87] The PTAB can resolve challenges in adjudicatory proceedings and has limited power to make these resolutions precedential, thereby affecting applicant behavior.[88] However, like the majority in *Thryv*, these scholars are myopically focused on the nebulous goal of "weed[ing] out bad patent claims."[89] Notwithstanding that the focus should be to not grant bad patents in the first place, the arguments for expanding *Chevron* deference for the PTO fail to recognize the impact it has on driving innovation.

III. THRYV INC. V. CLICK-TO-CALL TECHNOLOGIES

The *Thryv* case is the Court's latest foray into judicial reviewability of PTAB decision-making. The majority's decision is rooted in the belief that one of the main goals of the AIA is to create a forum to efficiently challenge the validity of patents, thereby improving the overall quality of existing patents. The dissent's argument looks at broader policy concerns and how IPRs can be used to harass patent owners. These competing concerns between efficiency and downstream effects on patentees sets the stage for Part IV's discussion of how deference to the PTO can deter innovation.

Three points are key before delving into this Part. First, IPR proceedings were created with efficiency and administrability in mind, primarily in an effort to have patent challenges adjudicated in a forum similar yet superior to what district courts can provide. Second, administrative law principles are based on assumptions of expertise and delegation of regulatory rulemaking power. And third, the PTO has never been fully delegated regulatory rulemaking power.

A. UNDERLYING DISPUTE AND PROCEDURAL HISTORY

On April 20, 2020, the Supreme Court held in *Thryv, Inc. v. Click-To-Call Technologies* that the Director's decision as to when and if the time bar under

86. *Cuozzo*, 136 S. Ct. at 2141.

87. Rebecca S. Eisenberg, *A Functional Approach to Judicial Review of PTAB Rulings on Mixed Questions of Law and Fact*, 104 IOWA L. REV. 2387, 2389–90 (2019).

88. Saurabh Vishnubhakat, *Disguised Patent Policymaking*, 76 WASH. & LEE L. REV. 1667, 1719–20 (2019).

89. *See Thryv*, 140 S. Ct. at 1374.

§ 315(b) was triggered is non-appealable.[90] Justice Ginsburg delivered the opinion of the Court and was joined by Chief Justice Roberts and Justices Breyer, Kagan, Kavanaugh, Thomas, and Alito (with the latter two refusing to join Part III-C).[91] Justice Gorsuch wrote a dissenting opinion, which Justice Sotomayor joined in part.[92]

The underlying dispute in this case concerned a patent, owned by respondent Click-to-Call, relating to a technology for anonymous telephone calls.[93] Thryv Inc. petitioned for an IPR in 2013, challenging several of the patent's claims as invalid.[94] Click-to-Call referred the Board to an infringement suit that was filed back in 2001 as evidence that Thryv was "served with a complaint alleging infringement of the patent."[95] Click-to-Call argued that this infringement suit triggered the § 315(b) time bar, which provides in relevant part: "An inter partes review may not be instituted if the petition requesting the proceeding is filed more than 1 year after the date on which the petitioner, real party in interest, or privy of the petitioner is served with a complaint alleging infringement of the patent."[96] The Board disagreed and said that a complaint dismissed without prejudice does not trigger the one-year limit to file an IPR.[97] The Board instituted the review and ultimately canceled thirteen claims as obvious or lacking novelty.[98] Click-to-Call challenged the Board's determination, arguing that the bar on appeal under § 314(d) is limited to the agency's threshold determination under § 314(a) that addresses only the question of whether the petitioner has a reasonable likelihood of prevailing.[99] The Federal Circuit dismissed the appeal for lack of jurisdiction due to § 314(d)'s bar on appeal, which provides that "[t]he determination by the [PTO] Director whether to institute an inter partes review under this section shall be final and nonappealable."[100] Click-to-Call then petitioned for certiorari, but the Supreme Court cited to their intervening decision in *Cuozzo*, vacated the judgment, and remanded the case back down to the Federal Circuit, which again dismissed the appeal on the same ground.[101]

---

90. *Id.* at 1377.
91. *Id.* at 1369.
92. *Id.* at 1378 (Gorsuch, J., dissenting). Justice Sotomayor joined Justice Gorsuch's dissent, except as to Part V.
93. *Id.* at 1371 (majority opinion).
94. *Id.*
95. *Id.* (quoting 35 U.S.C. § 315(b)).
96. 35 U.S.C. § 315(b).
97. *Thryv*, 140 S. Ct. at 1371.
98. *Id.*
99. *Id.*
100. *Id.* at 1370 (quoting 35 U.S.C. § 314(d)).
101. *Id.*

During this time, the Federal Circuit held en banc in another case that § 315(d)'s time bar determination is reviewable, despite the language of § 314(d).[102] The Federal Circuit determined that nothing displaces the usual presumption of favoring judicial review if a party appeals based on the time bar provision.[103] In light of this decision, the Federal Circuit reheard Click-to-Call's appeal and vacated the Board's final written decision, holding that the petition for IPR was untimely because the 2001 infringement complaint had started the one-year clock.[104] It is on this convoluted backdrop that Thryv petitioned the Supreme Court to hear the case.

B. THE MAJORITY WANTS TO KEEP OUT BAD PATENTS

Justice Ginsburg's majority opinion focused primarily on the statutory language of § 314(d), which renders "final and nonappealable" the Director's decision whether to institute an IPR.[105] The Court explained that the text of § 314(d) clearly states that determinations whether to institute IPR are final and non-appealable and "preclud[es] review of the Patent Office's institution decisions with sufficient clarity to overcome the strong presumption in favor of judicial review."'[106] The Court noted that its *Cuozzo* decision did not address whether § 314(d) would bar appeals reaching beyond the decision to institute an IPR; the Court declined to address whether appeals implicate constitutional questions.[107] Instead, the Court narrowed its holding to preclude appeal in instances where the PTAB's decision to institute an IPR involved questions that are "closely tied to the application and interpretation of statutes related to the Patent Office's decision to initiate inter partes review."[108]

The majority found that § 315(b)'s time limitation is an integral condition of institution of an IPR (i.e., a condition that determines whether an IPR proceeding should proceed) and sets forth a parameter under which an IPR may not be instituted.[109] Because the time bar provision "governs institution and nothing more," a challenge to a petitioner's timeliness in filing the petition falls

102. *Id.* at 1370; *see* Wi-Fi One, LLC v. Broadcom Corp., 878 F.3d 1364 (Fed. Cir. 2018) (en banc) (holding that, for § 315(b) appeals, § 314(d) does not dispel with the presumption of judicial review), *abrogated by Thryv*, 140 S. Ct. at 1367.

103. *Thryv*, 140 S. Ct. at 1372.

104. *Id.*

105. *Id.* at 1373.

106. *Id.* at 1373 (quoting Cuozzo Speed Techs., LLC v. Lee, 136 S. Ct. 2131, 2140–41 (2016)) (internal quotations omitted).

107. Cuozzo Speed Techs., LLC v. Lee, 136 S. Ct. 2131, 2141 (2016).

108. *Thryv*, 140 S. Ct. at 1373 (quoting *Cuozzo*, 136 S. Ct. at 2141).

109. *Id.*

under "'an ordinary dispute about the application of' an institution-related statute."[110] In essence, the Court held that the provisions of § 314(d) are not limited to a reasonable likelihood determination and instead encompass the *entire* determination to institute IPR.

To support its positions, the majority also pointed to the AIA's purpose and design of enabling a more efficient and streamlined patent system.[111] The majority reasoned that allowing § 315(b) appeals would go against that objective and waste resources on resolving patentability, potentially inducing patentees to use the appeal as a way to resurrect and save "bad patent claims."[112] This "unwind[ing] [of] the agency's merits decision" could potentially negate the Board's determination of patentability and "leave bad patents enforceable."[113] Moreover, the Court rationalized that allowing for judicial review of the time bar would not serve other goals, such as minimizing the burden and overlap between IPRs and district court litigation, something that both parties agreed is a crucial goal of IPR proceedings.[114]

The Court found Click-to-Call's argument that the bar on judicial review applies only to the threshold determination under § 314(a) unpersuasive in light of *Cuozzo*.[115] Nor was the Court persuaded by Click-to-Call's invocation of the decision in *SAS Institute Inc. v. Iancu*, where the Court held that if the Director initiates an IPR, the IPR must decide every claim challenged by the petitioner.[116] That decision, the Court said, was inapplicable to this case because *SAS* applies to challenges made *after* IPR institution, rather than the question of whether institution should have happened in the first place.[117]

Finally, the majority deftly dismissed Click-to-Call's alternative argument that § 319 permits appeal of the final written decision.[118] The Court held that respondents mischaracterized that provision in the sense that § 314(d) would still bar review of the time bar, since respondents' arguments still only concerned whether or not the Board should have instituted IPR.[119] The majority here brought multiple different rationales into its opinion and holding, including statutory interpretation, reading of congressional intent and purpose, and general policy concerns. In the end, the majority extended its *Cuozzo* holding

110. *Id.* (quoting *Cuozzo*, 136 S. Ct. at 2139).
111. *Id.* at 1374.
112. *Id.*
113. *See id.*
114. *Id.*
115. *Id.* at 1373–74.
116. *Id.* at 1376.
117. *Id.*
118. *Id.* at 1377.
119. *Id.*

and cabined the issue to specific timeframes: decisions made before institution of IPR versus decisions made after IPR has begun. However, the majority's decision left practitioners, patent owners, and challengers continually questioning the ever-changing legal effects of an already unstable patent system.

C. THE DISSENT WANTS TO PROTECT PATENT OWNERS

The *Thryv* dissent focused much of its analysis on upholding patent policy and long-held traditions of judicial review.[120] Justice Gorsuch authored the dissent with Justice Sotomayor joining in all but Part V. Justice Gorsuch stood alone in his views that patent rights are private, personal rights akin to property.[121]

First, the dissent was unable to find textual support in the statutory language for the majority's construction.[122] On its face, § 314(d) says that the Director's determination of whether to institute an IPR "under this section" is not reviewable.[123] The dissent argued that to interpret "under this section" to mean more than one determination invites both linguistic and practical nonsense.[124] Not only would a second determination require looking at a completely separate section, but § 314's "reasonable likelihood" determination relates to the *merits* and thus would naturally be reviewed by both the Board and the courts throughout the progression of the case.[125] Conversely, the Director's application of the time bar is a *procedural* issue.[126] Under the majority's view, application of § 315(b) effectively has no review. The Director has the final say on the matter, even if wrongly applied, and this could effectively close the IPR process altogether for some petitioners.[127]

Second, the dissent reminded the Court that the long-held presumption of judicial review serves an important purpose.[128] Here, the dissent posited that judicial review would provide some repose to the patent owner.[129] In doing so, the dissent recognized that "[p]atents typically last 20 years; what happens to the incentive to invent if litigation over them lasts even longer . . . ?"[130] The system of checks and balances between the three branches of government specifically tasked the judiciary with providing a remedy to aggrieved parties by

120. *Id.* at 1378–88 (Gorsuch, J., dissenting).
121. *Id.* at 1378.
122. *Id.* at 1380.
123. 35 U.S.C. § 314(d).
124. *Thryv*, 140 S. Ct. at 1381 (Gorsuch, J., dissenting).
125. *Id.*
126. *Id.*
127. *Id.*
128. *Id.* at 1383.
129. *Id.* at 1379.
130. *Id.* at 1379–80.

allowing review of decisions made in the other branches.[131] The majority's decision takes away this check on the system. Unlike the majority, the dissent rejected Thryv's argument that Congress's main purpose for creating the IPR was to "weed out poor quality patents."[132] Instead, the dissent said that Congress also sought to protect the rights of patent owners and preserve the value of their patents.[133] There is a balance, the dissent argued, whereby giving the PTAB power to cancel patents must be offset by at least some protections for patent owners.[134] Thryv asked the Court to consider the AIA as a whole in its argument but, in doing so, caused the Court to singularly focus on what the dissent deemed the efficient canceling of patents.[135]

Finally, the dissent addressed the issue of precedent, or rather the lack thereof.[136] The Court held in *Cuozzo* that claims closely related to the decision to institute IPR would be shielded from judicial review. Nonetheless, the dissent here felt that the strong presumption of judicial review should override the *Cuozzo* holding.[137] Standing alone in the final section of the dissent, Justice Gorsuch reiterated his view that the majority's decision "takes us further down the road of handing over judicial powers involving the disposition of individual rights to executive agency officials."[138] This is not the first time that Justice Gorsuch has expressed this view. He also dissented from the Court's decision in *Oil States Energy Services LLC v. Greene's Energy Group, LLC*.[139] There, the Justice lamented the consequences of the majority's decision that patent rights are public rights rather than private property, saying that society would not agree to allow the government to "cancel[]" a person's right to his farm, and thus society should not allow the Patent Office to do so for patents.[140] Justice Gorsuch worried that through this line of cases, deference had been misapplied and that in reality, the Court was abdicating its judicial duty to an agency that is wholly unable to provide inventors the type of protection that incentivizes

131. *Id.* at 1389 ("[T]he Executive Branch assumed responsibilities long reserved to the Judiciary. In so doing, we denied inventors the right to have their claims tried before independent judges and juries.").

132. *Id.* at 1384 (quoting Brief for Petitioner at 24, *Thryv*, 140 S. Ct. 1367 (No. 18-916)).

133. *Id.* at 1385.

134. *Id.* at 1384.

135. *Id.* at 1385.

136. *Id.* at 1385–86.

137. *Id.* at 1386.

138. *Id.* at 1387.

139. *See* Oil States Energy Servs., LLC v. Greene's Energy Grp., LLC, 138 S. Ct. 1365, 1380–81 (2018) (Gorsuch, J., dissenting).

140. *Id.*

the patent system.[141] Justice Gorsuch was particularly concerned that the majority's holding would subject patent owners to the mercy of a political bureaucratic system, without recourse in the courts.[142]

While the *Thryv* majority focused on the minutiae of the time bar and the immediate goals of the AIA, the dissent took a much broader view and focused on the bigger goals of the patent system as a whole. In a way, the dissent better understood that the ultimate goal of the patent system is to provide incentives for inventors to share their ideas and promote further innovation. The focus is the individual inventor, and it is important that they be provided a fair forum in which to protect their already-issued patent rights. This dynamic between inventors and incentives to invent is something the majority missed in its holding, focusing instead on the institutionalization of the patent system and the administration of patent laws.[143] The majority's systematic and rather clinical view of patent law did not factor in the behavior of the very people that make up the patent system. This can lead to consequences that threaten to further destabilize and undermine the patent regime.

IV. JUDICIAL REVIEW PROMOTES A FAIR AND STABLE PATENT SYSTEM

A lack of judicial review allows litigants to take advantage of IPR proceedings, especially with relation to parallel district court proceedings. Currently, rules surrounding IPRs lack certain limits, and this creates loopholes that litigants can take advantage of to continually try to invalidate patents. The concerns raised in Justice Gorsuch's *Thryv* dissent point to how this lack of judicial review leads to an unreliable patent system that can ultimately disincentivize innovation. A patent exceptionalism theory based in contract law, where patent law is exempt from administrative law deference principles, could help to address these concerns and provide some protection for patent owners against relentless challenges to their patents.

A. NO JUDICIAL REVIEW LEADS TO IMPROPER STRATEGIC GAMING

1. *How IPRs Intersect with District Courts*

The main goal of the AIA and of the new post-grant review process was not only efficiency, but also "effective[ness] at thwarting some strategic behavior in patent litigation."[144] With the advent of the new fast-tracked proceedings,

141. *Thryv*, 140 S. Ct. at 1386–87 (Gorsuch, J., dissenting).
142. *Id.* at 1378.
143. *See supra* Section III.B.
144. 152 CONG. REC. 16998 (2006) (statement of Sen. Patrick Leahy) (discussing S. 3818, 109th Cong. (2006), one of the many predecessor bills to the AIA).

litigants are now faced with the question of whether, when, and where to challenge the validity of patent claims.[145] Administrative ex post review by the PTAB was designed to be an efficient and preferable substitute for Article III litigation.[146] Notably, IPRs allow for the petitioner to potentially preemptively challenge patent validity at the PTAB before any district court litigation, or even to bring retaliatory challenges.[147] The interplay between an IPR and district court litigation is also exemplified in estoppel effects.[148] The final decision resulting from an IPR precludes the petitioner from asserting any claim the PTO, the federal courts, or the International Trade Commission that the petitioner raised or could have raised in the IPR proceeding.[149] If a petitioner has previously filed a civil action challenging a patent, then the petitioner cannot challenge that same patent in an IPR.[150] In addition, IPRs trigger automatic stays of co-pending or later-filed declaratory judgment litigation by that petitioner (or real party in interest).[151] If a party files a civil action in district court *after* the IPR petition, that action is automatically stayed.[152] The stay can only be lifted at the patent owner's request if the patent owner claims or counterclaims infringement against the petitioner or if the petitioner dismisses its civil action.[153] District courts also have the discretion to stay existing infringement litigation brought by a patent owner pending the outcome of an IPR.[154] In other words, there is no mandatory stay of district court litigation when a parallel IPR is filed. Consequently, stays are not always granted and often result in costly duplication, rather than efficiency.[155] This is especially true in districts that have gained a reputation for moving fast on patent validity determinations.[156]

---

145. *See* 35 U.S.C. § 315.

146. 152 CONG. REC. 16997 (2006) (statement of Sen. Orrin Hatch) ("I believe that by adopting a more robust post-grant review proceeding we are providing a more efficient means of challenging a patent's validity in an administrative proceeding.").

147. *See* Saurabh Vishnubhakat, Arti K. Rai & Jay P. Kesan, *Strategic Decision Making in Dual PTAB and District Court Proceedings*, 31 BERKELEY TECH. L.J. 45, 48 (2016) (noting two co-existing frameworks of resulting patent disputes: infringement litigation and declaratory judgment actions in Article III courts vs. administrative invalidation actions in the PTAB).

148. 35 U.S.C. § 315(e).

149. *Id.*

150. *Id.* § 315(a)(1).

151. *Id.* § 315(a)(2); *see* Vishnubhakat et al., *supra* note 147, at 50.

152. 35 U.S.C. § 315(a)(2).

153. *Id.*

154. Vishnubhakat et al., *supra* note 147, at 63.

155. *Id.* at 56.

156. *Id.* at 53 (noting that certain districts see a disproportionately greater number of patent cases and have gained the reputation of being "rocket dockets" that resolve patent cases quickly).

All of these rules impact how parties approach litigation strategy. One study in particular looks in part at how the time-bar affects litigant behavior.[157] Mapping this effect onto what we know about common human behavior, the study suggests that giving discretionary and non-appealable control over § 315(b)'s time bar to the PTO may lead to lax enforcement over time, leading to deleterious downstream effects.[158]

2. *Strategic Decision Making Between the Agency or the Court*

Having litigants make a conscious choice between the PTAB and the courts is a process of mutual substitution that Congress intended with the passing of the AIA.[159] There are two approaches: (1) standard substitution, where a defendant in district court litigation defensively challenges the validity of the asserted patent in an IPR; or (2) nonstandard substitution, where the petitioner of the IPR has not yet been sued in district court and challenges the patent preemptively.[160] Descriptive statistics performed during the first few years of IPR proceedings saw a large and rapid increase in IPR petition filings. There were twenty petitions per month in September 2012, with the number rising to approximately 140 petitions per month by the end of June 2015.[161] During this time, nearly 70 percent of petitioners were standard petitioners, meaning the parties were previously sued on the patents they were now challenging at the PTAB.[162] The same study also saw an increase in district court patent cases filed, rising from 150 cases per month in September 2011 to over 500 cases filed per month in June 2015.[163] More recent data show that roughly 80% of IPRs involve patents that are currently in litigation in parallel district court cases, while 20% of IPRs involve non-litigated patents.[164] These results suggest that many petitioners are using the PTAB as a means to challenge patent validity in connection with the threat or reality of infringement litigation.

---

157. *See id.* at 105 fig. 17 (demonstrating that most IPR petitions are distributed symmetrically on a median lag of six months, with a modal spike at the one-year deadline).

158. *See* Motion for Leave to File Brief of *Amici Curiae* and Brief of *Amici Curiae* Professors of Patent and Administrative Law in Support of Respondent Click-To-Call Technologies, LP at 13–14, Thryv., Inc. v. Click-To-Call Techs., LP, 140 S. Ct. 1367 (2020) (No. 18-916).

159. Vishnubhakat et al., *supra* note 147, at 49 ("A major normative argument for administrative ex post review is that it should be an efficient, accessible, and accurate substitute for Article III litigation over patent validity.").

160. *Id.* at 49.

161. *Id.* at 69.

162. *Id.* at 73.

163. *Id.* at 69.

164. *See* Roshan S. Mansinghani & Robert K. Jain, *Frequent Filers? What the Data Says About Parallel PTAB Cases*, 12 LANDSLIDE, Sept./Oct. 2019, at 41, 42 ("80 to 90 percent of AIA challenges involve patents asserted in district court litigation . . . .").

As discussed in the previous Sections, the statute dictates that a petition for IPR cannot be filed more than a year after a petitioner has been sued for infringing a particular patent unless the petition includes a request for joinder.[165] Empirical data show that most IPR petitions fall within the zero-to-one year range, distributed symmetrically with a median lag time of six months.[166] More importantly, lag time between the filing of the district court suit and the filing of the IPR petition produces a notable spike at the one-year mark, jumping from around 100 IPR petitions sought at 0.0 to 0.7 years to over 300 IPR petitions sought at 0.9 to 1.0 years.[167] These data are consistent with petitioners seeking IPR within the one-year statutory deadline set by § 315(b).[168] By these numbers alone, the IPR's one year time bar plays an important role in determining litigant behavior and strategy. This is especially poignant given that the data show the majority of IPR validity challenges are a defensive response to existing litigation.[169] Currently, litigants must adhere to the one-year bar and adjust their strategies regarding the PTAB and district courts accordingly. However, this could all change if a judicially unsupervised agency is allowed to weaken or completely ignore the enforcement of the time bar.[170]

If the one-year time bar is not strictly enforced, the lag time between PTAB and district court filings could potentially increase because a petitioner will take as much time as the PTAB allows.[171] The dispute in *Thryv* centered on the Board's determination that an infringement suit dismissed without prejudice does not trigger the time bar.[172] This holding further insulates from review a Director's discretionary interpretation of the time bar, which opens the door to strategic attacks. If a suit dismissed without prejudice is allowed to bypass this gate, what else could the Director turn a blind eye to, effectively giving petitioners more proverbial bites at the apple? If these discretionary decisions cannot be reviewed by a court of law, patent owners have little redress against these constant attacks from challengers.

The statutory language of IPRs currently does not have a standing requirement, nor does it preclude potential collective petitioner action through the

165. *See* 35 U.S.C. § 315(b).

166. Vishnubhakat et al., *supra* note 147, at 76.

167. *Id.* at 105.

168. *Id.* at 76.

169. *Id.*

170. *See* Motion for Leave to File Brief of *Amici Curiae* and Brief of *Amici Curiae* Professors of Patent and Administrative Law in Support of Respondent Click-To-Call Technologies, LP at 13, Thryv, Inc. v. Click-To-Call Techs., LP, 140 S. Ct. 1367 (2020) (No. 18-916).

171. *Id.* at 14.

172. *Thryv*, 140 S. Ct. at 1371.

use of joinder to an existing petition.[173] Recent IPR filings in the pharmaceutical industry demonstrate how parties can use this as a tactical advantage. The biotech giant Genentech owns a U.S. patent on a method for making humanized antibodies. This patent has been challenged in no less than twenty-one separate actions, including eleven district court litigations.[174] Notably, this patent has been the subject of ten IPR petitions filed by various pharmaceutical companies, such as Mylan Pharmaceuticals and Pfizer.[175] This example shows how IPRs can lead to a patent being challenged *too* many times with potentially no upper limit. As such, administrative adjudicatory bodies like the PTAB may end up providing an additional forum in which to harass patent owners. A challenger (or group of challengers) can relentlessly and repeatedly attack a patent until it is held to be invalid.[176]

Allowing petitioners to take advantage of an important checkpoint like the IPR time bar has downstream effects on both parties and the courts. Patent owners who cannot afford the time and expense of a drawn-out litigation in court find themselves in parallel disputes in the PTAB.[177] Even patent owners with monetary means can nonetheless be whittled down by constant attacks. Courts may find themselves drained of their resources if petitioners can simply start over again at the PTAB.[178] This effect would hardly further Congress's intent of reducing both litigation costs and judicial waste when it passed the AIA.[179]

B. NO JUDICIAL REVIEW LEADS TO AN UNSTABLE AND UNPREDICTABLE PATENT SYSTEM

The *Thryv* majority spent a great deal of its opinion on statutory interpretation of the phrase "under this section" to confirm that § 314(d) should apply

173. Vishnubhakat et al., *supra* note 147, at 59.

174. *See Patent 6407213 Invented by Carter*, DOCKET NAVIGATOR, https://search.docketnavigator.com/patent/patent/53705/1.

175. *See id.*

176. *See, e.g.*, Gregory Dolin, *Dubious Patent Reform*, 56 B.C. L. REV. 881 (2015); Jay P. Kesan, *Carrots and Sticks to Create a Better Patent System*, 17 BERKELEY TECH. L.J. 763, 783 (2002) (discussing the dangers of delay and harassment in post-issuance patent office proceedings); Raymond A. Mercado, *Ensuring the Integrity of Administrative Challenges to Patents: Lessons from Reexamination*, 14 COLUM. SCI. & TECH. L. REV. 558, 606–07 (2013); Jonathan Masur, *Patent Inflation*, 121 YALE L.J. 470, 522 (2011) ("[I]nter partes review could potentially be abused by parties interested only in delaying and harassing competitors.").

177. Brief of the Pharmaceutical Research and Manufacturers of America as *Amicus Curiae* in Support of Respondents at 15, *Thryv*, 140 S. Ct. 1367 (No. 18-916).

178. *Id.*

179. *See* Dolin & Manta, *supra* note 24, at 737-42 (discussing various post-issuance proceedings under the AIA.

to other sections as well.[180] However, imbuing the PTO and PTAB with this level of authority allows the agency an unprecedented degree of discretion that threatens the predictability and stability of the entire patent system.

1. *A Need for a Neutral Adjudicator*

A fundamental purpose of Article III courts is to provide a forum presided over by a neutral adjudicator, a judge who is "free from potential domination by other branches of government."[181] The Framers wanted to create a system different from the old one, where the King of Great Britain had "made Judges dependent on his Will alone, for the tenure of their offices, and the amount and payment of their salaries."[182] The Framers sought to create a system in which judges would not be beholden to nor impacted by the other branches of the federal government.[183] This goal is served by giving Article III judges a lifetime tenure on the bench, as well as financial security.[184]

Like Article III judges, PTAB judges perform a similar adjudicatory purpose, but PTAB judges operate under a completely different framework. PTAB judges are appointed by the Secretary of Commerce with consultation from the Director.[185] Any decisions made by the Board must be under the supervision of the Director, who is ultimately responsible for final decisions.[186] The Director in turn is a political appointee, appointed by the President and under the supervision of the Secretary of Commerce, who is also a political appointee.[187] Not only are the Board members supervised and paid by the Director, but the Director is also allowed to select which members will hear a particular patent challenge and how many will participate in the hearing.[188] Should the Director disagree with the decision of the hearing, he can add more members to the panel (which includes himself) to rehear the case.[189] This broad control by one executive appointee is a large part of what concerned Justice Gorsuch in the *Thryv* dissent.[190]

---

180. *Thryv*, 140 S. Ct. at 1373–74.

181. Commodity Futures Trading Comm'n v. Schor, 478 U.S. 833, 848 (1986) (quoting United States v. Will, 449 U.S. 200, 218 (1980)).

182. Stern v. Marshall, 564 U.S. 462, 484 (2011) (quoting THE DECLARATION OF INDEPENDENCE para. 11 (U.S. 1776)).

183. *Id.*

184. *Id.* at 483–84.

185. 35 U.S.C. § 6(a).

186. *Id.* § 318.

187. *Id.* §§ 3(a)(1), (4).

188. *Id.* § 6(c).

189. *Id.*

190. *See* Thryv, Inc. v. Click-To-Call Techs., LP, 140 S. Ct. 1367, 1378 (2020) (Gorsuch, J., dissenting).

2. *Discretionary Control Undermines Stability and Predictability*

Allowing discretionary control by an agency without any oversight further destabilizes the patent system. There are a few different ways in which discretionary control over the IPR time bar leads to unpredictable results for litigants due to the unique way IPRs intersect with district courts.

The time bar of § 315(b) serves as a statutory limit on agency authority.[191] It specifically dictates what the agency is not allowed to do.[192] Unlike § 314, where the language is directed to internal agency considerations and decisions related to the petition's merits, § 315 addresses the relationship between IPRs and other proceedings.[193] The reason the time bar is so significant is that during that one year when an IPR petition can be filed, a court is most likely to stay pending district court litigation to avoid parallel litigation.[194] After one year, the court can confidently proceed with its litigation. This serves the purpose of preventing duplicative judicial waste in the courts from having a decision be decided and later overturned. Because the time bar is a gate that prevents cross-traffic accidents, allowing the PTO to exercise sole authority over it would undermine the adjudication of pending infringement actions.[195] It would also mean that petitioners would remain unsure of what triggers the time bar, i.e., like in *Thryv*. If judicial review of the time bar is *never* permitted, it would create a strange situation in which the Federal Circuit retains judicial review over the merits of a final agency decision that the court understands should have never been granted in the first place under § 315(b)'s time bar. The result is a patent that should have been excluded from PTO review by Congress's design, all the while wasting litigation resources and time in district court.

For inventors seeking patents to protect their inventions, relying on a patent system that can have contradictory or discretionary outcomes is not an attractive option. Inventors would be disincentivized to seek patent protection if they are unsure about which patent laws apply. The Supreme Court itself has said that uniformity would encourage innovation by providing certainty to both the patentee and any potential infringers about the reach of the patent.[196] Additionally, if inventors disagree with the PTO's assessment of their invention, no judicial review of subsequent PTAB decisions precludes inventors from getting a second opinion and potentially any redress. Given the time and

191. 35 U.S.C. § 315.

192. *Id.* §§ 315(a)(1), (b) ("An inter partes review may not be instituted . . . .").

193. *See id.*

194. *Id.* § 315(a)(2).

195. *See* Brief of the Pharmaceutical Research and Manufacturers of America, *supra* note 177, at 10.

196. Markman v. Westview Instruments, Inc., 517 U.S. 370, 390–91 (1996).

cost that goes into pursuing innovation, the "carrot" of having exclusionary patent rights may be reduced by all of these different "sticks," ultimately impeding the Patent Act's goal of incentivizing innovation.

C. PATENT EXCEPTIONALISM SHOULD APPLY

Patent exceptionalism is the idea that the PTO, and by extension the PTAB, is different from other administrative agencies and thus should be precluded from application of deferential administrative law principles.[197] The role of the PTO is likened to the role of an agent in an offeror/offeree contractual relationship. By understanding the PTO's role as more ministerial, unlike other administrative agencies that are delegated rulemaking authority, the deferential standards from administrative law would not make sense.[198] This then leaves the PTO/PTAB open to judicial review, which can help to decrease the strategic litigation strategies and discretionary control problems articulated above.[199]

1. A Patent Law Theory Based in Contract Law

The administrative state was created to help ease the burden on Congress.[200] However, the patent law system predates this creation, so it is puzzling as to why administrative law principles should apply to patent law.[201] Rather, it is more helpful to use a framework presented by Professor Orin Kerr, who analogizes Congress's grant of a patent to a contractual offer, with the PTO as Congress's agent in that offer.[202] The patent applicant is thus the offeree and accepts the terms of the offer by meeting the statutory requirements of the Patent Act.[203] Thinking of the patent system in terms of a contractual relationship[204] can also help make clear why de novo judicial review should be applied to the PTAB to help protect patent owners in a system in which increasing patent challenges threaten the incentives to seek a patent.

---

197. *See* Orin Kerr, *Opinion: The Case for Patent Law and Criminal Law Exceptionalism in the Administrative State*, WASH. POST (Feb. 12, 2016, 1:51 AM), https://www.washingtonpost.com/news/volokh-conspiracy/wp/2016/02/12/the-case-for-patent-law-and-criminal-law-exceptionalism-in-the-administrative-state/.

198. *See* Kerr, *supra* note 47, at 131.

199. *See supra* Sections IV.A, IV.B.

200. *See supra* Section II.C.1.

201. *See* Robert P. Merges, *The Hamiltonian Origins of the U.S. Patent System, and Why They Matter Today*, 104 IOWA L. REV. 2559, 2561 (2019).

202. *Id.* at 138.

203. *Id.*

204. *See* J.E.M. Ag Supply, Inc. v. Pioneer Hi-Bred Int'l, Inc., 534 U.S. 124, 142 (2001).

Under Article I, Section 8, Clause 8, Congress has the power "[t]o 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."[205] To induce inventors to invent, Congress offers a binding contract: powerful exclusionary rights for a limited amount of time to each inventor who has created something deemed worthy of patentability.[206] If the patent application can satisfy the Patent Act, then an offer is made and accepted and the contract exists.[207] If the application does not satisfy the Patent Act, no offer is extended, and no contract exists.[208] The PTO's role in this exchange is only to *evaluate* whether the application meets the contractual requirements.[209] Unlike other regulatory agencies, the PTO has no regulatory rulemaking authority since it does not determine the actual *terms* of the contract.[210]

If it is accepted that Congress creates the offer through the creation of patent laws, and the PTO reviews the patent applications to determine if the inventors have met the requirements and accepted Congress's offer, it follows that the PTO serves the ministerial role of applying the patent law determined by Congress and the courts to the facts presented in the patent applications.[211] While the PTO can give guidance on how the law will be applied to the facts, i.e., if the application meets the offer's requirements, the agency cannot decide *what* those laws are. That power resides solely with the courts and Congress. Neither the patent examiners who assess the applications nor the PTAB adjudicating adverse decisions have any substantive power to interpret the terms of the offer, change the terms, or exercise any *discretion* to decide whether an application should be granted a patent. This idea of a contract law patent system is very different from the regulatory function that makes up administrative law, and it explains why administrative law deference should not extend to patent law.

Professor Kerr provides a useful way to think about the above: Imagine that A says to B, "I will give you $100 if you walk across the Brooklyn

205. U.S. CONST. art. I, § 8, cl. 8.

206. Kerr, *supra* note 47, at 135–38.

207. *Id.* at 130.

208. *Id.*

209. *Id.* at 138–40.

210. *Id.*

211. *See, e.g.*, Nobelpharma AB v. Implant Innovations, Inc., 930 F. Supp. 1241, 1253 (N.D. Ill. 1996) (noting that a patent examiner is a "ministerial official whose function is to apply policy set by others" (quoting James B. Kobak, Jr., Professional Real Estate Investors *and the Future of Patent-Antitrust Litigation:* Walker Process *and* Handgards *Meet* Noerr-Pennington, 63 ANTITRUST L.J. 185, 209 (1994))), *aff'd*, 141 F.3d 1059 (Fed. Cir. 1998).

Bridge."[212] In this scenario, A is the offeror (i.e., the PTO acting as an agent of Congress), and B is the patent applicant. Once B walks across the bridge, A must decide whether to pay the $100. A's decision here is purely ministerial in that A's obligation to pay over $100 is independent of A's subjective intent.[213] Once the offer to cross the bridge has been made and B has accepted and completed the terms, A does not have the power to restructure the legal relationship after the fact.[214] In other words, A has no discretion to decide whether to pay B the $100 that was promised.[215] But what happens if A has the discretion to change the terms and restructure the legal relationship?

2. *Reliance on the Offer*

An applicant's reliance on Congress's offer is undermined if the agent is allowed to step in and subsequently usurp the terms of the offer.[216] Suppose A, from the above example, adds a condition to the offer, one that says A has the right to determine whether B completed the terms of the offer and walked across the bridge.[217] This would give B a reason for pause. What if A decides that B walked too fast and really *jogged* across the bridge? Under normal circumstances, B could seek legal redress in the form of de novo review from the courts, a clean second look from a neutral third party. However, if A retains interpretive rights on the terms of the offer, B's prospects of earning $100 would depend entirely on how A will decide to interpret the offer. Notably, B will not know the outcome until *after* B has already made it across the bridge. What if making it across the bridge requires a great deal of time and effort on B's part? The uncertainty of A's offer caused by the shift of this interpretive power from the courts to A will discourage B from relying on A's offer and attempting to cross the bridge in the first place.[218]

Applied to patent law, the terms dictate that should the applicant meet certain standards of novelty, then the applicant will be granted the patent. These novelty standards are spelled out in the Patent Act and signed off by Congress. However, if the PTO retains the right to determine what novelty means, then the applicant cannot be sure how the PTO will interpret the standards of novelty. Practically, the applicant only has rights to the patent whenever

212. Kerr, *supra* note 47, at 140; *see also* I. Maurice Wormser, *The True Conception of Unilateral Contracts*, 26 YALE L.J. 136, 136–37 (1916).

213. Kerr, *supra* note 47, at 140.

214. *Id.*

215. *Id.*

216. *Id.* at 183–87.

217. *Id.* at 184.

218. *Id.*

the PTO feels like giving them.[219] Although it is more likely that the PTO will give the applicant the patent if the standards of novelty are met, the applicant cannot be completely sure that this will be the result, because the prospects of getting the patent depend on how the PTO will interpret the standards down the line. Again, the PTO's interpretation will come *after* the applicant has spent time and money to file the application in accordance with the novelty standards.[220] Therefore, the applicant is now left with the fact that the PTO may choose to interpret the standards differently. Would taking this chance be worth all the time and money expended on innovating? This shift of interpretive power from Congress, the original offeror, to the PTO will discourage the applicant from relying on Congress's offer and innovating in the first place.[221] Giving the PTO power to restructure the terms of patentability would create a risk that the research and development efforts used to meet today's interpretation of the Patent Act may end up as useless and invalidating based on tomorrow's interpretation.[222] To an applicant, this would only increase the risks associated with innovating and deter them from trying to pursue such endeavors.[223] Professor Kerr likens this to contract law doctrines, where contracts that grant one party unilateral rights to back out of the contract are disfavored because they discourage other parties from entering into the contract in the first place.[224]

3. *Contract Law Standards of Review*

In terms of reviewing the contractual relationship between Congress and the patent applicant, the question becomes whether a court should defer to the offeror's judgment on whether the applicant's conduct has satisfied the terms of the offer.[225] Under the contract theory, a standard of review of law determines whether the court should defer to the offeror's interpretation of his own offer, while the standard of review of fact determines whether the court should defer to the offeror's assessment of the offeree's efforts to accept the offer.[226] When the offeror's agent (here, the PTO examiner) has denied an applicant a patent, the applicant is allowed to appeal that decision.[227] Similar to the options available to prove a breach of contract claim, the applicant's options are to

219. *Id.*
220. *Id.*
221. *Id.*
222. *Id.* at 186.
223. *Id.*
224. *Id.*
225. *Id.* at 146.
226. *Id.* at 146–47.
227. *Id.* at 143.

either challenge the PTO's construction of the offer (i.e., the PTO's findings of fact) or to challenge the PTO's assessment of the applicant's attempts to meet the statutory requirements of patentability (i.e., the PTO's interpretation of the legal requirements of patentability).[228]

A court that reviews a breach of contract claim applies a de novo standard of review in a contract dispute.[229] A court reviewing a patent adjudication decision from the PTAB should also apply a de novo standard of review. Deferring to the PTAB would be analogous to deferring to the offeror's refusal to give consideration, usurping the offeree's reliance on that offer.[230]

In sum, patent exceptionalism can be understood from a contract-like theory of patent law. This theory compares a grant of a patent to the formation of a contract between Congress and a patent applicant through the PTO. Because the PTO plays a purely ministerial role and has no power to interpret for itself nor change the terms of the contract, it warrants no deference when it comes to reviewing its decisions. Not allowing the PTO interpretive or discretionary control over the terms of the contract also means that applicants can be sure the terms of the contract will not change and can depend on a set outcome if they meet those terms. This is especially important for applicants who are either seeking patents on inventions that require a lot of expensive up-front costs or are counting on licensing deals that insist on settlement of title and patent validity.[231] This reliability in the offer incentivizes the applicant to make the initial investment in seeking a patent and furthers Congress's goal of promoting innovation. Finally, applying de novo review to the PTO, the same standard as applied to district courts, means that litigants no longer have an incentive to try to game the system between using the PTAB or district court.

D. WHERE DO WE GO FROM HERE? THE FUTURE OF PATENT DEFERENCE

While the AIA brought on new and efficient changes to the patent system, its emphasis on efficient ways to challenge patent validity through IPRs may have been near-sighted, leaving many patentees with few protections. Judicial review is a way to restore that balance.

228. *Id.* at 147–48.

229. *Id.* at 149.

230. *Id.* at 148.

231. *See* Jake Berdine & Matt Rosenberg, *Creating Leverage: A Practitioner's Guide to Inter Partes Review and its Effects on Intellectual Property License Negotiations*, 44 AIPLA Q.J. 1, 6–7 (2016) (highlighting the importance for negotiators to weigh various factors during licensing negotiations, including settlement value of infringement and validity suits).

Due to its unique history and development, patent law should receive a specific carve-out and be exempted from administrative law principles. Specifically, the courts should not apply *Chevron* deference for questions of law and mixed questions of law and fact when reviewing PTAB decisions. This Part explores and responds to some of the most common objections to judicial review of the PTO/PTAB. It concludes that congressional action is unlikely in the short term. Thus, de novo judicial review is the best tool to provide protection to patent owners, while preserving patent law's ultimate goal of incentivizing innovation.

1. Responses to Objections

a) a) The Misclassification of Patent Rights as Public Rights

One of the arguments against de novo judicial review of the PTAB lies in the classification of patent rights as per se public rights. This classification greatly impacts the separation of powers debate as public rights are generally thought of as a grant of special favor by the government, as exemplified by public services like utilities, lands, the post office, and interstate commerce.[232] Congress thus has greater discretionary authority to delegate adjudication of public rights to administrative agencies.[233] In contrast, a private right stems from a relationship between individuals and involves classic common law doctrines like contract, property, and tort.[234] Adjudication of private rights is traditionally thought to belong to Article III courts.[235]

Patents and other intellectual property rights are commonly and mistakenly classified as public rights simply because the rights first came from statutes. Modern day courts have perpetuated this misconception and effectively reduced the public-private rights distinction to merely a question of whether the law stemmed from statutes or from common law.[236] In 2018, the Supreme Court in *Oil States v. Greene's Energy Group* upheld the constitutional legitimacy of IPR proceedings, effectively giving the PTAB definitive authority to review and cancel issued patents.[237] In his opinion, Justice Clarence Thomas argued

232. Crowell v. Benson, 285 U.S. 22, 31 (1932).

233. *See generally* Oil States Energy Servs., LLC v. Greene's Energy Grp., LLC, 138 S. Ct. 1365 (2018) (discussing the public rights doctrine).

234. *See Crowell*, 285 U.S. at 51 ("The present case . . . is one of private right, that is, of the liability of one individual to another under the law as defined.").

235. *Id.*

236. Adam Mossoff, *Statutes, Common Law Rights, and the Mistaken Classification of Patents as Public Rights*, 104 IOWA L. REV. 2591, 2601 (2019).

237. *See Oil States*, 138 S. Ct. at 1365 (holding that patents are public rights and thus the PTAB does not violate either the separation of powers doctrine or a patent owner's Seventh Amendment right to a jury trial when canceling patents via its administrative proceedings).

that because patents are a "creature of statute law" and "did not exist at common law," patents are to be considered public rights.[238] In one sense this characterization makes sense, as sources of patent law come from statutes that have been enacted since the Patent Act of 1790. However, the evolution of patent law has seen the creation of many common law legal doctrines. Some examples include the nonobviousness doctrine, the exhaustion doctrine, literal infringement versus equivalents infringement, and exclusions from patent eligibility.[239] Congress has subsequently codified many of these judicially created doctrines, while doing away with others.[240] It is therefore a mistake to characterize patent rights as solely public rights based on the idea that these rights stem from statutes and other legislation. Doing so limits our understanding of patent rights and improperly precludes application of helpful private rights mechanisms, like judicial review.[241]

b) Patent Law's Presumption of Validity Has Important Policy Rationales and Hints at a Private Right

A doctrinal example that demonstrates that patent rights should be thought of as private rights is the presumption of validity doctrine. The presumption of validity is a legal doctrine in which the courts are obligated to defer to the PTO's initial determination that an invention qualifies for a patent unless the challenger can show by "clear and convincing" evidence that the PTO erred in its judgment to grant the patent.[242] Challengers must meet this higher standard when persuading courts to overrule that determination of validity.[243] Historically, patent examiners are granted great deference when it comes to questions of validity due to their supposed expertise.[244] Judges are less likely to second-guess these experts when it comes to ruling on technologies as these examiners have more knowledge and experience.[245] However, the realities of hundreds of thousands of patent applications filed per year and a

238. *See id.* at 1373–74.

239. Mossoff, *supra* note 236, at 2611–12.

240. *See id.* at 2614 (citing as an example § 103 of the 1952 Patent Act that reversed the "flash of creative genius" test).

241. A more complete discussion of the categorization of patent rights can be found in Robert P. Merges, *What Kind of Rights Are Intellectual Property Rights?*, *in* THE OXFORD HANDBOOK OF INTELLECTUAL PROPERTY LAW 57 (Rochelle Dreyfuss & Justine Pila eds., 2018). This Note will focus on how understanding patent rights through the lens of private rights allows for judicial review. This Note leaves the discussion of whether patent rights fit into private or public rights, or both, to other scholars.

242. Doug Lichtman & Mark A. Lemley, *Rethinking Patent Law's Presumption of Validity*, 60 STAN. L. REV. 45, 47 (2007).

243. *Id.*

244. *Id.*

245. *Id.*

limited number of patent examiners result in examiners spending very little time evaluating any given patent.[246] While patent examiners assigned to evaluate applications are chosen due to having backgrounds generally related to the technology of the applications, examiners rarely have the *exact* expertise required to fully understand the invention.[247] This, coupled with the sheer volume of applications and the time pressure examiners face to get through an application, results in the granting of so-called bad patents. Worse still, these bad patents end up being vigorously enforced later due to this presumption of validity.

In light of the above, deference to the PTO hardly seems warranted, but the presumption of validity lives on due to powerful policy reasons. First, it is not practical to have a more aggressive first-round review. It would be extremely cost-prohibitive to hire experts with specific expertise in the technology at hand.[248] The cost of hiring such experts would likely fall on the patent applicants themselves, resulting in cost-prohibitive filing fees. Second, and relatedly, an aggressive first-round review and exorbitant fees would chill applicants from filing in the first place.[249] Researching and developing an invention can take an extraordinary amount of time, effort, and money. If inventors lack confidence that their work could result in a patent, or if they believe that they would need to fight an uphill battle to gain patent rights, it would similarly disincentivize inventors from becoming applicants. The patent system was created as an incentive-driven system, and keeping the initial gates relatively open allows innovation to move forward.

The presumption of validity also hints at the idea that patent rights can be thought of as private rights. This presumption has long been a fixture of common law.[250] Like other private rights, the presumption works under the belief that before the patentee can be stripped of his or her patent rights, there must be proof shown through competent evidence of some deliberate wrong.[251]

---

246. *See* U.S. PAT. & TRADEMARK OFF., FY 2020 UNITED STATES PATENT AND TRADEMARK OFFICE PERFORMANCE AND ACCOUNTABILITY REPORT 188 tbl.1 (2020) (noting that the number of total patent applications filed during 2016–2020 averaged around 653,697 per year, with an average pendency of 24 months per application).

247. Lichtman & Lemley, *supra* note 242, at 53.

248. *Id.* at 54 (estimating that even a modest salary for hiring these experts would result in aggregate costs upwards of $3 billion annually).

249. *Id.*

250. *See* 35 U.S.C. § 282; Cellspin Soft, Inc. v. Fitbit, Inc., 927 F.3d 1306, 1319 (Fed. Cir. 2019) ("[P]atents granted by the Patent and Trademark Office are presumptively valid.").

251. *See* Therasense, Inc. v. Becton, Dickinson & Co., 649 F.3d 1276, 1290 (Fed. Cir. 2011) (en banc) (finding that an accused infringer claiming inequitable conduct must prove that the patentee "acted with the specific intent to deceive the PTO").

This belief is similar to beliefs in other private rights regimes, such as property. Indeed, infringement is often thought of as a type of trespass.[252]

2. *Judicial Review is Currently the Best Tool for Change*

The AIA was an extremely popular piece of legislation and has mostly changed our patent system for the better. Despite widespread agreement on its basic principles, there are still some areas, like the IPR proceeding, that would benefit from legislative reform. However, this legislative reform is unlikely to occur anytime soon. While the traditional congressional lawmaking process lends itself well to dealing with well-understood problems, it becomes less effective in the technology space, where there can be fast and unpredictable changes.[253] Big tech companies have become comfortable with the tactical advantages IPRs give them against non-practicing entities (commonly referred to as patent trolls) and are less likely to push for legislative reform that would make it harder for them to challenge patents.[254] On the other side of the spectrum, players like the ones in the pharmaceutical industry are fighting to protect their patent rights and the value of their patent portfolios.[255] Judicial review is a tool that is currently available to provide protection to patent owners, while still allowing for an efficient patent system.

Going back to an era of patent exceptionalism and allowing for judicial review of PTAB decisions would be relatively simple. It would only require reviewing courts to discontinue the trend of applying administrative law deference standards (unless specifically required by Congress via statute). Under this new regime, the Federal Circuit would act as the primary gatekeeper and ultimate interpreter of patent law. Congress can continue to implement policy changes, working with both the PTO and the courts to ensure patentees meet the standards for patentability *and* receive a full judicial review. After all, the

---

252. Eric R. Claeys, *The Conceptual Relation Between IP Rights and Infringement Remedies*, 22 GEO. MASON L. REV. 825, 851 (2015).

253. *See* Wulf A. Kaal & Erik P.M. Vermeulen, *How to Regulate Disruptive Innovation—From Facts to Data*, 57 JURIMETRICS J. 169, 173 (2017) (noting that where there is rapidly evolving innovation, "by the time legal issues are addressed, new and different legal issues are created").

254. *See* Ashley Gold, *Why Big Tech is Suing the Patent Office*, AXIOS (Sept. 3, 2020), https://www.axios.com/why-big-tech-is-suing-the-patent-office-543809c5-f277-4316-aa66-b7804149fae4.html (pointing to a recent action by Apple, Google, Cisco, and Intel against the PTO that challenges the agency's rule of refusing to adjudicate patent claims while there is parallel district court litigation, because "[t]he companies say the rule hurts innovation and their legal rights, letting invalid patents stay on the books while lawsuits slowly wend their way through court").

255. *See supra* Section IV.A.2; *see generally*, Patricia Carson & Ashley Cade, *Implications of Inter Partes Review Proceedings on Biotech/Pharma Litigation*, 8 PHARM. PAT. ANALYST 65, (2019).

Federal Circuit is unique among appellate courts as the only reviewing court with jurisdiction based on subject matter.[256]

De novo review of PTAB decisions would also not deter litigants from participating in post-grant review proceedings, such as IPRs. Litigants would still be incentivized to partake in IPR proceedings for all the same reasons that Congress intended: speediness, affordability, and full participation of the challenger in the proceedings. The most recent IPR statistics show that of all IPRs that go to final written decision, 62% end up with all claims being found unpatentable.[257] This advantage for the challenger can be evened out by allowing owners to be able to appeal to a reviewing court and receive judicial review of any contested decisions, including review of provisions, like the time bar, that are not clearly barred from judicial review by statute. Judicial review would achieve the correct amount of balance between providing efficiency of patent challenges and protection for patent owners.

Ratcheting up a standard of review would be relatively simple for a court to achieve. Despite the trajectory of patent-related administrative law precedent at the Supreme Court, the Court has not definitively required application of *Chevron* or any other deferential standard to appellate review of questions of law or mixed questions of law and fact. Stability and fairness have always been judicial goals. Harmonization with district courts is not a new exercise for the PTO. The issue of uniformity recently came up in the context of claim interpretation. Until recently, the PTAB gave claim language its broadest reasonable interpretation, consistent with other PTO proceedings.[258] Meanwhile, the Federal Circuit required district courts to give claims their ordinary meaning, as understood by a person with skill in the art.[259] A broader interpretation standard leads to broader claims, which then leads to a higher likelihood of reading on prior art. Thus, the broader interpretation standard made it easier to invalidate patent claims at the PTAB compared to district courts.[260] In 2018, the PTO recognized the dangers of this lack of uniformity and changed its claim

256. *Court Jurisdiction*, U.S. CT. OF APPEALS FOR THE FED. CIR., http://cafc.uscourts.gov/the-court/court-jurisdiction.

257. PAT. TRIAL & APPEAL BD., U.S. PAT. & TRADEMARK OFF., TRIAL STATISTICS: IPR, PGR, CBM 11 (2020), https://www.uspto.gov/sites/default/files/documents/trial_statistics_20200930.pdf [https://perma.cc/V284-VQJW].

258. 37 C.F.R. § 42.100(b) (2020).

259. Phillips v. AWH Corp., 415 F.3d 1303, 1314 (Fed. Cir. 2005) (en banc).

260. The risk of inconsistent interpretations by the PTAB and district courts was noted by Justice Breyer in *Cuozzo Speed Technologies, LLC v. Lee*. 136 S. Ct. 2131, 2146 (2016).

construction standard for PTAB adjudication to conform with the ordinary meaning standard used in district courts.[261]

The tradition of judicial review has a much longer and stronger history than administrative deference. The PTAB and robust post-grant proceedings, like the IPR, are even younger in judicial history. Concerns about the constitutionality of IPR[262] and administrative patent judges[263] have been voiced during each of the cases that the Supreme Court has taken on regarding patent deference. These arguments signal that at least some on the Court question continual deference.

V. CONCLUSION

Reviewing courts should not grant administrative deference to the PTO (and by extension the PTAB) because the agency is not charged with interpreting rules passed by Congress, unlike the vast majority of administrative agencies created after the enactment of the APA. The current trend from the Supreme Court appears to favor granting the PTO more and more deference. This deference is not warranted and needs to stop, especially in the context of IPRs. Given that current IPR regulations favor patent challengers and lack important statutory limits, reviewing courts can step in and offer some protection to patent owners. To understand this fully, a contract law theory is a better way to view the role of the PTO and provides a way for private law mechanisms to justify judicial review of the PTO/PTAB. Allowing judicial review of PTAB decisions and harmonizing the review standard with those used for district courts solves problems of unfair strategic gaming and also leads to a much more stable and predictable patent system. This further incentivizes inventors to participate in the patent system, ultimately promoting innovation.

261. *See* Changes to the Claim Construction Standard for Interpreting Claims in Trial Proceedings Before the Patent Trial and Appeal Board, 83 Fed. Reg. 51340, 51340 (Oct. 11, 2018) (to be codified at 37 C.F.R. pt. 42).

262. Oil States Energy Servs., LLC v. Greene's Energy Grp., LLC, 138 S. Ct. 1365, 1370 (2018).

263. United States v. Arthrex, Inc., 141 S. Ct. 1970, 1983 (2021) (finding that "the unreviewable executive power exercised by APJs is incompatible with their status as inferior officers" and that final written decisions of APJs must be subject to review by the Director of the USPTO).

DIGITAL BLACKBEARDS: COPYRIGHT INFRINGEMENT BY STATES AND THE "CONGRUENCE AND PROPORTIONALITY" TEST IN *ALLEN V. COOPER*

Tom James†

I. INTRODUCTION

Today's "[s]treaming [w]ars" have brought many new players into the market for subscriber dollars, with companies like Disney, Apple, and Amazon wading into an arena once originally dominated by Netflix and Hulu.[1] However, one might not expect to see the State of North Carolina enter the fray with its own streaming platform. Even more surprising might be the realization that, should North Carolina or any other state begin filling its treasury by charging consumers to stream unlicensed films, television shows, songs, or other copyrighted works on their websites, those states could do so without compensating copyright holders. Yet that is the implication of the Supreme Court's recent decision in *Allen v. Cooper*.

The facts of the case concerned only a handful of copyrighted photographs and videos added to the State of North Carolina's website without permission. The Justices, however, fully understood that the case's implications could reach all copyright holders.[2] After all, if a state could post a photographer's copyrighted images to its website without fear of facing monetary damages, it might be able to do the same with any number of other copyrighted works. Perhaps, noted Justice Breyer, a state might conceive of a "wonderful money-raising" venture whereby it charged consumers five dollars to watch the latest blockbuster on its own website.[3] Under such a scheme, he continued, artists that created those works would "unfortunately receive nothing because

DOI: https://doi.org/10.15779/Z38W37KW8B

† J.D. 2022, University of California, Berkeley, School of Law.

1. *See* Jason Murdock, *Disney+, Netflix Streaming Wars Will Usher in New Age of Online Piracy, Warn Experts*, NEWSWEEK (Oct. 12, 2019, 6:30 AM), https://www.newsweek.com/streaming-wars-apple-tv-plus-disney-plus-netflix-hulu-piracy-torrenting-risk-1464648 (discussing the significance of the arrival of new players in the entertainment streaming market).

2. *See* Transcript of Oral Argument at 36, Allen v. Cooper, 140 S. Ct. 994 (2020) (No. 18-877).

3. *Id.*

everyone will have seen [their films] on the state's own streaming device."[4] Similarly, Justice Kavanaugh speculated that a holding for North Carolina risked encouraging "rampant" copyright infringement by states, while Justice Sotomayor called the circumstances of the case "deeply troubling."[5]

Despite the harsh words and extreme hypotheticals during oral arguments, the Court returned a unanimous decision in favor of North Carolina.[6] The holding protects even the most blatant copyright infringement when the pirates are states, giving them free reign to use copyrighted works however they choose, without the need to compensate creators.[7] It is difficult to reconcile the Court's ultimate decision with the policy concerns that were evidently on the Justices' minds during oral arguments. The answer, however, rests in the doctrine of sovereign immunity.

Sovereign immunity gives state copyright pirates a tool far stronger than any sword wielded by Blackbeard, Jack Sparrow, or Long John Silver. Put simply, the doctrine prevents suits brought against non-consenting states in federal court. Though Congress has tried and, at times, succeeded in abrogating sovereign immunity, the Supreme Court made resoundingly clear in *Allen v. Cooper* that copyright holders do not have a recourse for obtaining damages against states under either Article I of the Constitution or any existing legislation passed through Congress's Fourteenth Amendment enforcement power.[8]

Sovereign immunity was built into the original constitutional framework of the United States at the behest of states worried that their treasuries would be depleted by countless lawsuits.[9] The doctrine's abrogation can be permitted, however, where Congress enacts federal laws to combat unconstitutional behavior by states. However, Congress's power has its own limits. In *City of Boerne v. Flores*, the Court developed a "congruence and proportionality" test for assessing whether a given congressional action has overstepped its authority in curtailing unconstitutional behavior by states.[10] In application, this "congruence and proportionality" test has proven challenging, yielding results

4. *Id.* at 37.
5. *Id.* at 39, 43.
6. *See Allen*, 140 S. Ct. at 999.
7. *See id.*
8. *See id.*
9. *See* Alden v. Maine, 527 U.S. 706, 720 (1999).
10. *See* City of Boerne v. Flores, 521 U.S. 507, 519–20 (1997), *superseded by statute*, Religious Land Use and Institutionalized Persons Act of 2000, Pub. L. No. 106-274, 114 Stat. 803.

that have been contradictory and frustrating.[11] This Note will explore why *Allen v. Cooper* is no exception.

After an overview of *Allen v. Cooper* in Part II, this Note will discuss the history, purpose, and abrogation of sovereign immunity in Part III. Part IV argues that the Supreme Court misapplied its congruence and proportionality test. While the Court maintained that its own precedent "prewrote" its decision in *Allen v. Cooper*, its reasoning leaves too much on the table and unconsidered.[12] The Court failed to grapple with the errors of its previous decisions and subsequently ignored differentiators that separated *Allen v. Cooper* from previous landmark cases, including differences between types of intellectual property and nuances of the relevant legislative record. Ultimately, in its overly rigid misapplication of its own precedent, the Court's analysis became so heavily quantitative that it lost sight of the congruence and proportionality test's genuine purpose. Part IV of the Note will then also discuss available congressional remedies, before concluding in Part V.

II. *ALLEN V. COOPER*

In a single calendar year, copyrighted works contribute over $1.3 trillion dollars to the national GDP, amounting to roughly 7% of the U.S. economy.[13] This number only stands to grow significantly in the future. From 2014 to 2017, industries revolving around copyrighted works grew at an annual aggregate rate of 5.23%, outpacing the rest of the U.S. economy by more than 137%.[14] Contributing to this output is a workforce of nearly 5.7 million

11. *See* Michael J. Neary, Note, *Reversing a Trend: An As-Applied Approach Weakens the* Boerne *Congruence and Proportionality Test*, 64 MD. L. REV. 910, 936–38 (2005) (claiming that the Court's use of the test in *Lane v. Tennessee*, 541 U.S. 509 (2004), conflicted with its application in *Florida Prepaid Postsecondary Education Expense Board v. College Savings Bank*, 527 U.S. 627 (1999)).

12. *See Allen*, 140 S. Ct. at 1007 ("*Florida Prepaid* all but prewrote our decision today.").

13. Since 1990, the International Intellectual Property Alliance has measured the economic impact of copyright-related industries on the economy of the United States in a series of regular reports. This figure comes from their 2018 report, which is the most recent report, and is a conservative estimate only including "core copyright industries," defined as those whose "primary purpose is to create, produce, distribute, or exhibit" materials that qualify for copyright protection. *See* STEPHEN E. SIWEK, INT'L INTELL. PROP. ALL., COPYRIGHT INDUSTRIES IN THE U.S. ECONOMY: THE 2018 REPORT 3 (2018), https://iipa.org/files/uploads/2018/12/2018CpyrtRptFull.pdf. When the figure is expanded to include the "total" copyright industries, meaning those in which only aspects or portions of the products qualify for copyright protection, the number swells to $2.2 trillion and roughly 11.59% of the United States economy. *Id.*

14. Like the previous figure, this one only pertains to "core" copyright industries. *Id.*

Americans.[15] One of them is Frederick Allen, a videographer based in North Carolina.[16]

Starting in the mid-1990s, Allen spent nearly a decade working on a unique project for his home state.[17] In 1996, Intersal, Inc., a marine salvage company, stumbled upon the sunken remains of a French slave ship that was captured by infamous pirate Blackbeard in 1717, rechristened as the pirate ship *Queen Anne's Revenge,* and run aground on a sandbar off the coast of Beaufort, North Carolina roughly one year later.[18] Under federal and state law, the wreck belonged to the State, but North Carolina commissioned Intersal to lead the recovery of the sunken ship.[19] Intersal, in turn, retained Allen to document the process, which he dutifully accomplished by capturing photographs and videos of the recovery.[20] He copyrighted each of his resulting works.[21]

Much like the ship he photographed, Allen soon found himself the victim of a modern-day pirate. In 2013, Allen noticed that North Carolina had uploaded his photographs onto its website without his permission.[22] He sued for copyright infringement, and the State settled, agreeing to pay Allen $15,000 for its infringement and laying out the parties' respective rights to the images moving forward.[23] Within two years, Allen alleged that the State violated the terms of the settlement when it posted five of Allen's videos online and used

15. *Id.*

16. *Allen*, 140 S. Ct. at 999.

17. *Id.*

18. *Id.*

19. *See* 43 U.S.C. § 2105(c); N.C. GEN. STAT. § 121–22 (2021). Notably, under the fifteen-year salvage agreement between North Carolina and Intersal, both parties acknowledged the state's ownership of the shipwreck and the company's salvage rights, which included the ability to "retain a designated portion of the financial proceeds arising from the sale of media relating to the *Revenge* and replicas of its artifacts." Allen v. Cooper, 895 F.3d 337, 343 (4th Cir. 2018), *aff'd*, 140 S. Ct. at 994. Intersal retained "exclusive right to make and market all commercial narrative (written, film, CD Rom, and/or video) accounts" of the recovery. *Id.*

20. *Allen*, 140 S. Ct. at 999.

21. *Id.*

22. *Id.*

23. *Id.* The exact terms of the agreement are beyond the scope of this Note. Generally, the agreement divided Allen's footage into different categories, labeling some footage as "commercial documentaries" and some as "non-commercial media." *See Allen*, 895 F.3d at 344. Intersal (with Allen as its designee) retained the exclusive right to produce a documentary about the recovery, which it would allow to be shown in public museums or for educational purposes. *See id.* Non-commercial media would be watermarked with the website for Allen's production company and was not to be displayed anywhere but the official website of the North Carolina Department of Natural and Cultural Resources, where a link to Intersal's and Allen's websites would be "clearly and visibly displayed." *Id.* All remaining archival footage would be returned to Allen. *Id.*

one photograph in a newsletter.[24] Allen filed suit in federal court in 2015; this time, North Carolina refused to back down.[25] In an effort to get the case dismissed, the State cited the doctrine of sovereign immunity under the Eleventh Amendment, which prevents suits against non-consenting states in federal court.[26]

The case would ultimately make its way to the Supreme Court, with Allen attempting to circumvent sovereign immunity in two ways. He argued first under Article I of the Constitution, which empowers Congress to grant both patents and copyrights in order to "promote the Progress of Science and useful Arts."[27] The Court, however, had already rejected this reasoning. In its 1996 decision in *Seminole Tribe of Florida v. Florida*, the Court concluded that Article I could not be used to "circumvent" the limitations that sovereign immunity placed upon federal jurisdiction.[28] This was affirmed in 1999, when the Court's decision in *Florida Prepaid Postsecondary Education Expense Board v. College Savings Bank* held unconstitutional the Patent Remedy Act's attempt at abrogating sovereign immunity when states infringed on the rights of patent holders.[29] Allen pointed to *Central Virginia Community College v. Katz*, which allowed for the abrogation of sovereign immunity in bankruptcy proceedings under Article I.[30] The Court, however, concluded that *Katz* did not overrule *Seminole* or *Florida Prepaid* and was limited to the realm of bankruptcy proceedings.[31] If Allen had a road to the abrogation of sovereign immunity, it would not be through Article I.

Allen's second attempt at abrogation relied on Section Five of the Fourteenth Amendment. He pointed to the Copyright Remedy Clarification Act (CRCA) of 1990, which, he argued, appropriately leveraged Congress's enforcement power under the Fourteenth Amendment to ensure that no state "deprive[s] any person of life, liberty, or property, without due process of

24. Allen's complaint identified six "infringing works," five of which were videos posted to YouTube and the other an image used in an article within a state-sponsored newsletter. *Id.* at 345.

25. *Allen v. Cooper*, 244 F. Supp. 3d 525 (E.D.N.C. 2017), *rev'd and remanded*, 895 F.3d at 337, *aff'd*, 140 S. Ct. at 994.

26. *Allen*, 140 S. Ct. at 999.

27. *Id.* at 1001 (quoting U.S. CONST. art. I, § 8, cl. 8).

28. *Id.* at 1002 (quoting Seminole Tribe of Fla. v. Florida, 517 U.S. 44, 73 (1996)).

29. 527 U.S. 627 (1999).

30. Cent. Va. Cmty. Coll. v. Katz, 546 U.S. 356 (2006). Far from abrogating sovereign immunity, the Bankruptcy Clause of Article I approached states as "competing sovereigns," among whom federal courts could settle disputes in the face of "wildly divergent schemes" for discharging debt in different states. *Allen*, 140 S. Ct. at 1002–03 (quoting *Katz*, 546 U.S. at 365, 373).

31. *See Allen*, 140 S. Ct. at 1002–03.

law."[32] Unlike Article I's Intellectual Property Clause, Section Five has been used by the Court to abrogate sovereign immunity, but only in certain circumstances.[33] Whether or not Allen's case fell within one such circumstance required applying a test from *City of Boerne v. Flores*, under which congressional action was "appropriate" under Section Five only if there was "a congruence and proportionality between the injury to be prevented or remedied and the means adopted to that end."[34] This "congruence and proportionality" test functioned as an assessment of whether Congress's means matched the ends of addressing a particular problem.[35] The assessment required looking at two variables. First, the Court needed to determine "the nature and the extent" of the constitutional problem facing Congress using the relevant legislative record as the clearest evidence.[36] Second, the Court considered the "scope" of Congress's chosen response by asking whether the law it ultimately passed went beyond addressing actual constitutional violations.[37] Under this framework, only "reasonably prophylactic legislation" can pass muster.[38]

In applying the "congruence and proportionality" test to Allen's case, the Court claimed to have little wiggle room due to its established precedent.[39] Deciding to begin with the question of scope, the Court again turned to *Florida Prepaid*, its 1999 decision on the Patent Remedy Act.[40] Much like the CRCA in Allen's case, the Patent Remedy Act aimed to abrogate sovereign immunity without limiting its scope to especially egregious cases involving non-negligent infringement or infringement explicitly authorized by state policy.[41] The Court found that allowing for abrogation in instances where a state's infringement had only been unintentional made the Patent Remedy Act's scope unforgivably "indiscriminate."[42] Since the CRCA was effectively the copyright equivalent of the same statute, the Court reasoned that its scope was necessarily "identical" to that of the Patent Remedy Act, which is to say, likewise indiscriminate.[43]

---

32. *Id.* at 1003 (quoting U.S. CONST. amend. XIV, § 1).
33. *Id.* at 1004.
34. *Id.* (quoting City of Boerne v. Flores, 521 U.S. 507, 519 (1997), *superseded by statute*, Religious Land Use and Institutionalized Persons Act of 2000, Pub. L. No. 106-274, 114 Stat. 803).
35. *Id.*
36. *Id.*
37. *Id.*
38. *Id.* (quoting Kimel v. Fla. Bd. of Regents, 528 U.S. 62, 88 (2000)).
39. *Id.* at 1007.
40. *Id.* at 1004.
41. *Id.* at 1005.
42. *Id.* (quoting Fla. Prepaid Postsecondary Educ. Expense Bd. v. Coll. Sav. Bank, 527 U.S. 627, 646 (1999)).
43. *Id.* at 1006.

Apparently, no further analysis of copyright law or the statute itself was necessary.

Given this presumably broad scope, the CRCA could only survive the congruence and proportionality test if it had been passed off the back of exceptionally strong evidence of widespread constitutional violations.[44] Upon review, the Court was unsatisfied by the evidence. Once again, the Court's reasoning followed the example of *Florida Prepaid*. The previous decision had placed a significant amount of emphasis on the number of concrete examples of patent infringement, which were too few to "identify a pattern of unconstitutional" behavior.[45] The Court acknowledged a "significant disparity" between how Congress had created its legislative record when it passed the CRCA and the Patent Remedy Act.[46] The Court's conclusion regarding the evidence of the record, however, remained the same. Unlike the Patent Remedy Act, the CRCA was passed after Congress commissioned and reviewed a report from the then Register of Copyrights, Ralph Oman, who argued over the course of 158 pages that "copyright proprietors have demonstrated they will suffer immediate harm if they are unable to sue infringing states in federal court."[47] Despite the larger argument being made, the only data of interest to the Court were the dozen examples of infringement it counted throughout the document, only two of which it considered sufficiently intentional so as to be unconstitutional.[48] This figure, the Court concluded, was "exceedingly slight."[49]

Comparing the "slight" evidence of harm to the "indiscriminate scope" of the statute meant that the CRCA could not survive *Boerne*'s "congruence and proportionality" test as it had been applied in *Florida Prepaid*.[50] Having failed to abrogate sovereign immunity under Article I or Section Five, the CRCA was struck down, taking Allen's claim with it.[51] The precedent was sufficiently clear that no Justice dissented, with Justices Breyer and Ginsburg only underlining in their concurrence that they found the Court's decisions in both *Seminole* and *Florida Prepaid* to be misguided but nevertheless controlling.[52] Justice Thomas,

---

44. *Id.*
45. *Id.* at 1005.
46. *See id.* at 1006.
47. *Id.* (quoting REGISTER OF COPYRIGHTS, U.S. COPYRIGHT OFF., COPYRIGHT LIABILITY OF STATES AND THE ELEVENTH AMENDMENT 103 (1988)).
48. *Id.*
49. *Id.* at 1007.
50. *Id.*
51. *Id.*
52. *Id.* at 1009 (Breyer, J., concurring).

meanwhile, took issue with relatively minor aspects of the majority's opinion in his own concurrence.[53]

If there was hope for copyright holders, it did not come from any dissenting voice but rather from a closing note in Justice Kagan's majority opinion.[54] The demise of the CRCA did not mean that Congress would never be able to pass a valid copyright law abrogating sovereign immunity; Congress would simply need to do so in light of the Court's recent decisions.[55] The CRCA had been passed before the decisions in *Seminole*, *Boerne*, and *Florida Prepaid*. Congress would now need to draft new legislation with a keen awareness of the hurdles these decisions placed in its path.[56] *Seminole*'s holding that Article I did not abrogate sovereign immunity meant that Congress needed to pass a new law properly abrogating the doctrine under Section Five. This could be achieved by intentionally creating a legislative record that linked the scope of the new law's abrogation to the unconstitutional injuries Congress was able to prove as per *Boerne*, *Florida Prepaid*, and now *Allen v. Cooper*.[57] Such a "tailored statute" could, the Court concluded, "effectively stop States from behaving as copyright pirates" and ultimately "bring digital Blackbeards to justice."[58] Reviewing the case and its precedent further, however, suggests that alternative recourses were possible.

III. ABROGATION OF SOVEREIGN IMMUNITY

A. INTRODUCING SOVEREIGN IMMUNITY

Understanding the legal arguments that ultimately doomed Allen's case requires understanding the history and context from which the Eleventh Amendment emerged. While sovereign immunity is sometimes referred to as "Eleventh Amendment immunity," the phrase is only a "convenient shorthand" and ultimately something of a "misnomer."[59] The actual text of the Amendment only bars suits "against one of the United States by Citizens of another State, or by Citizens or Subjects of any Foreign State."[60] It does not

53. *See id.* at 1007–08 (Thomas, J., concurring) (taking issue with the majority's conception of binding precedent as demanding "special justification[s]" in order to be overruled, disagreeing with the majority's dicta on future copyright legislation, and questioning whether it was established law that copyrights are property under the Fourteenth Amendment).

54. *Id.* at 1007 (majority opinion).

55. *See id.*

56. *Id.*

57. *Id.*

58. *Id.*

59. Alden v. Maine, 527 U.S. 706, 713 (1999).

60. U.S. CONST. amend. XI.

explicitly preclude suits brought against a state by one of its own citizens, like Frederick Allen bringing a claim against his home state of North Carolina. Ultimately, it is not the text but rather the history of the Eleventh Amendment that gives it its teeth. Thus, as with much of American law, the story of *Allen v. Cooper* begins not with a photographer's copyright infringement claim or even the pirate ship he photographed, but with the Founders themselves and their understanding of the Constitution.

In the eighteenth century, it was largely taken for granted that sovereigns (both literal monarchs and governing states) could not be hauled into court.[61] The notion came from English law. Under the feudal system, no lord could be sued by his vassal in his own court, but each lord could face a suit in the court of a higher lord.[62] Since the king sat at the top of the feudal pyramid, there was no higher court that could try him, meaning that sovereign immunity "rested primarily on the structure of the feudal system and secondarily on a fiction that the King could do no wrong."[63] This led Blackstone, whose works were the primary authority on the English legal landscape at the time of the Founders, to describe the concept as a common law rule that "no suit or action can be brought against the king, even in civil matters."[64] Under Blackstone's conception, immunity and sovereignty were inextricably linked. Jurisdiction necessarily implied a "superiority of power" because a court's authority to try someone suggested a similar authority to enforce its judgment, and giving domestic courts such power to command a sovereign would necessarily undermine that sovereign's power.[65] As a result, by the time of the Founders, the concept of bringing suit against one's sovereign was a "thing unknown to the law," and so much so that explicitly forbidding it was considered "hardly necessary to be formally asserted" when the Framers drafted the Constitution.[66]

Nonetheless, discussions surrounding sovereign immunity became heated in the early days of independence. Following the Revolutionary War, several states were deeply in debt and concerned that the creation of a new federal sovereign would "subject them, like lower English lords, to suits in the courts of the 'higher' sovereign," leaving their treasuries vulnerable.[67] The subsequent

61. *See* Hans v. Louisiana, 134 U.S. 1, 16 (1890) (discussing the Founders' perspective on sovereign immunity).

62. Nevada v. Hall, 440 U.S. 410, 415 (1979), *overruled by* Franchise Tax Bd. of Cal. v. Hyatt, 139 S. Ct. 1485 (2019).

63. *Id.*

64. 1 WILLIAM BLACKSTONE, COMMENTARIES *235.

65. *Id.*

66. *Hans*, 134 U.S. at 16.

67. *Hall*, 440 U.S. at 418.

debate focused on the scope of judicial powers under Article III of the Constitution, which gave federal courts jurisdiction over "[c]ontroversies . . . between a State and Citizens of another State."[68] Antifederalists considered it "humiliating and degrading" that a state might be called to answer "the suit of an individual" in federal court.[69] Federalists meanwhile maintained that this was neither the intent nor the implication of the Article, which would only allow for suits in which states were plaintiffs attempting to recover from individuals, not ones in which they were themselves defendants.[70] In his Federalist Papers, Alexander Hamilton attempted to placate worried antifederalists by affirming his commitment to sovereign immunity, which he saw as "inherent in the nature of sovereignty" and "the general practice of mankind."[71] Forcing a state to both meet claims brought by individuals and then adhere to the decisions of a federal court, he continued, was akin to "waging war" against the state and "altogether forced and unwarrantable."[72] Given these assurances, certain state conventions agreed to ratify the Constitution, but only with the understanding that their immunity remained intact.[73] The Constitution was, of course, eventually ratified along with the controversial language of Article III, despite antifederalist fears.

Only five years later, these fears were confirmed by the Supreme Court's decision in *Chisholm v. Georgia*. The case, which concerned a citizen of South Carolina attempting to recover payment for goods sold to Georgia during the Revolutionary War, concluded with the stunning finding that a state could, in

68. U.S. CONST. art. III, § 2, cl. 1.

69. Franchise Tax Bd. of Cal. v. Hyatt, 139 S. Ct. 1485, 1494 (2019) (quoting noted Antifederalist "Brutus," the anonymous author of a series of essays designed to discourage New Yorkers from ratifying the proposed Constitution from 1787 to 1788).

70. Noting the "unusual vehemence" with which Article III had been met, John Marshall proclaimed, "I hope no gentleman will think that a state will be called at the bar of the federal court. . . . The intent is, to enable states to recover claims of individuals residing in other states." Alden v. Maine, 527 U.S. 706, 718 (1999) (quoting 3 JONATHAN ELLIOT, DEBATES ON THE FEDERAL CONSTITUTION 555–56 (2d ed. 1854)). Echoing the point that states would only be plaintiffs but not defendants in such federal cases, James Madison underscored that it was "not in the power of individuals to call any state into court" at the Virginia ratification convention. *See id.* at 716–17 (quoting 3 ELLIOT, *supra*, at 533).

71. *Id.* (quoting THE FEDERALIST NO. 81, at 411 (Alexander Hamilton) (Ian Shapiro ed., 2009)).

72. *Id.* (quoting THE FEDERALIST NO. 81, at 412 (Alexander Hamilton) (Ian Shapiro ed., 2009)).

73. The New York Convention declared that its understanding was "[t]hat the judicial power of the United States, in cases in which a state may be a party, does not . . . authorize any suit by any person against a state" and underlined that it ratified the Constitution only "[u]nder these impressions, and declaring that the rights aforesaid cannot be abridged or violated, and that the explanations aforesaid are consistent with the said Constitution." *Id.* at 718–19 (quoting 1 ELLIOT, *supra* note 70, at 329).

fact, be held liable in suit to a private individual from another state.[74] The Court's reasoning hinged largely on the use of the word "between" in Article III, which seemed to grant federal courts jurisdiction in any case involving both a state and an out-of-state opponent, regardless of which party was the plaintiff.[75] Imagining various possible government misdeeds, the Court presented its holding as a check on a state government's power in the name of "love of morality" and insisted that the nation, having just recently escaped "the vices of old Governments," could readily appreciate that "individuals had been victims to the oppression of States."[76] However, it seemed that even the Justices themselves understood that the holding would likely be met with criticism. One pointed out that if the Constitution was "inconvenient in practice," then perhaps an amendment was in order but that in the meantime, "all offices Legislative, Executive, and Judicial, both of the States and of the Union, are bound by oath to support it."[77] Indeed, an amendment was just around the corner.

The country's reaction to *Chisholm v. Georgia* demonstrated just how deeply ingrained the concept of sovereign immunity was to the founding generation. Newspapers condemned the "unexpected blow" to state sovereignty and worried that there might be a sudden raid on state treasuries.[78] The Massachusetts legislature characterized the decision as "repugnant to the first principles of a federal government."[79] The state House of Representatives in Georgia, the losing defendant in *Chisholm*, quickly passed a bill providing that anyone attempting to enforce the decision would be "guilty of [a] felony and shall suffer death, without benefit of clergy, by being hanged."[80] The most effectual response occurred in the U.S. Congress. The day after the *Chisholm*

---

74. Chisholm v. Georgia, 2 U.S. (2 Dall.) 419 (1793), *superseded by constitutional amendment*, U.S. CONST. amend. XI.

75. The Court also dismissed the idea that the order of the parties in Article III was somehow instructive, writing,

> The order in this instance, works no difference. In common language too, it would not violate the substantial idea, if a controversy, said to be between A. B. and C. D. should appear to be between C. D. and A. B. Nay the opportunity fairly occurs in two pages of the judicial article, to confine suits to States, as plaintiffs; but they are both neglected, notwithstanding the consciousness which the convention must have possessed, that the words, unqualified, strongly tended at least to subject States as defendants.

Id. at 421.

76. *Id.* at 422–23.

77. *Id.* at 468 (Cushing, J., concurring).

78. *See Alden*, 527 U.S. at 720.

79. *Id.* at 720 (quoting 15 ALEXANDER HAMILTON, PAPERS OF ALEXANDER HAMILTON 314 (Harold C. Syrett & Jacob E. Cooke eds., 1969)).

80. *Id.* at 721 (internal quotation marks omitted).

decision was announced, representatives in the House proposed an amendment to the Constitution, and within two months it had made its way through both houses.[81] The result was the Eleventh Amendment.

The Eleventh Amendment was thus a product of the specific circumstances under which it was adopted, and this bore out in its language. Its text only speaks to suits brought against states by out-of-state plaintiffs, as had been the case in *Chisholm* due to the apparent loophole left open by Article III.[82] "Congress chose not to enact language codifying the traditional understanding of sovereign immunity but rather to address the specific provisions of the Constitution that had raised concerns during the ratification debates and formed the basis of the *Chisholm* decision."[83] Presumably, Congress thought it did not need to "draft with a broader brush" by providing sweeping language about the general sovereignty of states and the immunity it allowed, since the national consensus was that states already had that sovereignty in the first place.[84] Rather than change the system, the Eleventh Amendment simply restored the original constitutional design to its pre-*Chisholm* status.[85]

In the centuries since Congress adopted the Eleventh Amendment, it has been understood to stand "not so much for what it says, but for the presupposition…which it confirms."[86] While the text of the Amendment might appear to restrict only Article III diversity jurisdiction of federal courts, the Court has consistently rejected strict readings of the Amendment's text and instead offered unequivocal endorsements of the doctrine of sovereign immunity.[87] As a result, the Court has prohibited suits against non-consenting

---

81. *Id.*

82. *See* U.S. CONST. amend. XI.

83. *Alden*, 527 U.S. at 723.

84. *Id.* at 724.

85. *See id.* at 722.

86. Seminole Tribe of Fla. v. Florida, 517 U.S. 44, 54 (1996) (quoting Blatchford v. Native Vill. of Noatak, 501 U.S. 775, 779 (1991)).

87. The Court deemed one rigidly textualist reading of the Amendment "an attempt to strain the [C]onstitution and the law to a construction never imagined or dreamed of." Hans v. Louisiana, 134 U.S. 1, 15 (1890). It continued, asking,

> Can we suppose that, when the [E]leventh [A]mendment was adopted, it was understood to be left open for citizens of a state to sue their own state in the federal courts, while the idea of suits by citizens of other states, or of foreign states, was indignantly repelled? Suppose that [C]ongress, when proposing the [E]leventh [A]mendment, had appended to it a proviso that nothing therein contained should prevent a state from being sued by its own citizens in cases arising under the [C]onstitution or laws of the United S[t]ates, can we imagine that it would have been adopted by the states? The supposition that it would is almost an absurdity on its face.

Id.

states in many different circumstances.[88] Consistent through many of these decisions is a commitment to protecting state treasuries from otherwise limitless litigation.[89] While *Chisholm v. Georgia* was a lightning rod that brought sovereign immunity to the forefront of the national conversation, sovereign immunity in the United States "neither derives from, nor is limited by," the text of the Amendment it inspired.[90]

If Frederick Allen wanted his infringement claim against the state of North Carolina to succeed, he would first need to circumvent the Eleventh Amendment before the suit could be considered on its merits.

B. ABROGATION OF SOVEREIGN IMMUNITY

Fortunately for Frederick Allen, the doctrine of sovereign immunity can and has been abrogated in limited circumstances. Generally, abrogation falls into three buckets.

First, the *Ex parte Young* exception to the Eleventh Amendment holds that sovereign immunity does not bar citizens from enjoining state officers or officials from enforcing laws that violate the Constitution. Such a suit is not considered to be "against the state" but rather against the individual.[91] This exception is limited, however, because it only applies to injunctions and cannot be leveraged to gain a retroactive award or compensate for monetary losses resulting from the state official's misbehavior.[92] This route was thus

88. *See* Franchise Tax Bd. of Cal. v. Hyatt, 139 S. Ct. 1485, 1496 (2019) (collecting cases).

89. The Court's treatment of sovereign immunity cases has hinged on protecting a state's treasury from actions brought by individuals, with its firm stance being that whenever a suit is "in essence one for the recovery of money from the state, [then] the state is the real, substantial party in interest and is entitled to invoke its sovereign immunity from suit." Ford Motor Co. v. Dep't of Treasury, 323 U.S. 459, 464 (1945), *overruled by* Lapides v. Bd. of Regents of Univ. Sys. of Ga., 535 U.S. 613 (2002); *see* Edelman v. Jordan, 415 U.S. 651, 663 (1974).

90. *Alden*, 527 U.S. at 713.

91. *See Ex parte* Young, 209 U.S. 123, 154 (1908) (quoting Smyth v. Ames, 169 U.S. 466, 519 (1898)).

92. Though the *Ex Parte Young* exception might still have "an ancillary effect on the state treasury" that is "permissible and often an inevitable consequence," it does not permit "any form of relief [that] may be awarded against a state officer, no matter how closely it may in practice resemble a money judgment payable out of the state treasury, so long as the relief may be labeled 'equitable' in nature." *See Edelman*, 415 U.S. at 666, 668; *see also* Hagood v. Southern, 117 U.S. 52 (1886); *Ex parte* Ayers, 123 U.S. 443 (1887) (both showing the historical basis for the Court's holding that the Eleventh Amendment can still bar actions aiming for equitable relief against state officials).

unavailable to Allen, who aimed to recover monetary damages for an infringement that had already occurred.[93]

The second bucket of exceptions concerns states implicated by in rem bankruptcy proceedings. The Court has held that in ratifying the Bankruptcy Clause of Article I of the Constitution, states agreed not to assert sovereign immunity when faced with bankruptcy proceedings.[94] Far from undermining the sovereignty of the states, however, the Bankruptcy Clause approached them as "competing sovereigns," among whom federal courts can settle disputes in the face of "wildly divergent schemes" for discharging debt in different states.[95] Bankruptcies are the exception that proves the rule. Allen argued that since the Bankruptcy Clause of Article I abrogated sovereign immunity, then perhaps the Intellectual Property Clause of Article I might too, but the Court quickly rebuked his argument.[96] Bankruptcy proceedings thus did not provide a workable route for Allen's attempt at abrogation.

The third bucket is the most relevant to *Allen v. Cooper* and also the most nebulous: abrogation by congressional action. In order to abrogate sovereign immunity, Congress must first have "unequivocally expresse[d] its intent to abrogate [sovereign] immunity" within a given piece of legislation.[97] Second, Congress must have acted "pursuant to a valid exercise of power."[98] Determining whether Congress has genuinely acted within the boundaries of its own power has proven complicated, even with the advent of the congruence and proportionality test.

---

93. An injunction—rather than monetary relief—would be insufficient to copyright holders like Allen after states had already infringed on their copyrights. Justice Breyer illustrated this point during the case's oral arguments, offering a hypothetical in which a state streamed an individual's copyrighted film in order to flood its treasury with rental fees. An injunction that prevented the state from continuing to stream the film tomorrow would do little for the copyright holder who had missed out on the streaming revenue from yesterday. *See* Transcript of Oral Argument at 38, Allen v. Cooper, 140 S. Ct. 994 (2020) (No. 18-877).

94. *See Central Virginia Community College v. Katz*, 546 U.S. 356, 375–77 (2006) (allowing for the abrogation of sovereign immunity in bankruptcy proceedings under Article I).

95. *See id.* at 365–70, 373 (contrasting the difficulties that arose from conflicting bankruptcy laws between states, which the Court saw as "peculiar to the American experience," with the clarity of English law, where having a single sovereign meant "a single discharge could protect the debtor from his jailer and his creditors").

96. *See Allen*, 140 S. Ct. at 1002–03 (explaining that the Bankruptcy Clause did not simply allow Congress to abrogate sovereign immunity but rather abrogated sovereign immunity in and of itself). Further action by Congress was unnecessary in the realm of bankruptcy because states had already consented to the abrogation of sovereign immunity. *See id.* This did not, the Court continued, invite the clause-by-clause examination of Article I that Allen relied upon. *See id.*

97. Seminole Tribe of Fla. v. Florida, 517 U.S. 44, 55 (1996) (quoting Green v. Mansour, 474 U.S. 64, 68 (1985)).

98. *Id.*

1. *The Congruence and Proportionality Test*

The congruence and proportionality test is fundamentally a check on congressional authority, which was itself expanded by the Fourteenth Amendment.[99] Enacted in the wake of the Civil War, the Amendment ensured that no state shall "deprive any person of life, liberty, or property, without due process of law."[100] Meanwhile, Section Five gives the Amendment its enforcement power, granting Congress the authority to "enforce" it "by appropriate legislation."[101] This Amendment expanded the reach of the federal legislature into the realm of state prerogatives and thus "fundamentally altered the balance of state and federal power struck by the Constitution."[102]

That this federal power would have limits was understood and demanded even before the Amendment was adopted.[103] The Amendment was not, after all, intended to give Congress free reign to trample state autonomy but merely "allow[ed] Congress to correct the unjust legislation of the States."[104] Broad interpretations of this purpose, which developed in the nineteenth century and survived into the twentieth, pushed a conception of Congress as able to pass any legislation, so long as the "end[s] be legitimate" and the means "not prohibited, but consist[ent] with the letter and spirit of the [C]onstitution."[105] Other cases more firmly maintained that "[a]s broad as the congressional enforcement power is, it is not unlimited."[106] Ultimately, the Court came to describe Congress's power under Section Five as "remedial" rather than

99. *See Allen*, 140 S. Ct. at 1003.

100. U.S. CONST. amend. XIV, § 1.

101. *Id.* § 5.

102. *See Seminole Tribe*, 517 U.S. at 59.

103. The Amendment's original wording granted Congress power to "make all laws which shall be necessary and proper to secure to the citizens of each State all privileges and immunities of citizens," but that proposal was resoundingly defeated by congressmen like William Stewart of Nevada, who felt that the language went too far in granting Congress so much power that "there would not be much left for the State Legislatures." *See* City of Boerne v. Flores, 521 U.S. 507, 520–21 (1997) (quoting CONG. GLOBE, 39th Cong., 1st Sess. 1034, 1082 (1866)), *superseded by statute*, Religious Land Use and Institutionalized Persons Act of 2000, Pub. L. No. 106-274, 114 Stat. 803.

104. *See id.* at 522 (quoting CONG. GLOBE, 39th Cong., 1st Sess. 2459 (1866)).

105. *See* Katzenbach v. Morgan, 384 U.S. 641, 650–51 (1966) (resurrecting the language of Chief Justice Marshall in *McCulloch v. Maryland*, 17 U.S. (4 Wheat.) 316, 421 (1819), more than a century later to conclude that the Fourteenth Amendment was a "positive grant of legislative power" that enabled Congress to "exercise its discretion in determining whether and what legislation is needed to secure the guarantees of the Fourteenth Amendment").

106. *See* Oregon v. Mitchell, 400 U.S. 112, 128 (1970) ("[T]he power granted to Congress was not intended to strip the States of their power to govern themselves or to convert our national government of enumerated powers into a central government of unrestrained authority over every inch of the whole Nation.").

"substantive."[107] The distinction suggests that Congress can use its enforcement power remedially in order to prevent unconstitutional actions by states but cannot pass "general legislation upon the rights of the citizen" by substantively defining new rights under the guise of the Fourteenth Amendment.[108]

The line between congressional action that is substantive and that which is remedial is admittedly murky. The *City of Boerne v. Flores* "congruence and proportionality" test was an attempt to provide clarity.[109] While admitting that the distinction between the remedial and the substantive was "not easy to discern" and that Congress should be granted "wide latitude" to determine where that distinction lies, the Court nevertheless underlined that "the distinction exists and must be observed."[110] The rule it offered was ostensibly simple. Writing for the majority, Justice Kennedy explained, "There must be a congruence and proportionality between the injury to be prevented or remedied and the means adopted to that end. Lacking such a connection, legislation may become substantive in operation and effect."[111] At its core, the congruence and proportionality test as explained in *Boerne* was simply a tool for assessing overall appropriateness. Strong remedial measures that might be justified when implemented to address a problem in one context might be found entirely out of proportion and entirely unwarranted when used to address a lesser problem.[112] As a result, the *Boerne* Court advocated for considering Congress's "remedial measures . . . in light of the evil presented."[113]

In actually applying the congruence and proportionality test in *Boerne*, the Court eschewed a confusing list of enumerated factors and instead engaged in a qualitative and holistic assessment of the legislation at hand. At issue was the Religious Freedom Restoration Act (RFRA), passed in 1993 in an effort to combat any state law that targeted religious beliefs and practices.[114] The Act prevented a state from "substantially burden[ing]" an individual's exercise of their religion unless the government could demonstrate that the burden represented the "least restrictive means" of furthering a "compelling governmental interest."[115] The Court weighed different facets and implications

107. *See* Civil Rights Cases, 109 U.S. 3, 13 (1883); *Boerne*, 521 U.S. at 519–20.
108. *Boerne*, 521 U.S. at 525 (quoting *Civil Rights Cases*, 109 U.S. at 13).
109. *Id.* at 520.
110. *Id.* at 508.
111. *Id.*
112. *Id.* at 530.
113. *Id.*
114. *Id.* at 529.
115. 42 U.S.C. § 2000bb-1, *invalidated by* City of Boerne v. Flores, 521 U.S. 507 (1997).

of the Act. It began (but notably did not conclude) its analysis with an assessment of the historical record that Congress considered when it passed the Act, finding that no examples of the religious persecution that the Act allegedly addressed had occurred in the forty years prior to its enactment.[116] Since the Act applied to any burden experienced by a religious group, it stood to impact a large swath of state laws, providing "[s]weeping coverage" that would "ensure its intrusion at every level of government, displacing laws and prohibiting official actions of almost every description and regardless of subject matter."[117] In addition, the Court found that the RFRA's "stringent" nature was overly burdensome for states because it would likely prove difficult to counter a plaintiff's allegation that their religious expression had been burdened.[118] Meanwhile, the test required the state to show it had used the least restrictive means for furthering its interest.[119] All this amounted to "the most demanding test known to constitutional law."[120] In short, the RFRA's reach was far, its scope enormous, and its burden onto states quite considerable—all in the name of combating an evil that might not have been a meaningful problem in the first place. There was no congruence or proportionality between the problem and its attempted solution, so the Court concluded that Congress, in passing the RFRA, had acted beyond "its sphere of power and responsibilities" under the Fourteenth Amendment.[121]

In the years since *Boerne*, the congruence and proportionality test has failed to bring the clarity that the Court aspired for in its majority opinion. Academics have criticized the inconsistent manner in which the test has been applied, as well as the "divergent and arbitrary" decisions it has prompted.[122] *Allen v. Cooper* continued this tradition.

---

116. *Boerne*, 521 U.S. at 530.
117. *Id.* at 532.
118. *Id.* at 533–34.
119. *Id.*
120. *Id.* at 534.
121. *Id.* at 535.
122. *See* Y. Frank Ren, Note, *Fixing Fourteenth Amendment Enforcement Power: An Argument for a Rebuttable Presumption in Favor of Congressional Abrogation of State Sovereign Immunity*, 94 B.U. L. REV. 1459, 1476 (2014) (arguing that the "congruence and proportionality" test be abandoned altogether); *see also* Neary, *supra* note 11, at 936–38 (claiming that the Court's use of the test in *Tennessee v. Lane*, 541 U.S. 509 (2004), was a novel "as-applied" approach that focused on whether the remedy was justified in one specific context, as opposed to assessing Congress's chosen remedy "as a whole"). Neary noted that if the *Florida Prepaid* Court had used this methodology, the Patent Remedy Act would likely have passed the congruence and proportionality test. Neary, *supra* note 11, at 938.

IV. CONGRUENCE AND PROPORTIONALITY IN *ALLEN V. COOPER*

Frederick Allen's claim in *Allen v. Cooper* sat at the fulcrum of precariously balanced doctrines involving the competing rights of states and the authority of Congress. *Boerne* did not offer any bright-line rules, involved no exhaustive list of factors, and did not suggest that any variable should be weighed as more significant than any other.[123] Notably, the test involved some consideration of the relevant legislative record, but if examining historical examples of unconstitutional activity was the starting point of the Court's analysis, it was by no means its end point.[124] In *Boerne*, the congruence and proportionality test helped the Court broadly consider the scope, reach, and stringency of a congressional action to arrive at a holistic conclusion.[125]

Neither *Florida Prepaid* nor *Allen v. Cooper* received this level of considered analysis. As a result, both cases were poorly decided. Both decisions eschewed holistic considerations of the statutes at hand in favor of a strict review of the relevant historical record. This ignored the intent of the congruence and proportionality test and reduced it to a rote and overly quantitative exercise in counting historical examples. As Justices Breyer and Ginsburg suggested in their concurrence, the Court "erred" in *Florida Prepaid*.[126] However, the fact that their "longstanding view has not carried the day" did not necessarily doom Allen's case in the manner the Justices suggested.[127] The nature of the historical record and facets of intellectual property law should have distinguished Allen's case, making the Court's holding in *Allen v. Cooper* an even more egregious mistake than *Florida Prepaid*. The ultimate effect of the decision is a waste of congressional resources for the sake of a test with an increasingly unclear benefit.

A. APPLYING THE CONGRUENCE AND PROPORTIONALITY TEST

Applying the congruence and proportionality test as it was conceived in *Boerne* requires looking first at the problem that the CRCA aimed to address and then at the scope of the solution it offered. Nuances of the legislative record suggest that evidence of state infringement should not be as casually dismissed as it was by the Court. Meanwhile, distinctions between patent and copyright suggest that the scope of the CRCA was not entirely settled by

123. *See supra* text accompanying notes 114–21.
124. *See supra* text accompanying notes 114–21.
125. *See supra* text accompanying notes 114–21.
126. Allen v. Cooper, 140 S. Ct. 994, 1009 (2020) (Breyer, J., concurring).
127. *See id.*

Florida Prepaid and demands its own consideration, a step in the congruence and proportionality analysis that the Court declined to perform.

1. *The Nature and Extent of the "Evil Presented"*

As described in *Boerne*, the congruence and proportionality test demands analyzing the appropriateness of a remedial action "in light of the evil presented," which in turn requires identifying the "evil" or injury that Congress was aiming to prevent.[128] The Court has typically done so by turning to the legislative findings that compelled Congress's action in the first place.[129]

While a close look at the relevant legislative record is important, it is not a dispositive inquiry. *Boerne* is very clear that the legislative record is but one factor, not the deciding factor in the congruence and proportionality test.[130] Even in *Boerne*, where the Court found no examples of generally applicable laws passed because of religious bigotry, it still maintained that the hole in the legislative record was "not RFRA's most serious shortcoming."[131] Then again in *Florida Prepaid*, where the Court devoted all but the final three paragraphs of its case-specific analysis to the legislative record at hand, the Court still maintained that a "lack of support in the legislative record is not determinative."[132] The congruence and proportionality test in *Boerne* does not offer a specific quantitative tipping point at which a given number of examples in the legislative record suddenly render a remedy appropriate under Section 5. It is instead useful to approach the legislative record as a thumb on the scale, whereby a particularly severe or particularly frequent "evil" will demand a particularly forceful response from Congress.[133]

Yet *Allen v. Cooper* was explicitly decided on the nature of its legislative record. Having already decided that the scope of the CRCA was identical to

---

128. *See* City of Boerne v. Flores, 521 U.S. 507, 530 (1997), *superseded by statute*, Religious Land Use and Institutionalized Persons Act of 2000, Pub. L. No. 106-274, 114 Stat. 803; *see also* Neary, *supra* note 11, at 935 (explaining that the first step in the analysis of *Tennessee v. Lane*, 541 U.S. 509 (2004), was "identifying the rights that Congress intended to enforce through passage of Title II of the ADA and the congressional findings that compelled Congress to pass Title II").

129. *See Boerne*, 521 U.S. at 530.

130. *See id.* at 531 ("Judicial deference, in most cases, is based not on the state of the legislative record Congress compiles but 'on due regard for the decision of the body constitutionally appointed to decide.' " (quoting Oregon v. Mitchell, 400 U.S. 112, 207 (1970) (Harlan, J., concurring in part and dissenting in part))).

131. *Id.*

132. Fla. Prepaid Postsecondary Educ. Expense Bd. v. Coll. Sav. Bank, 527 U.S. 627, 646 (1999).

133. *See Boerne*, 521 U.S. at 530 ("Strong measures appropriate to address one harm may be an unwarranted response to another, lesser one." (citing South Carolina v. Katzenbach, 383 U.S. 301, 334 (1966))).

the "indiscriminate" scope of the Patent Remedy Act in *Florida Prepaid*,[134] the Court decided that the CRCA could only survive a congruence and proportionality inquiry if the legislative record included "stronger evidence" of infringement.[135] Upon review, the Court was unimpressed by what it found.[136] In its hastiness to dismiss the legislative record as too quantitatively slight, the Court lost sight of nuances within the evidence that suggest that Congress had identified a legitimate problem that was only likely to increase.

It is impossible to argue that there was no evidence of copyright infringement by states at the time that the CRCA was passed.[137] One might claim, and the Court certainly did, that the evidence was not overwhelming, but state-sanctioned infringement of copyrights was certainly occurring, even if it was not overwhelming.[138] The Court examined a 158-page report compiled by the then-Register of Copyrights, Ralph Oman ("Oman Report"), who maintained after a year-long inquiry that "copyright proprietors have demonstrated they will suffer immediate harm if they are unable to sue infringing states in federal court."[139] Rather than engage with the reasoning or qualifications within the Oman Report, the Court simply counted the number of examples of infringement. It was willing to concede that there were a dozen instances of state infringement, counting seven court cases and five "anecdotes," of which it felt only two were intentional.[140] The Court was willing to acknowledge that unconstitutional copyright infringement had occurred even at the time that the CRCA was passed but wagged its finger that so few examples were not "the stuff from which Section 5 legislation ordinarily arises."[141] It did not, however, go so far as to offer any guidance regarding how many examples it would need to see before it could permit Congress to intervene.[142] If one dozen was insufficient, it is unclear if two dozen would have been acceptable, or three or four dozen. Simply counting the number of cases makes for an unsatisfying application of the congruence and proportionality test because *Boerne* eschewed any specific quantitative threshold in favor of a general assessment of appropriateness that demands

134. *See infra* Section IV.A.2 (examining the Court's analysis of the scope of the CRCA); Allen v. Cooper, 140 S. Ct. 994, 1005–06 (2020).

135. *Allen*, 140 S. Ct. at 1006.

136. *Id.*

137. *See supra* text accompanying notes 44–49.

138. The Court's contention was that the legislative record did not feature enough examples, but it did concede that some of the examples evidenced genuine unconstitutional infringement by states. *See Allen*, 140 S. Ct. at 1006.

139. *Id.* (quoting U.S. COPYRIGHT OFF., *supra* note 47, at 103).

140. *Id.*

141. *Id.*

142. *See id.*

more than counting. As a result, the Court's application strays from the heart of the congruence and proportionality inquiry.

Evidence within the Oman Report may have been limited, but there were reasons for these limitations. First, the evidence before Congress was from a limited period of time.[143] Congress commissioned the Oman Report in the wake of the Court's decision in *Atascadero State Hospital v. Scanlon* in 1985, which held that "general authorization for suit in federal court is not the kind of unequivocal statutory language sufficient to abrogate the Eleventh Amendment."[144] The decision was seen as a departure from previous decisions, which had been more flexible in the abrogation of sovereign immunity.[145] In response to *Atascadero*, courts across the country suddenly concluded that the Copyright Act lacked the sufficiently unequivocal language necessary for abrogating sovereign immunity.[146] Thus, that states were immune from liability for copyright infringement was still a relatively new idea when Congress commissioned the Oman Report in 1987 and later passed the CRCA in 1990.[147] Therefore, the dozen or so examples cited in the Oman Report, which included two recent cases before the Fourth and Ninth Circuits, had all occurred within the last decade before the CRCA.[148] By comparison, the record of patent infringement by states at issue in *Florida Prepaid* showed only eight examples of infringement suits prosecuted against states in the 110 years between 1880 and 1990.[149] Thus, the Court's conclusion that the evidence motivating the CRCA and the Patent Remedy Act were "equivalent" is reductive.[150] The former's evidence suggested a growing trend, while the latter showed scant evidence over the course of a century. At the very least, the CRCA stood on firmer evidentiary ground than had the Patent Remedy Act.

The Oman Report should be understood as representing merely the "tip of the iceberg" of copyright infringement by states, due to the limitations of

143. *See* Brief of Ralph Oman as Amicus Curiae Supporting Petitioners at 14, *Allen*, 140 S. Ct. 994 (No. 18-877) [hereinafter Oman Brief].

144. Atascadero State Hosp. v. Scanlon, 473 U.S. 234, 246 (1985), *superseded by statute*, Rehabilitation Act Amendments of 1986, Pub. L. No. 99-506, § 1003, 100 Stat. 1807, 1845.

145. *See* Oman Brief, *supra* note 143, at 6–7 (citing *Mills Music, Inc. v. Arizona*, 591 F.2d 1278, 1284–85 (9th Cir. 1979), as an example of a circuit court permitting before *Atascadero* the abrogation of sovereign immunity under the Copyright Act).

146. *See id.* (referencing several decisions).

147. *See id.* at 14.

148. *See id.* at 7, 14.

149. *See* Fla. Prepaid Postsecondary Educ. Expense Bd. v. Coll. Sav. Bank, 527 U.S. 627, 640 (1999) (noting that the legislative record of the Patent Remedy Act only included two examples and that it was actually the Federal Circuit's opinion that offered additional research to bring the count to eight).

150. *See* Allen v. Cooper, 140 S. Ct. 994, 1007 (2020).

Oman's methodology.[151] As head of the Copyright Office, Oman was well informed on the subject matter but did not have the subpoena power necessary to create a genuinely comprehensive catalogue of infringement by states.[152] Rather than seeking public comment, he instead depended on a "modest request for information" sent to a "relatively small group of individuals and organizations savvy enough to be aware of the notice and to prepare and submit responsive comments."[153] At the time, this did not strike Oman as a problem. He understood his task as helping Congress understand the "practical problems" faced by copyright holders under the Eleventh Amendment, as opposed to creating an inscrutable record that would survive a congruence and proportionality test not yet developed by the Court.[154] As a result, he prepared an overview intended as "illustrative rather than exhaustive."[155] The Oman Report should thus be read as providing a flavor of the kinds of infringement experienced by copyright holders and the range of industries being threatened. When read in that light, it is easier to appreciate the diversity of the data in the report, which included complaints from individuals, small and large businesses, and companies spanning numerous industries, including healthcare, the arts, education, and financial data.[156] In simply counting the number of examples, the Court took the Oman Report as overly literal, imagining that it showcased the full extent of state infringement, which was never its intended function.

The Oman Report successfully highlighted a real problem that only grew in the decades after the CRCA. In October 1990, Congress passed the CRCA, citing the Oman Report as "persuasive" evidence of an increasingly serious problem that would "continue to occur if this legislation . . . is not enacted."[157] Congress's predictive judgment has proven accurate, especially after the Court's own decisions signaled that the CRCA might be overturned. The Fifth Circuit was first to overturn the CRCA in 2000, correctly predicting that it was "doomed in the wake of" the Court's 1999 decision in *Florida Prepaid*.[158] The decision likely discouraged other plaintiffs from mounting cases against infringing states, and yet a recent compilation identified more than 150 copyright cases filed against states in the twenty years since the Fifth Circuit's

151. Oman Brief, *supra* note 143, at 13.
152. *Id.*
153. *Id.*
154. *See id.*
155. *Id.* at 14.
156. *Id.* at 12.
157. Oman Brief, *supra* note 143, at 19 (quoting H.R. REP. NO. 101-282, at 8 (1989)).
158. Chavez v. Arte Publico Press, 204 F.3d 601, 608 (5th Cir. 2000).

decision.[159] The cases cover widely different industries. Public universities and schools distributed unauthorized electronic copies of books, with one instance leading to the infringing distribution of nearly 100 copyrighted books in a single academic year.[160] Software company Oracle submitted its own amicus brief detailing its legal battle with the state of Oregon, in which the state pirated millions of dollars' worth of copyrighted software after reneging on a contractual waiver of its Eleventh Amendment rights.[161] Getty Images, the British-American media company, identified fifty instances of copyright infringement by states.[162] The Software & Information Industry Association, a trade association representing 800 technology companies, undertook its own review and discovered seventy-seven matters involving infringement by state entities over the course of just six years.[163] Even if the data within the legislative record of the CRCA was limited, the intervening years have made clear that infringement of copyright by state actors continued to damage copyright holders even before the Act was formally struck down by the Court in *Allen v. Cooper*. In short, the Oman Report's predictive analysis was correct.

By placing so much emphasis on a quantitative review of the Oman Report, the Court too casually dismissed evidence that Congress weighed when it passed the CRCA in 1990. Interpreting the Oman Report as an illustrative snapshot of infringement by states, rather than a comprehensive catalog, demonstrates that the evil of state copyright infringement existed at the time of the CRCA and has remained pervasive ever since.

2. *The Scope of Congress's Remedial Action*

Having determined that the legislative history of the CRCA was lacking, Justice Kagan devoted little time in her majority opinion in *Allen v. Cooper* to assessing the scope of the CRCA.[164] The congruence and proportionality test only demands an appropriateness between the problem identified and the means used to address it. As a result, a slight or particular problem might still be appropriate if addressed through a necessarily slight or particular remedy. As such, even a limited legislative record might still satisfy the congruence and proportionality test if it corresponds to a deliberately limited congressional

159. Oman Brief, *supra* note 143, at 19–20.

160. *See* Brief of the Copyright Alliance and the Chamber of Commerce of the United States as Amici Curiae in Support of Petitioner at 15, Allen v. Cooper, 140 S. Ct. 994 (2020) (No. 18-877).

161. *See* Brief of Oracle America, Inc. as Amicus Curiae in Support of Petitioners at 1–3, *Allen*, 140 S. Ct. 994 (No. 18-877).

162. *See* Brief of the Copyright Alliance and the Chamber of Commerce, *supra* note 160, at 13–16.

163. *See id.*

164. *Allen*, 140 S. Ct. at 1006–07.

remedy. Deciding that the legislative record of the CRCA was limited did not necessarily render it inherently unconstitutional and should not have been the extent of the Court's analysis.

The Court did not assess the scope of the CRCA because it took for granted that its scope was "identical" to that of the Patent Remedy Act.[165] Having already decided in *Florida Prepaid* that the scope of the Patent Remedy Act was "indiscriminate," meaning that "[a]n unlimited range of state conduct" would trigger abrogation of sovereign immunity, the Court took at face value that the same was true of the CRCA.[166] However, the scope of the CRCA was not as "indiscriminate" as the Court alleged, at least not by the standards dictated by the *Boerne* test.[167] In both *Florida Prepaid* and *Allen v. Cooper*, the Court limited its assessment of scope to the fact that neither the Patent Remedy Act nor the CRCA confined the abrogation of sovereign immunity to a particular subset of infringement cases.[168] This reasoning fails to consider the other parameters that limit the scope of the CRCA, particularly those that are unique to copyright law and had not already been examined in *Florida Prepaid*.

The CRCA was first limited in terms of subject matter. The CRCA applied exclusively to copyright infringement cases against states, meaning it did not intrude upon a large swath of unrelated state laws. This set both the CRCA and the Patent Remedy Act apart from the Religious Freedom Restoration Act (RFRA) originally at issue in *Boerne*, where the Court first developed the congruence and proportionality test. In contrast, the RFRA was not limited to any subject matter or cause of action and imposed a test upon all laws that might burden the exercise of any individual's religion.[169] Under the RFRA, a potential plaintiff could have brought a claim against any "branch, department, agency, instrumentality, and official"[170] citing any federal or state law, "statutory or otherwise," adopted before or after the RFRA's enactment.[171] In *Boerne*, the Court distinguished the RFRA from the Voting Rights Act of 1965, which had also been passed under Congress's Section Five enforcement power but was smaller in its reach and scope, only affecting a "discrete class" of laws (in that case, state voting laws).[172] Here, the CRCA was more like the Voting

165. *Id.* at 1007.

166. *See* Fla. Prepaid Postsecondary Educ. Expense Bd. v. Coll. Sav. Bank, 527 U.S. 627, 646–47 (1999).

167. *Allen*, 140 S. Ct. at 1007.

168. *Id.* at 1005.

169. 42 U.S.C. § 2000bb-1, *invalidated by* City of Boerne v. Flores, 521 U.S. 507 (1997).

170. *Id.* § 2000bb-2.

171. *Id.* § 2000bb-3.

172. *See Boerne*, 521 U.S. at 531–33 (discussing *South Carolina v. Katzenbach*, 383 U.S. 301, 315 (1966)).

Rights Act than the RFRA because it only stood to impact a discrete set of claims, copyright infringement suits brought against states. In terms of its impact on state laws, the CRCA was even more limited than both the RFRA and the Voting Rights Act because it had no impact at all on any substantive rule of state law.[173] Unlike the RFRA, the CRCA identified a single problem and enabled a single judicial route for addressing that problem without imposing newfound scrutiny of existing state laws governing unrelated interests.

In confining its scope to the realm of copyright, the CRCA was also limited because copyright is an inherently limited doctrine, even more so than patent law. Unlike patents, which grant protection over ideas, copyrights grant protection over the original expression of an idea.[174] As a result, copyright law leaves room for various permitted uses that patent law does not. For example, independent creation by a second comer who has not actually copied another's work as a factual matter will not lead to liability.[175] Similarly, the merger doctrine holds that copyright protection cannot extend to the expression of an idea that can only reasonably be expressed in so few ways that protecting the expression will result in protecting the idea itself.[176] Furthermore, the existence of a robust fair use defense allows the repurposing of copyrighted works for "criticism, comment, news reporting, teaching . . . , scholarship, or research."[177] These various acceptable uses amount to safe harbors in which a state could not be held liable due to the innate limitations of copyright protection. Similar acceptable uses are not available in the realm of patent law, making copyright liability a more limited cause of action than patent

173. In *Florida Prepaid*, Justice Stevens made a similar argument that the Patent Remedy Act had "no impact whatsoever on any substantive rule of state law, but merely effectuates settled federal policy to confine patent infringement litigation to federal judges." *See* Fla. Prepaid Postsecondary Educ. Expense Bd. v. Coll. Sav. Bank, 527 U.S. 627, 662 (1999) (Stevens, J., dissenting).

174. *See* Fred Fisher, Inc., v. Dillingham, 298 F. 145, 151 (S.D.N.Y. 1924) (discussing an illustrative example involving two mapmakers, who collect firsthand evidence to create maps of the same terrain but will not infringe on one another's copyrights even if their maps are practically identical); Mazer v. Stein, 347 U.S. 201, 217 (1954) (exploring the distinction between copyrights and patents).

175. 4 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 13.01[B] (release no. 116, 2022) ("[W]hen two works are substantially similar with respect to protectible expression, if the defendant did not copy as a factual matter, but instead independently created the work at issue, then infringement liability must be denied."); *see* Kewanee Oil Co. v. Bicron Corp., 416 U.S. 470, 478 (1974).

176. *See* BUC Int'l Corp. v. Int'l Yacht Council Ltd., 489 F.3d 1129, 1143 (11th Cir. 2007) (noting that something as basic as using a circle with a line drawn through it to connote that a certain activity or object is not permitted in a given location cannot be copyrighted, because there are so few ways to express that idea).

177. 17 U.S.C. § 107.

infringement.[178] By extension, the CRCA was necessarily tailored to a more limited range of misconduct than was the Patent Remedy Act at issue in *Florida Prepaid*, which undermines the Court's conclusion that the "scope of the two statutes is identical."[179] In reality, the limitations of copyright law built limitations into the scope and reach of the CRCA that were not present in *Florida Prepaid*. For the sake of assessing congruence and proportionality, these limitations decrease the scope of the CRCA to ensure it is proportional to the problem of state copyright infringement.

Further acceptable uses are spelled out under Title 17 of the United States Code, offering states other circumstances in which they could evade liability without citing the doctrine of sovereign immunity. Under § 110, Congress had already provided states with additional protections for government uses of copyrighted materials intended to benefit the public.[180] Specifically, the provision exempts government bodies from liability when performing various works at agricultural or horticultural fairs,[181] the "systematic mediated instructional activities of a governmental body,"[182] and in performances "specifically designed for and primarily directed to blind or other handicapped persons."[183] Meanwhile, public schools can evade liability for the "performance or display of a work" by either students or their teachers "in the course of face-to-face teaching activities."[184] These additional safe harbors further speak to the inherent limitations of copyright that reduce the proportionality of the CRCA's impact on state actors. The CRCA may not have built-in original limitations on state liability, but that does not render the scope of the Act as "indiscriminate" as the Court imagined in its holding.[185] Instead, its scope was already defined by the limitations of copyright law itself.

Where the reach of the CRCA is broad, it is so by necessity. *Boerne* noted that the presence of termination dates or geographic restrictions within Section Five legislation will "tend to ensure Congress' means are proportionate to ends

178. *See* Brief of the Copyright Alliance and the Chamber of Commerce, *supra* note 160, at 18.

179. Allen v. Cooper, 140 S. Ct. 994, 1007 (2020).

180. 17 U.S.C. § 110.

181. *Id.* § 110(6).

182. *Id.* § 110(2).

183. *Id.* § 110(8).

184. *Id.* § 110(1).

185. Allen v. Cooper, 140 S. Ct. 994, 1007 (2020); *see* Brief of the Copyright Alliance and the Chamber of Commerce, *supra* note 160, at 18 (discussing both the sections of 17 U.S.C. § 110 above, as well as §§ 108 and 504(c)(2), pertaining to research performed in libraries and reproductions by educational institutions, respectively).

legitimate under" its enforcement power.[186] In the context of the Voting Rights Act, for example, this took the form of additional procedural requirements that would only remain in effect for seven years and were imposed upon only jurisdictions with a history of discriminatory practices.[187] Admittedly, no such restrictions were present in the CRCA, but it is difficult to imagine what they might actually have been. Unlike voting rights, which are tied to a particular place because one can only vote in one location, the rights of copyright holders require broad protection across different geographical regions because of the vast availability of copyrighted works in both electronic and hard copy formats. As a result, limiting the CRCA's impact to certain states or a given temporal window would be "impractical to the point of [being] self-defeating" because it would undermine the legislation's ability to secure "uniform, dependable, nationwide protection."[188] Furthermore, while the *Boerne* Court noted that such restrictions might be helpful for measuring the proportionality of congressional action, "[t]his [was] not to say, of course, that [Section] 5 legislation requires termination dates, geographic restrictions, or egregious predicates."[189] As a result, it should not count against the CRCA's proportionality that it did not include any temporal or geographic restrictions.

The Court was too quick to assume that the scope of the CRCA was necessarily overly broad. By relying on the similarities between copyright and patents, it oversimplified its inquiry. The Court too heavily relied on the fact that it had previously found the scope of the Patent Remedy Act to be cripplingly indiscriminate in *Florida Prepaid*, thus overlooking contextual limitations that explain why the relevant legislative record was limited. Together, these factors suggest relevant nuances that the Court did not meaningfully consider in its application of the congruence and proportionality test.

B. CONGRESSIONAL REMEDIES FOR THE COURT'S DECISION IN *ALLEN V. COOPER*

Ultimately, *Allen v. Cooper* leaves Congress with a command to go back to the drawing board. Acknowledging that the "congruence and proportionality" test had not been invented until after the CRCA was enacted, the Court invited the legislature to "enact a proportionate response" now that it "know[s] [the] rules" for abrogating sovereign immunity under its enforcement powers.[190] If

186. City of Boerne v. Flores, 521 U.S. 507, 533 (1997), *superseded by statute*, Religious Land Use and Institutionalized Persons Act of 2000, Pub. L. No. 106-274, 114 Stat. 803.

187. *See id.*

188. Reply Brief for Petitioners at 24, *Allen*, 140 S. Ct. 994 (No. 18-877).

189. *Boerne*, 521 U.S. at 533.

190. *See Allen*, 140 S. Ct. at 1007.

Congress jumps through judicial loopholes, the Court suggested, a new law might "bring digital Blackbeards to justice."[191] While useful as a call to action, this advice showcases the shortcomings of the Court's analysis in *Allen v. Cooper*.

The Court, or at least the majority opinion authored by Justice Kagan, seemed confident that Congress can, and perhaps even should, pass a new law that will effectively duplicate the CRCA to achieve its intended purpose.[192] The Court offered that when Congress passed the CRCA, it "likely did not appreciate the importance of linking the scope of its abrogation to the redress or prevention of unconstitutional injuries—and of creating a legislative record to back up that connection."[193] Should Congress conduct its research again and build a longer list of constitutional violations, "it may enact a proportionate response," and "[t]hat kind of tailored statute can effectively stop States from behaving as copyright pirates."[194] Notably, the Court did not dispute that state copyright infringement is a problem or even that it can be remedied under Congress's Section Five powers. The Court did not consider abrogation under these circumstances fundamentally unjustified or somehow an affront to the balance between state and federal powers. Instead, while it may have found the CRCA's legislative record too slight, the Court's frustration was not that the record's shortcomings suggest a lack of unconstitutional violations but rather that those violations simply were not appropriately documented as of 1990. Implicit within Justice Kagan's majority opinion is an acknowledgement that curtailing state infringement is a worthwhile endeavor but one that will require Congress return to a topic it already identified, studied, and addressed through legislation a full three decades ago—all in the name of satisfying a vague congruence and proportionality test with uncertain quantitative demands.

Congress, it seems, heard the Court's advice and is willing to follow Justice Kagan's prescription, even if doing so is not an especially productive use of congressional resources.[195] Understanding the significance of the Court's holding in *Allen v. Cooper* and noting that they were "concerned about the impact [it] may have on American creators and innovators," Senators Thom Tillis (R-North Carolina) and Patrick Leahy (D-Vermont) took up Kagan's

191. *See id.*
192. *Id.*
193. *Id.*
194. *See id.*
195. Letter from Sens. Thom Tillis & Patrick Leahy to Acting Register of Copyrights and Dir. of the U.S. Copyright Off. Maria Strong (Apr. 28, 2020), https://www.copyright.gov/policy/state-sovereign-immunity/letter.pdf.

challenge.[196] Barely a month after the Court's holding, they sent a pair of letters to both the U.S. Copyright Office and the U.S. Patent and Trademark Office, requesting a study on the impact of infringement by states.[197] Grateful that *Allen v. Cooper* offered a "blueprint for how to validly abrogate State sovereign immunity from certain copyright infringement claims," the Senators asked for help in "identifying a pattern of unconstitutional infringement" with consideration for the extent to which such infringement resulted from "intentional or reckless conduct."[198] The Senators asked that the results be delivered so that "Congress can evaluate whether legislative action needs to be taken."[199] Regardless of those results, neither *Boerne*, nor *Florida Prepaid*, nor *Allen v. Cooper* provided a quantitative benchmark for developing a sufficient legislative record. As a result, Senators Tillis and Leahy were left with a relatively complicated task: creating a robust legislative record that must include a certain number of examples (at the risk of sinking the legislation otherwise) but without a clear mandate as to how many examples will prove sufficient.

A new legislative record and a more tailored statute may satisfy the congruence and proportionality test, but it is far from clear why satisfying that test is useful. Considering that the Court appeared willing to recognize that the CRCA addressed a legitimate problem, contending only that it failed under the parameters of the congruence and proportionality test, one must wonder whether it is the test and not the legislation that is insufficient. The test requires that congressional actions abrogating sovereign immunity be assessed only as they existed at the moment of their enactment, which curtails Congress's ability to pass prophylactic legislation. Even where Congress successfully identifies a problem barreling forward like a train on its tracks, the congruence and proportionality test prevents Congress from stepping out of its way and instead demands that Congress only provide a remedy after the point of collision.[200]

Additional scholarship should examine whether or not the "congruence and proportionality" test is the proper tool for assessing the abrogation of sovereign immunity in the first place.[201] One can imagine many different

196. *Id.* at 1.

197. *Id.*

198. *Id.* at 1–2.

199. *Id.* at 2.

200. Limiting Congress's Fourteenth Amendment powers to actions justified by a historical record would require that the legislature "await maximal state violations before laboring to fashion minimalist, patchwork federal remedies that *might* withstand judicial review." *See* Reply Brief for Petitioners, *supra* note 188, at 2.

201. For discussion and criticism of the "congruence and proportionality" test, see, for example, Ren, *supra* note 122, at 1461; Randolph J. Haines, *Getting to Abrogation*, 75 AM. BANKR. L.J. 447 (2001); Joseph M. Pellicciotti & Michael J. Pellicciotti, *Sovereign Immunity &*

models that might be more efficient. Alternatives might involve recalibrating the Court's analysis to place less emphasis on the historical record, allowing Congress to be more forward-thinking with its legislation, instead of being constrained by the past. Perhaps rather than asking the Court to second-guess Congress, the test should be replaced with judicial deference toward a rebuttable presumption in favor of congressional abrogation, as one scholar has advocated.[202] To the other extreme, should the Court remain committed to its quantitative analysis, its approach would be more actionable if it were accompanied by an explicit benchmark. Regardless of the exact nature of the solution, there is ample room for improvement.

Ultimately, the Court's acknowledgement that a new and tailored law might bring digital Blackbeards to justice signaled both that Congress can abrogate sovereign immunity to protect copyright holders from infringing states and that doing so might benefit copyright holders.[203] Congress seems poised to try to do so, even at the expense of its own time and energy.

V. CONCLUSION

Senators Tillis and Leahy may have found the Court's decision in *Allen v. Cooper* to be a useful blueprint,[204] but that assessment is generous. In applying *Boerne*'s congruence and proportionality test, an inquiry originally intended to assess the general appropriateness of congressional efforts to abrogate, sovereign immunity became rigid and quantitative. The Court's decision skirted meaningful analysis of the CRCA's scope, deciding instead to gesture to a decision from twenty years ago in *Florida Prepaid*, which concerned a different strand of intellectual property law, without considering the meaningful differentiators that separate copyrights from patents. Such distinctions could have distinguished *Allen v. Cooper* from *Florida Prepaid* and should not have been left unconsidered. Having dismissed the question of scope entirely, the Court placed the entire weight of its decision-making on a purely quantitative inquiry into the CRCA's legislative record. Yet, while the Court suggests that congruence and proportionality can be addressed by simple counting, it does not offer a quantitative threshold it might instead have

---

Congressionally Authorized Private Party Actions Against the States for Violation of Federal Law: A Consideration of the U.S. Supreme Court's Decade Long Decisional Trek, 1996–2006, 59 BAYLOR L. REV. 623 (2007); William J. Rich, *Hiding in Plain View: A Path Around Sovereign Immunity for State Government Employees*, 25 WM. & MARY BILL RTS. J. 1179 (2017).

202. This is the solution for which Y. Frank Ren advocates in his own Note, arguing that a rebuttable presumption would be a "return[] to a deferential approach that nevertheless retains some space for state sovereignty." Ren, *supra* note 122, at 1461.

203. *See* Allen v. Cooper, 140 S. Ct. 994, 1007 (2020).

204. *See supra* text accompanying notes 195–99.

found satisfying. It also ignored relevant nuances that explain the shortcomings of the legislative record.

Senators Tillis and Leahy are left to construct a new legislative record that they have been told must include more examples in order to meet an imagined threshold, though exactly how many more examples are required remains unclear. In the meantime, while we may not actually see *Pirates of the Caribbean* and other blockbusters streaming on www.nc.gov, artists and inventors alike will remain adrift and at the mercy of whichever state pirates decide to flex their Eleventh Amendment muscles.

DOES TRAINING AI VIOLATE COPYRIGHT LAW?

Jenny Quang†

I. INTRODUCTION

From targeted advertising to search optimization, machine learning (ML) algorithms are increasingly prevalent in our daily lives.[1] Artificial intelligence (AI) and copyright law intersect when copyrighted data are used to train machines to learn, reason, and act as humans do. The development of some AI technologies, such as autonomous vehicle software, facial recognition algorithms, and smart assistants, requires massive data mining of video, photos, or text that may be subject to copyright.[2] However, the downloading and storage of copyrighted data to train machine learning models may violate copyright law and impose undue liability on AI developers. This Note identifies how Congress can drive AI innovation by clarifying copyright law and adopting a safe harbor for data mining.

Recently, a 3-D imaging firm filed copyright claims accusing Meta and Princeton University of illegally downloading its data for use in scene-recognition AI projects.[3] Although it is uncertain how the court will rule on this novel issue, it is clear that the stakes are high—the market for scene-recognition technology is estimated to reach $60 billion by 2025, and the cost of the data collection at issue is in the millions of dollars.[4] Alongside potentially steep damages, the plaintiffs are seeking injunctive relief from further acts of copyright infringement and destruction of all infringing copies.[5] The outcome of this case, and others like it that may follow, could handicap new AI technologies.

Most copyright stakeholders believe that the fair use doctrine is sufficient to defend data mining uses in the United States.[6] However, fair use is an undesirable form of protection because the doctrine has been stretched beyond

---

DOI: https://doi.org/10.15779/Z38XW47X3K

† J.D., 2022, University of California, Berkeley, School of Law.

1. *See* ANDREW NG, MACHINE LEARNING YEARNING 6 (2018).
2. *See infra* Part II.
3. *See* Complaint for Copyright Infringement at 27–30, UAB "Planner5D" v. Facebook, Inc., No. 3:20-cv-08261 (N.D. Cal. Nov. 23, 2020).
4. *Id.* at 3.
5. *Id.* at 30.
6. U.S. PAT. & TRADEMARK OFF., PUBLIC VIEWS ON ARTIFICIAL INTELLIGENCE AND INTELLECTUAL PROPERTY POLICY 26 (2020) ("Most commenters found that existing law does

its limits and is often unpredictable in practice. This uncertainty disproportionately handicaps smaller actors; the possibility of copyright liability may intimidate new AI developers aiming to compete with established tech giants and could influence how researchers conduct their work. In a recent example, an AI-based legal research startup shut down amid financial pressures brought on by a copyright infringement lawsuit.[7] And statutory fines, which range from $200 to $150,000 per work,[8] are unnecessarily crippling, especially when a single machine learning model may be trained using thousands to millions of works.[9] Establishing a clear right to use copyrighted materials for data mining is consistent with the goals of copyright law and would remove barriers to innovation.

A safe harbor for data mining is warranted because the use of data to develop functional AI technologies is *fundamentally* not an act of infringement. The seminal Supreme Court copyright case *Baker v. Selden* distinguished a copyrighted work from its material form and showed that not all uses of a work's material form are acts of copyright infringement.[10] Copyright infringement requires not just copying of a work's material form but also the unauthorized use of the work for its expressive purpose. Merely technical or non-communicative uses are not uses of a work for its expressive purpose and therefore are not copyright infringement.

---

not require modification, as fair use is a flexible doctrine and is capable of adapting to the use of copyrighted works in an AI context.").

7. Lyle Moran, *ROSS Intelligence Will Shut Down Amid Lawsuit from Thomson Reuters*, ABA J. (Dec. 11, 2020, 11:50 AM), https://www.abajournal.com/news/article/ross-intelligence-to-shut-down-amid-thomson-reuters-lawsuit ("Litigation is expensive—no matter how speculative the claims against you nor how worthy your position[.] . . . With our company ensnared by this legal battle, we have been unable to raise another round of funding to fuel our development and marketing efforts. Our bank account is running out, and we must cease operations in the new year." (quoting ROSS Intelligence)).

8. 17 U.S.C. § 504.

9. *See, e.g.*, Tom B. Brown et al., Language Models are Few-Shot Learners 1 (July 22, 2020) (unpublished manuscript), https://arxiv.org/pdf/2005.14165.pdf ("[T]his method still requires task-specific fine-tuning datasets of thousands or tens of thousands of examples."); Chen Sun, Abhinav Shrivastava, Saurabh Singh & Abhinav Gupta, Revisiting Unreasonable Effectiveness of Data in Deep Learning Era 3 (Aug. 4, 2017) (unpublished manuscript), https://arxiv.org/pdf/1707.02968.pdf (training on a dataset of "300M images and 375M labels").

10. *See* Baker v. Selden, 101 U.S. 99 (1879); ABRAHAM DRASSINOWER, WHAT'S WRONG WITH COPYING? 88-100 (2015). In *Baker*, the copyrighted work was a book that explained a novel accounting method, and its material form included the accounting forms used as part of the explanation. *See* 101 U.S. at 100–01. The defendant's non-communicative copying of the accounting forms to perform the novel accounting method did not constitute infringement. *See id.* at 107.

Likewise, to download copyrighted images and text for data mining is to make copies for a different purpose. Training a machine learning model with this copyrighted data does not infringe because the data are not redistributed or recommunicated to the public. Copyright protects creative expression, but model training extracts unprotectable ideas and patterns from data. Thus, data mining uses of copyrighted works need not even be subject to a fair use analysis.

Copyright law distinguishes between creative expression and unprotectable ideas.[11] In this Note, "data mining" will specifically refer to the mining of *expressive* data (i.e., literary works, photographs, video) for *functional*, or non-expressive, purposes. Expressive applications of data mining (i.e., AI-generated art, music, and literature) are outside of the scope of this analysis. Due to an increasing interest in artificial intelligence technologies,[12] this Note will focus on copyright and data mining in an artificial intelligence and machine learning context.

This Note argues that Congress should adopt a safe harbor for data mining of copyrighted works because (1) it is fundamentally not an act of infringement and (2) legal uncertainty currently exists with the fair use doctrine. The Note presents the argument as follows. Part II explains how data mining implicates copyright law and reviews how courts have applied the fair use doctrine unevenly in similar technological contexts. Next, Part III argues that data mining is fundamentally not copyright infringement and that a safe harbor is needed to clear up legal uncertainty that results from the fair use doctrine. Finally, Part IV reviews data mining exceptions in foreign jurisdictions and outlines a legislative proposal for a safe harbor in U.S. copyright law.

II. BACKGROUND

A. DATA MINING DEFINITIONS

Broadly, data mining involves improving future decisions by finding patterns in data collected from past events.[13] Interest in the field has rapidly grown with advancements in data collection and storage, the use of machine learning to process this data, and the falling cost of computational power.[14]

11. *See* 17 U.S.C. § 102(b).
12. *See, e.g.*, Exec. Order No. 13,859, 84 Fed. Reg. 3967 (Feb. 14, 2019).
13. *See* Tom M. Mitchell, *Machine Learning and Data Mining*, 42 COMMC'NS ACM 30, 30 (1999).
14. *See id.*

Machine learning is a process of "[u]sing data to answer questions."[15] Data is key to this process, as machine learning algorithms use statistics to find insights hidden in massive amounts of data.[16] These algorithms are also responsible for the majority of recent advancements and applications in artificial intelligence, a field that trains machines to learn, reason, and act as humans do.[17]

Two subfields of AI learn from expressive data and will be used as examples throughout this Note. Computer vision is a subfield of AI that studies visual data.[18] A computer is taught how to understand the world through images and video.[19] Applications of such technology include autonomous vehicles, facial recognition, and medical imaging.[20] Natural language processing (NLP) uses computational techniques to understand and represent human languages.[21] Large datasets of text are required to train models to understand and generate new text.[22] Applications include translation, information extraction, and question answering.[23]

The majority of artificial intelligence applications require machine learning algorithms and models.[24] A machine learning algorithm will find patterns from past data and output a machine learning model that captures these patterns.[25] The model can then be used to make predictions on new data.[26] After first defining the problem (i.e., what the model should predict), an AI developer must collect data, prepare the data (including labeling or annotating), choose a model architecture; train the model with the annotated data ("training data")

15. Yufeng G, *What is Machine Learning?*, TOWARDS DATA SCI. (Aug. 24, 2017), https://towardsdatascience.com/what-is-machine-learning-8c6871016736.

16. *Id.*

17. Karen Hao, *What is Machine Learning?*, MIT TECH. REV. (Nov. 17, 2018), https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart; Karen Hao, *What is AI? We Drew You a Flowchart to Work it Out*, MIT TECH. REV. (Nov. 10, 2018), https://www.technologyreview.com/2018/11/10/139137/is-this-ai-we-drew-you-a-flowchart-to-work-it-out.

18. Ben Dickson, *What is Computer Vision?*, PCMAG (Feb. 9, 2020), https://www.pcmag.com/news/what-is-computer-vision.

19. *Id.*

20. *Id.*

21. K. R. CHOWDHARY, FUNDAMENTALS OF ARTIFICIAL INTELLIGENCE 604 (2020).

22. *Id.* at 608.

23. *Id.* at 608–09.

24. *See* Hao, *What is AI?*, *supra* note 17.

25. *Training ML Models*, AMAZON WEB SERVS., https://docs.aws.amazon.com/machine-learning/latest/dg/training-ml-models.html.

26. *Id.*

using a machine learning algorithm, and test the model with new data ("testing data").[27]

In practice, AI developers will often fine-tune open source models that have been pre-trained on extremely large datasets.[28] For example, a developer who wants to train a computer to recognize the movements of a surgeon's hand during an operation will start with a model that has already been trained using ImageNet, a database containing over 14 million images of common objects (i.e., cat, dog, car).[29] This pre-trained model will already be able to find edges and shapes and recognize common objects.[30] The developer will then compile her own dataset of surgical images (containing hundreds to thousands of photographs) and fine-tune the model on that training data. The new model would then be able to better recognize surgery-specific objects, such as surgical tools and gloved hands. This fine-tuning ("transfer learning") process contributes to the democratization of AI technology by allowing smaller innovators to more easily develop their own products when they build off previous work.[31]

However, the threat of copyright liability has the potential to stifle this democratization process. Starting with a pre-trained model can reduce the data points needed from millions to thousands.[32] But an AI developer will still need to download some data to fine-tune her model. Because machine learning

---

27. *See id.*; Zaid Alissa Almaliki, *Do You Know How to Choose the Right Machine Learning Algorithm Among 7 Different Types?*, TOWARDS DATA SCI. (Mar. 19, 2019), https://towardsdatascience.com/do-you-know-how-to-choose-the-right-machine-learning-algorithm-among-7-different-types-295d0b0c7f60; Yufeng G, *The 7 Steps of Machine Learning*, TOWARDS DATA SCI. (Aug. 31, 2017), https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e.

28. *See* Naveen Joshi, *How to Fine-tune Your Artificial Intelligence Algorithms*, ALLERIN (Jan. 13, 2020), https://www.allerin.com/blog/how-to-fine-tune-your-artificial-intelligence-algorithms; Aidan Boyd, Adam Czajka & Kevin Bowyer, *Deep Learning-Based Feature Extraction in Iris Recognition: Use Existing Models, Fine-tune or Train From Scratch?*, 10 IEEE INT'L CONF. ON BIOMETRICS THEORY, APPLICATIONS AND SYS., SEPT. 2019, at 1.

29. *Summary and Statistics*, IMAGENET (Apr. 30, 2010), http://image-net.org/about-stats.

30. *See* Sebastian Ruder, *NLP's ImageNet Moment Has Arrived*, THE GRADIENT (July 8, 2018), https://thegradient.pub/nlp-imagenet ("Importantly, knowledge of edges, structures, and the visual composition of objects is relevant for many CV tasks . . . . A key property of an ImageNet-like dataset is thus to encourage a model to learn features that will likely generalize to new tasks in the problem domain.").

31. *See, e.g.*, Vikash Gupta et al., Democratizing Artificial Intelligence in Healthcare: A Study of Model Development Across Two Institutions Incorporating Transfer Learning. 1 (Sept. 25, 2020) (unpublished manuscript), https://arxiv.org/ftp/arxiv/papers/2009/2009.12437.pdf ("[I]f a well-tested deep learning model from another institution is available, it can be adopted for use in model fine-tuning on a relatively smaller local dataset, thereby allowing institutions with fewer resources to also directly participate in AI development.").

32. *See* Brown et al., *supra* note 9, at 1 ("[T]his method still requires task-specific fine-tuning datasets of thousands or tens of thousands of examples.").

models are trained on massive amounts of data, this step is often automated.[33] In computer vision, images and videos may be scraped from the internet (e.g., Google Images, YouTube) and incorporated into large datasets for training.[34] In NLP, corpora of text are usually mined from the internet (e.g., Wikipedia, Gmail).[35] The downloading of data that may be subject to copyright presents a potential violation of copyright law.

B. COPYRIGHT AND DATA MINING

The Intellectual Property Clause of the U.S. Constitution grants Congress the power to "promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries."[36] Copyright protection serves a dual purpose: to motivate the creative activity of authors and artists and to advance public welfare through access to expressive works.[37] In exchange for the creation of literary and artistic works, authors and artists are granted certain exclusive rights, including the rights to reproduction, distribution, and preparation of derivative works, for a limited period.[38] To enforce these rights, copyright owners can sue alleged infringers for monetary relief, including statutory damages, which can range from $200 to $150,000 per work depending on the willfulness of infringement.[39] These steep fines provide an incentive for copyright owners to enforce copyright interests when harm from infringement can be difficult to prove.[40]

---

33. *See* Lewis Chou, *Four Basic Ways to Automate Data Extraction*, TOWARDS DATA SCI. (Oct. 12, 2019), https://towardsdatascience.com/four-basic-ways-to-automate-data-extraction-3151064dc110.

34. *See, e.g.*, Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar & Li Fei-Fei, *Large-scale Video Classification with Convolutional Neural Networks*, 2014 PROC. IEEE CONF. ON COMPUT. VISION & PATTERN RECOGNITION 1725 (using dataset of 1 million YouTube videos); Adrian Rosebrock, *How to Create a Deep Learning Dataset Using Google Images*, PYIMAGESEARCH (Dec. 4, 2017), https://www.pyimagesearch.com/2017/12/04/how-to-create-a-deep-learning-dataset-using-google-images.

35. *See, e.g.*, Evgeniy Gabrilovich & Shaul Markovitch, *Wikipedia-based Semantic Interpretation for Natural Language Processing*, 34 J. A.I. RSCH. 443 (2009); Mia Xu Chen et al., Gmail Smart Compose: Real-Time Assisted Writing 3 (May 17, 2019) (unpublished manuscript), https://arxiv.org/pdf/1906.00080.pdf (using dataset of "user-composed e-mails").

36. U.S. CONST. art. I, § 8, cl. 8.

37. *See* 1 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 1.03 (2021).

38. 17 U.S.C. § 106.

39. 17 U.S.C. § 504.

40. Peter S. Menell, *This American Copyright Life: Reflections on Re-equilibrating Copyright for the Internet Age*, 61 J. COPYRIGHT SOC'Y U.S.A. 235, 306–07 (2014).

Artificial intelligence and copyright law intersect when expressive data is used to train machines to learn, reason, and act as humans do. Under § 106, the reproduction right grants a copyright holder the exclusive right to make copies of the protected work.[41] As explained in Section II.A, developers often use images, video, and text downloaded from the internet to train machine learning models.[42] The downloaded data are essentially copies that are stored via hard drives, cloud storage, or other data repositories.[43] Given the large volume of data—often scraped from the internet en masse—that is needed to train a machine learning model, it is likely that some of that training data is protected by copyright.[44] Because copyright infringement is a strict liability offense,[45] it does not matter if a developer was unaware that copyrighted works existed in the dataset.

The download and storage of data creates copies that may be subject to copyright. The Copyright Act states that copies are "material objects . . . in which a work is fixed *by any method now known or later developed*, and from which the work can be perceived, reproduced, or otherwise communicated, either directly or *with the aid of a machine or device*."[46] Thus, data that are solely used by machines during model training would be treated just the same as works perceived by humans under copyright law. Further, a work is "fixed" in a tangible medium when its embodiment in a copy is "*sufficiently permanent or stable* . . . for a period of more than transitory duration."[47]

Copyright caselaw dealing with memory storage and RAM shows that training data downloaded onto a computer's hard drive would be sufficiently fixed under copyright law. In *Stern Electronics, Inc. v. Kaufman*, the Second Circuit held that audiovisual works of a video game that are permanently embodied in "memory devices" of the game were protectable under copyright law.[48] Later, in *MAI Systems Corp. v. Peak Computer, Inc.*, the Ninth Circuit held that loading

41. 17 U.S.C. § 106.

42. *See supra* Section II.A.

43. *See* Jim Dowling, *Guide to File Formats for Machine Learning: Columnar, Training, Inferencing, and the Feature Store*, TOWARDS DATA SCI. (Oct. 25, 2019), https://towardsdatascience.com/guide-to-file-formats-for-machine-learning-columnar-training-inferencing-and-the-feature-store-2e0c3d18d4f9.

44. Contemporary photographs are presumptively under copyright. Brammer v. Violent Hues Prods., LLC, 922 F.3d 255, 266 (4th Cir. 2019).

45. 17 U.S.C. § 106; *see, e.g.*, *Brammer*, 922 F.3d at 265 ("As a basic matter, copyright infringement is a strict liability offense, in which a violation does not require a culpable state of mind.").

46. 17 U.S.C. § 101 (emphases added).

47. *Id.* (emphasis added).

48. 669 F.2d 852, 855–56 (2d Cir. 1982).

of copyrighted software into a computer's random-access memory (RAM) created an unauthorized reproduction under the Copyright Act.[49] Thus, fixation of a copy in either permanent or temporary memory storage is sufficiently permanent and stable to satisfy the Copyright Act.

So, even if training data is not to be *permanently* stored on a hard drive, like the audiovisual works in *Stern Electronics* (a developer could, in theory, delete the training data from her computer once she finished training her model), the storage could still constitute unauthorized copying because a hard drive is more stable than RAM.[50] And what about training data that is not stored on a personal hard drive, but on a cloud storage service?[51] Technically, cloud storage is still the storage of digital data on physical hard drives, but in off-site locations that are accessed via the internet or a private network connection.[52] The Copyright Act also accounts for methods of fixation "now known or later developed," which would allow for extrapolation to future data storage devices, like solid-state drives.[53] Thus, whether training data is downloaded onto a personal computer or on the cloud, the copies should be sufficiently fixed in a medium to implicate copyright law.

An allegedly infringing developer could argue that although she downloaded copyrighted images to train her computer vision model, there is no trace of the copyrighted works in her final model. If a human can learn from reading books without infringing copyright, why can't a machine similarly learn from training data? However, when that training data is comprised of data downloaded from the internet, copies are necessarily created in the process of training a machine learning model.

It is this intermediate copying that differentiates machine learning from human learning and explains why the former implicates copyright law. The Ninth Circuit in *Sega Enterprises v. Accolade, Inc.* held that the intermediate cop-

49. 991 F.2d 511, 519 (9th Cir. 1993). *But see* Cartoon Network LP, LLLP v. CSC Holdings, Inc., 536 F.3d 121, 130 (2d Cir. 2008) (holding that a period of 1.2 seconds was not sufficiently fixed to constitute copyright infringement).

50. *See Stern Elecs.*, 669 F.2d at 856; *MAI Sys. Corp.*, 991 F.2d at 519; John Cruickshank, *The Difference Between RAM and Hard Drive*, TECHTORIUM (May 8, 2020, 8:49 PM), https://techtorium.ac.nz/the-difference-between-ram-and-hard-drive.

51. *See, e.g.*, *Working with Cloud Storage*, GOOGLE CLOUD (Nov. 16, 2020), https://cloud.google.com/ai-platform/training/docs/working-with-cloud-storage.

52. *See* IBM Cloud Education, *Cloud Storage*, IBM (June 24, 2019), https://www.ibm.com/cloud/learn/cloud-storage.

53. 17 U.S.C. § 101; *see SSD vs. HDD: Which is Best for You?*, INTEL, https://www.intel.com/content/www/us/en/products/docs/memory-storage/solid-state-drives/ssd-vs-hdd.html.

ying of protected computer code could constitute copyright infringement, regardless of whether the end product of the copying also infringed.[54] Applying that precedent, the District of Nevada in *Tiffany Design, Inc. v. Reno-Tahoe Specialty* found that intermediate copying through the scanning of protected photographs constituted copyright infringement as a matter of law, without even determining whether the end product, an artistic depiction of the Las Vegas Strip, was substantially similar.[55]

Despite the intermediate copying, the defendants in *Sega* and *Sony Computer Entertainment v. Connectix Corp.*, a more recent computer case also involving the intermediate copying of protected code, prevailed under the fair use defense.[56] The fair use doctrine, although riddled with imperfections, currently represents an AI developer's best chance at defeating a claim of copyright infringement in court.

C. THE FAIR USE DOCTRINE

Courts have applied the fair use doctrine to find non-infringement in technology cases. Codified in 17 U.S.C. § 107, the fair use doctrine is an affirmative defense to a claim of copyright infringement.[57] The analysis considers four factors:

> (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
>
> (2) the nature of the copyrighted work;
>
> (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
>
> (4) the effect of the use upon the potential market for or value of the copyrighted work.[58]

The fair use doctrine has led to unexpected results when these four factors are applied to new technological contexts.[59]

For example, the Ninth Circuit in *Sega* ruled that Accolade's reverse engineering of Sega's Genesis video game programs qualified as fair use even

54. 977 F.2d 1510, 1519 (9th Cir. 1992).

55. 55 F. Supp. 2d 1113, 1121 (D. Nev. 1999).

56. 17 U.S.C. § 107; *see Sega*, 977 F.2d at 1514; Sony Computer Ent., Inc. v. Connectix Corp., 203 F.3d 596, 598 (9th Cir. 2000).

57. 17 U.S.C. § 107.

58. *Id.*

59. *See Sega*, 977 F.2d at 1527 ("We are not unaware of the fact that to those used to considering copyright issues in more traditional contexts, our result may seem incongruous at first blush.").

though the intermediate copying of protected code was necessary.[60] The court stated that the first factor, "the purpose and character of the use," weighed in favor of Accolade because Accolade's ultimate goal was the release of Genesis-compatible games for sale—a "legitimate, essentially non-exploitative purpose."[61] Accolade's copying of protected code was necessary for extracting the functional requirements for Genesis compatibility so that it could make its existing games compatible with the Genesis console.[62]

Relatedly, under the second factor, "the nature of the copyrighted work," Sega's video game programs were afforded a lower degree of protection than traditional literary works because they contained unprotected, functional aspects like Genesis compatibility.[63] The second factor analysis essentially incorporated the idea-expression dichotomy of copyright law—that ideas and functional concepts are not protected by copyright.[64] Additionally, the fourth factor, the effect of the use on the market, also weighed in favor of Accolade because its entry into the market for Genesis-compatible games would promote creative expression in video games.[65] Finally, under the third factor, the amount and substantiality of the copied portion, copying of the entire code was not enough to outweigh the effect of the other three factors.[66] Thus, the Ninth Circuit in *Sega* ruled that Accolade's reverse engineering was fair use.

More recently, the first factor of the fair use analysis has tended to overshadow the second and third factors because of the introduction of the doctrine of transformative use.[67] As explained by Judge Leval in his seminal law review article, central to the first factor of the fair use analysis is whether the challenged use is *transformative*, that is, whether the original is used as "raw material, transformed in the creation of new information, new aesthetics, new insights and understandings."[68] Since its introduction, the percentage of transformative use decisions in fair use cases has increased from 8% in 1991 to

60. *Id.* at 1520.
61. *Id.* at 1522–23.
62. *Id.*
63. *Id.* at 1524–26.
64. *See id.* at 1519–20; 17 U.S.C. § 102(b).
65. *Sega*, 977 F.2d at 1524.
66. *Id.* at 1526–27.
67. Jane C. Ginsburg, *Fair Use in the United States: Transformed, Deformed, Reformed?*, 2020 SING. J. LEGAL STUD. 265, 277 n.56 (2020).
68. Pierre N. Leval, Commentary, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1111 (1990).

around 90% in recent years.[69] And out of a large sample of decisions, 94% that found transformative use also ultimately found fair use.[70]

Relevant to this Note's data mining context, this transformative analysis has also been applied to cases involving internet search engines that functionally utilize text and images. For instance, the Ninth Circuit in *Perfect 10, Inc. v. Amazon.com, Inc.* held that the use of copyrighted images to display search engine results was fair use because the use of thumbnails was transformative.[71] While the images originally served "entertainment, aesthetic, or informative function[s]," the search engine transformed the image into a "pointer directing a user to a source of information."[72] Similarly, the Second Circuit in *Authors Guild, Inc. v. Google, Inc.* held that Google's digitizing of copyrighted texts was fair use because the Google Books search engine was highly transformative.[73] Although Google scanned full copyrighted texts, it displayed only snippets of text, which acted as "pointers directing users to a broad selection of books."[74] These cases suggest that courts are likely to find fair use exceptions to copyright infringement where the end product serves a functional purpose and provides social utility.

For an AI developer, the search engine cases provide some precedent for structuring a fair use defense. If internet search engines were determined to be transformative, it is easy to imagine how AI products could qualify as highly transformative uses of expressive data (i.e., surgery-assistant robots, self-driving cars, translation services). Images, video, and text are fed as "raw material" to train models that teach machines to make decisions as humans do—creating "new insights and understandings."[75] Even the massive scale reproductions of images or text performed on the computer are not prohibitive as Google scanned "tens of millions of books" for use in its search engine.[76] And *Authors Guild, Inc. v. Google, Inc.* showed that a defendant's highly commercial nature does not outweigh a product's transformative value.[77] Thus, if an AI developer were downloading thousands of copyrighted images to train a computer vision

69. Jiarui Liu, *An Empirical Study of Transformative Use in Copyright Law*, 22 STAN. TECH. L. REV. 163, 174 (2019).

70. *Id.* at 180.

71. 508 F.3d 1146, 1167 (9th Cir. 2007).

72. *Id.* at 1165.

73. 804 F.3d 202, 216 (2d Cir. 2015).

74. Authors Guild, Inc. v. Google Inc., 954 F. Supp. 2d 282, 291 (S.D.N.Y. 2013), *aff'd*, 804 F.3d at 202.

75. Leval, *supra* note 68, at 1111.

76. *Authors Guild*, 804 F.3d at 206.

77. *Id.* at 209.

model, perhaps she too could rely on the fair use defense, even if she intended to make a commercial product out of it.

However, the fair use doctrine fluctuates in its friendliness to technology.[78] In a more recent search engine case, the Ninth Circuit held that Zillow's use of copyrighted photographs in its apartment listing searches was not fair use.[79] Zillow's database featured photos of artfully-designed rooms that could be sorted using "various criteria, like room type, style, cost, and color."[80] Although the search engine made the photographs functionally searchable, the court found that it did not fundamentally change their original purpose to "artfully depict rooms and properties," and thus preserved the original photos' "inherent character."[81] Additionally, because the plaintiff was "actively exploring" a market for licensing its photographs, the fourth factor weighed against Zillow.[82] Similarly, the Second Circuit in *Fox News Network, LLC v. TVEyes, Inc.* found no fair use for a TV clip search engine that was "somewhat transformative" but usurped a potential market for Fox to license its works.[83]

These recent cases suggest that the fair use defense may not be so straightforward when applied to AI technologies.[84] Although it may have once been enough to lean on the first factor and point out that an end use is transformative, *VHT*, *Inc.* and *TVEyes* show that courts may heavily consider the potential effect that the use has on licensing markets.[85] Thus, a developer may be penalized under the fair use analysis for depriving a copyright owner of the opportunity to license her works to be used in training datasets. And the case law illustrates that copyrighted photographs do not receive weaker protection

78. *See* Michael W. Carroll, *Copyright and the Progress of Science: Why Text and Data Mining is Lawful*, 53 U.C. DAVIS L. REV. 893, 936 (2019).

79. VHT, Inc. v. Zillow Grp., Inc., 918 F.3d 723, 744 (9th Cir 2019). Another recent case weighs unfavorably on the use of copyrighted photos for informational purposes. In *Brammer v. Violent Hues Productions*, the Fourth Circuit held that a film website's use of a copyrighted photograph, found from a Google Images search, was not fair use. 922 F.3d 255, 269 (4th Cir. 2019). The defendant's claim that the photograph was used to provide information regarding a tourist attraction was not convincing, otherwise "virtually all illustrative uses of photography would qualify as transformative." *Id.* at 264. The court also noted that photographs have long received "thick copyright protection" even though they "capture images of reality." *Id.* at 266–67.

80. *VHT, Inc.*, 918 F.3d at 730.

81. *Id.* at 742.

82. *Id.* at 744.

83. *See* 883 F.3d 169, 178–80 (2d Cir. 2018).

84. *See VHT, Inc.*, 918 F.3d at 740 ("While we can discern certain animating principles bridging cases in this area, the doctrine has hardly followed a straight, or even slightly curved, line.").

85. *See id.* at 744; *TVEyes*, 883 F.3d at 180–81.

even if used for "informational" purposes.[86] Even if an AI developer tries to invoke the idea-expression dichotomy and argues that she uses the informational aspects of copyrighted photographs to teach machine learning models how to identify objects, a court may find that she did not alter the work's original purpose of depicting such objects in the process of compiling her training data set.

Rather than relying on a fair use defense, it would be more ideal for enterprising AI developers if a clear safe harbor in copyright law allowed the open use of expressive works in training data.

III. AN ARGUMENT FOR A DATA MINING SAFE HARBOR

Fundamentally, data mining is not copyright infringement but rather involves the lawful copying of unprotected material forms. Practically, providing a safe harbor for data mining on top of the existing fair use doctrine would provide legal certainty to innovators who may currently be deterred by the threat of litigation and hefty statutory fines.

A. DATA MINING IS NOT COPYRIGHT INFRINGEMENT

Not all copying is copyright infringement.[87] Professor Abraham Drassinower uses *Baker v. Selden* to illustrate the distinction between a copyrighted work and its material form, arguing that not all uses of the work's material form are acts of infringement.[88] In the classic Supreme Court case, Selden obtained a copyright on his book explaining the operation of a novel accounting system, which included accounting forms as part of the explanation.[89] Baker then used a similar system, which included copying the forms, as explained and illustrated in Selden's books.[90] The forms were unique in that they both described the accounting system and were also used to perform the accounting.[91] While copyright law protects the use of the forms as an explanation, it does not protect the use of the forms as an invention (which falls within the realm of patents).[92] The defendant escaped copyright liability because he used the plaintiff's forms as aspects of a novel accounting system, but not as aspects of an explanation of the accounting system.[93]

86. *See* Brammer v. Violent Hues Prods., LLC, 922 F.3d 255, 267–69 (4th Cir. 2019).
87. Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 361 (1991).
88. *See* DRASSINOWER, *supra* note 10, at 88–100.
89. Baker v. Selden, 101 U.S. 99, 100 (1879).
90. *Id.* at 101.
91. DRASSINOWER, *supra* note 10, at 89.
92. *Id.*
93. *Id.* at 93.

Thus, copyright infringement requires not just copying of a work's material form, but also the unauthorized use of the work for its expressive purpose.[94] By the same logic, merely technical or non-communicative uses are not uses of a work for its expressive purpose and therefore are not copyright infringement.[95] Because they are not expressive uses of a work at all, technical and noncommunicative uses need not even be subject to a fair use analysis.[96]

Applying Drassinower's logic to data mining, this Note argues that downloading images from the internet for use in training data is not copyright infringement, but rather lawful copying of the works' material forms. An AI developer does not redistribute, or re-communicate, the copyrighted images to the public but instead uses them to train a machine learning model. Because copyright protection does not extend to the material forms of works themselves, the simple act of downloading images does not encroach upon a protected use of copyrighted works. By reading too closely into fixation requirements for intermediate copying in software cases and too eagerly applying the fair use analysis, courts have overlooked the simple explanation that some acts of reproduction are not copyright infringement because they are non-expressive reproductions of material forms only, analogous to Baker's copying of Selden's accounting forms.[97]

Thus, imposing copyright liability on developers who use expressive works to train their functional models would be an overreach of creators' rights. Copyright law is intended to protect creative and expressive works.[98] The idea-expression dichotomy limits copyright protection to the expressive elements of a work and not the functional ideas contained within.[99] In building a machine learning training set, the developer is not interested in reproducing the expressive works but is instead interested in the functional content contained within the material forms. For example, NLP developers may be interested in using literary works as training data to extract foundational patterns of human

94. *See id.* at 94.

95. *Id.* at 87–88, 100–102.

96. *Id.* at 101.

97. *See* Baker v. Selden, 101 U.S. 99, 107 (1879).

98. *See* 17 U.S.C. § 102.

99. *See* Mazer v. Stein, 347 U.S. 201, 217 (1954) ("[A] copyright gives no exclusive right to the art disclosed; protection is given only to the expression of the idea—not the idea itself."); Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 556 (1985) ("[C]opyright's idea/expression dichotomy 'strike[s] a definitional balance between the First Amendment and the Copyright Act by permitting free communication of facts while still protecting an author's expression.'" (quoting Harper & Row Publishers, Inc. v. Nation Enters., 723 F.2d 195, 203 (2d Cir. 1983), *rev'd*, 471 U.S. at 539)).

speech.[100] And in computer vision, developers use videos and photographs of busy city streets to teach machines how to identify pedestrians.[101] In neither case is the machine copying the expressive character of the works (i.e., the expression in the writing itself or the artful depiction of city streets in a photograph).

Additionally, allowing the open use of copyrighted material as training data would further the broader goals of intellectual property law. As Judge Leval noted in *Authors Guild*, "[t]he ultimate goal of copyright is to expand public knowledge and understanding . . . . [T]he ultimate, primary intended beneficiary [of copyright] is the public."[102] A safe harbor in copyright law for data mining would provide public benefit by stimulating innovation. New applications of artificial intelligence are constantly conjured up: computer vision technology can help drones spot wildfire hazards in California,[103] detect bias and trends on TV,[104] and catch illegal elephant poachers in Africa.[105] AI-driven tools like Grammarly and Adobe Sensei even assist writers and artists in the creation of new content.[106] These beneficial applications could be more numerous and of higher quality if potential innovators had open access to copyrighted works as training data.[107] Allowing copyright to hinder the growth of beneficial technology that does not infringe upon creators' rights would ultimately impede the goals of intellectual property law.

---

100. *See* Ana Cristina Mendes & Cláudia Antunes, *Pattern Mining with Natural Language Processing: An Exploratory Approach*, *in* MACHINE LEARNING AND DATA MINING IN PATTERN RECOGNITION 266, 266 (Petra Perner ed., 2009) ("Articles from a Portuguese newspaper are the input . . . . Results . . . provided several evidences about the structure of the language.").

101. *See* Antonio Brunetti, Domenico Buongiorno, Gianpaolo Francesco Trotta & Vitoantonio Bevilacqua, *Computer Vision and Deep Learning Techniques for Pedestrian Detection and Tracking: A Survey*, 300 NEUROCOMPUTING 17, 20–21 (2018).

102. 804 F.3d 202, 212 (2d Cir. 2015).

103. John McCormick, *California Utilities Hope Drones, AI Will Lower Risk of Future Wildfires*, WALL ST. J. (Sept. 11, 2020, 5:30 AM), https://www.wsj.com/articles/california-utilities-hope-drones-ai-will-lower-risk-of-future-wildfires-11599816601?mod=hp_minor_pos4.

104. Sherin Shibu, *Who are the Biggest Attention Hogs on Cable News? Ask this new AI Tool*, PCMAG (Sept. 1, 2020), https://www.pcmag.com/news/who-are-the-biggest-attention-hogs-on-cable-news-ask-this-new-ai-tool.

105. Simorin Pinto, *AI-powered Camera to Stop Illegal Poaching*, INNOVATION ENTER. CHANNELS, https://channels.theinnovationenterprise.com/articles/ai-powered-camera-to-stop-illegal-poaching [https://perma.cc/93NU-43LJ].

106. *See, e.g.*, Grammarly, *How We Use AI to Enhance Your Writing | Grammarly Spotlight*, GRAMMARLY BLOG (May 17, 2019), https://www.grammarly.com/blog/how-grammarly-uses-ai; Marc DeAngelis, *Adobe Premiere Pro Can Automatically Reframe Your Videos*, ENGADGET (Sept. 13, 2019), https://www.engadget.com/2019-09-13-adobe-premiere-pro-reframes-video.html.

107. *See infra* Section III.B.3.

B. A SAFE HARBOR TO FIX LEGAL UNCERTAINTY

Because data mining falls outside the scope of copyright, a safe harbor for such uses should be enacted. Clarifying the law would promote innovation by providing legal certainty for smaller innovators.

1. *The fair use doctrine should not be applied to data mining because it would result in further stretching of the unpredictable doctrine.*

Most copyright stakeholders argue that fair use is currently sufficient because the "flexible doctrine . . . is capable of adapting to the use of copyrighted works in an AI context."[108] In practice, though, fair use has been referred to as "the most troublesome [doctrine] in the whole law of copyright" and described as a "billowing white goo."[109] Because data mining is not copyright infringement, the inquiry should end before the courthouse doors, rendering the nebulous fair use analysis moot. A safe harbor would not displace the fair use defense in the United States for all AI application but would provide an additional layer of certainty on top of the doctrine of functional uses. Certainty in the law is desirable because even the possibility of litigation may be a deterrent for smaller creators.[110]

Conceptually, the fair use doctrine should not even apply to data mining because downloading data from the internet for use in training data is not a use of protected works.[111] Drassinower differentiates this nonuse from fair use based on whether the use is communicative.[112] The fair use doctrine protects communicative uses of copyrighted works and appeals to the "equality as authors"—the defendant reproduces the plaintiff's work as a "reasonably necessary aspect of the defendant's own authorial engagement."[113] Fair use is rightly invoked for communicative uses like parody, art, and criticism.[114] Conversely,

---

108. *See* U.S. PAT. & TRADEMARK OFF., *supra* note 6, at 26; Stan Adams, *Comments On the USPTO's Intellectual Property Protection for Artificial Intelligence Innovation*, CTR. FOR DEMOCRACY & TECH. (Jan. 16, 2020), https://cdt.org/insights/comments-on-the-usptos-intellectual-property-protection-for-artificial-intelligence-innovation ("[F]air use offers advantages over other possible legal mechanisms for allowing the use of copyrighted works in the context of machine learning . . . . [T]he resulting [EU] exception for text and data mining (TDM) is so rigid and restrictive as to prevent many beneficial uses of datasets.").

109. VHT, Inc. v. Zillow Grp., Inc., 918 F.3d 723, 739 (9th Cir. 2019) (quoting Monge v. Maya Mags., Inc., 688 F.3d 1164, 1170–71 (9th Cir. 2012)).

110. *See infra* Section III.B.2.

111. *See supra* Section III.A.

112. *See* DRASSINOWER, *supra* note 10, at 100–01.

113. *Id.* at 108.

114. *See, e.g.*, Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 572 (1994); Blanch v. Koons, 467 F.3d 244, 246 (2d Cir. 2006); Salinger v. Random House, Inc., 811 F.2d 90, 92 (2d Cir. 1987).

nonuse stems out of the distinction between a work and its material form; the defendant is not liable because she only makes use of the material form of the work and not the work as a communicative act.[115] Nonuses include search engine thumbnail reproductions, temporary copies made during internet browsing, and digital copies of student papers made to detect plagiarism.[116] Courts confusing nonuses for fair uses have contributed to the confusion in the fair use doctrine.

Application of the fair use defense to digital technologies has also stretched the doctrine to the detriment of creators in general. As Professor Jiarui Liu has observed, lower courts stretch transformative use to new fact patterns, contributing to a "slippery slope progression."[117] He argues that the trend in intermediate copying decisions perfectly illustrates the phenomenon: logic applied to copies made and deleted during the reverse engineering of video games was stretched to justify search engines that continuously store and display verbatim copies.[118] Instead of relying on an overstretching of the fair use doctrine, courts should have permitted search engine uses of copyrighted works as nonuses of the protected works.[119]

Proponents of justifying data mining with fair use must also keep in mind that the fair use doctrine applies across the board to decisions involving literature, art, and music. Fair use decisions have wide-reaching implications for other creative mediums. For example, the Supreme Court suggested that a musical parody was unlikely to act as a substitute for an original song because the two works serve different market functions.[120] Expanding on that logic, the Second Circuit in *Cariou v. Prince* pointed to the differences in wealth in consumers to justify a finding of fair use.[121] In *Cariou*, the two artists served two different markets of art collectors: while the plaintiff earned just thousands of dollars in royalties and sold only to personal acquaintances, the defendant sold millions of dollars' worth of art to celebrities.[122] Ultimately, the reasoning used to justify digital technologies under the fair use doctrine could narrow the rights of authors in other areas.

115. DRASSINOWER, *supra* note 10, at 108–09.
116. *See id.* at 87, 102–04.
117. Liu, *supra* note 69, at 211.
118. *Id.*
119. DRASSINOWER, *supra* note 10, at 100–03.
120. Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 572 (1994).
121. *See* Cariou v. Prince, 714 F.3d 694, 709 (2d Cir. 2013); Liu, *supra* note 69, at 199–200.
122. *Cariou*, 714 F.3d at 709.

Liu also argues that the transformative analysis under the fair use doctrine has become a way for courts to decide new policy concerns under the "impression of *stare decisis*."[123] Given the malleability of the doctrine and the wide scope of case law, courts usually have little difficulty in finding a precedent that justifies fair use decisions.[124] This constant expansion of transformative use creates further uncertainty in copyright law. Rather than rely on district courts to dictate important policy issues involving emerging technologies, like the training of machine learning models with copyrighted data, Congress should be more proactive in enacting safe harbors for new technologies.

Practically, the fair use doctrine is an affirmative defense that must be raised during litigation.[125] And, because fair use is an affirmative defense, a claim of copyright infringement through data mining could survive a motion to dismiss because intermediate copying is infringement as a matter of law.[126] The fair use case law provides conflicting guidance for innovators to try to understand how their technology will be viewed by the courts.[127] The unpredictability of the fair use doctrine is especially apparent with technology cases; the acclaimed copyright scholar William F. Patry once predicted that Google Books was unlikely to stand the fair use test in a now-deleted blog post (he has since joined Google's legal team).[128] If great legal minds cannot make sense of the fair use doctrine, how can innovators be expected to rely on the defense when developing new technologies? There are already examples of this uncertainty among smaller innovators as discussed in the next Section.[129]

2. Uncertainty in copyright law disproportionately handicaps new innovators.

A safe harbor in copyright law for data mining is further warranted because current uncertainty in the law disproportionately handicaps small innovators. At the "Copyright in the Age of Artificial Intelligence" conference hosted by

---

123. Liu, *supra* note 69, at 172.

124. *Id.*

125. *See Campbell*, 510 U.S. at 599.

126. *See* Tiffany Design, Inc. v. Reno-Tahoe Specialty, Inc., 55 F. Supp. 2d 1113, 1121 (D. Nev. 1999) (holding that intermediate copying is copyright infringement as a matter of law).

127. *See supra* Section II.C.

128. Liu, *supra* note 69, at 169.

129. *See, e.g.*, Peter Ned, Comment to *In the US, is it Illegal to Train Neural Networks Using Copyrighted Images?*, QUORA (July 1, 2017), https://www.quora.com/In-the-US-is-it-illegal-to-train-neural-networks-using-copyrighted-images; farsass, Comment to *Is it Legal to Use Copyright Material as Training Data?*, REDDIT (July 1, 2016, 8:33 AM), https://www.reddit.com/r/MachineLearning/comments/4qrgh8/is_it_legal_to_use_copyright_material_as_training; bluboxsw, Comment to *[D] Are There Any Legal Issues with Training Machine Learning Models on Copyrighted Content?*, REDDIT (July 12, 2019, 10:50 AM), https://www.reddit.com/r/MachineLearning/comments/cc76us/d_are_there_any_legal_issues_with_training.

the U.S. Copyright Office and the World Intellectual Property Organization in 2020, Vanessa Bailey, the global director of IP policy for Intel Corporation, stated, "[C]opyright laws are still adequate . . . . [W]e're doing fine with what we have and . . . smart lawyers are figuring things out."[130] This statement illustrates the wealth of resources that large tech companies enjoy: access to enormous datasets and expensive legal teams. The argument would be different from the perspective of a smaller innovator because copyright law can create disparities in access to data, ability to litigate, and deterrent effects against using larger datasets.

Copyright law contributes to the disproportionate access to valuable training data between dominant tech companies and smaller innovators. Large tech platforms like YouTube and Meta operate with terms of service that provide them with access to copyright-protected text and data uploaded to their servers.[131] This data presents a treasure trove for training machine learning models. For example, Meta uses its data from over 2 billion users to calibrate newsfeeds, generate text for the visually impaired, and power its facial recognition technology.[132] Even when large companies do not have built-in systems for acquiring data, they can buy large datasets.[133] These acquisitions of large datasets can be extremely expensive; one such partnership IBM undertook for access to oncology data cost $50 million.[134] And though some large tech companies have released open source algorithms, it is rarer for them to release the underlying datasets.[135]

130. U.S. COPYRIGHT OFF., COPYRIGHT IN THE AGE OF ARTIFICIAL INTELLIGENCE 3, 294 (2020) (transcript of symposium).

131. *See Terms of Service*, YOUTUBE (Nov. 18, 2020), https://www.youtube.com/static?template=terms [https://perma.cc/3NTH-KB9V] ("By providing Content to the Service, you grant to YouTube a worldwide, non-exclusive, royalty-free, sublicensable and transferable license to use that Content (including to reproduce, distribute, prepare derivative works, display and perform it) . . . ."); *Terms of Service*, FACEBOOK (Oct. 22, 2020), https://www.facebook.com/terms.php [https://perma.cc/Z7MS-4D3R]("Specifically, when you share, post, or upload content that is covered by intellectual property rights on or in connection with our Products, you grant us a non-exclusive, transferable, sub-licensable, royalty-free, and worldwide license to host, use, distribute, modify, run, copy, publicly perform or display, translate, and create derivative works of your content . . . .").

132. Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579, 606–07 (2018).

133. *Id.* at 606.

134. *Id.* at 608.

135. *Id.* at 599.

Thus, smaller innovators do not have equivalent access to datasets, which are an important resource for developing good algorithms.[136] Although open source datasets do exist, they may not be comparable in size or may be prone to bias.[137] However, if a safe harbor in copyright law clarified that mining copyrighted content for machine learning training purposes was legal, these smaller innovators would enjoy freer access to valuable data, allowing them to develop innovative products that could allow them to better compete against tech giants. Increased competition would benefit the public by providing access to more numerous and higher quality AI products.

Notwithstanding the data that large tech companies legally have access to, it is "plausible, if not probable, that dominant AI players create unauthorized copies of protectable works to use as training data for AI systems."[138] Even if large tech companies run afoul of copyright law in training their machine learning algorithms, they are better equipped with lawyers to defend themselves from liability. As illustrated previously, the doctrine of fair use is highly fact-specific and arguably unpredictable, which creates room for smart (i.e., creative, expensive) lawyering.[139] The tech giant Google secured a victory in the landmark fair use case *Authors Guild v. Google, Inc.* after ten years of litigation.[140] Conversely, litigation costs can drain the bank accounts of startups and lead them to cease operations.[141] The support of affluent institutions in copyright cases like *Cariou v. Prince* further hints at the influence of wealth in securing favorable verdicts.[142]

---

136. *See* Steven Levy, *Inside Facebook's AI Machine*, WIRED (Feb. 23, 2017, 12:00 AM), https://www.wired.com/2017/02/inside-facebooks-ai-machine/ ("One huge factor in building machine learning systems is getting quality data—the more the better.").

137. *See id.* ("When you have over a billion people interacting with your product every day, you collect a *lot* of data for your training sets . . . ."); *infra* Section III.B.3.

138. Levendowski, *supra* note 132, at 599.

139. *See supra* Part II, Section III.B.1.

140. *See* Adam Liptak & Alexandra Alter, *Challenge to Google Books is Declined by Supreme Court*, N.Y. TIMES (Apr. 18, 2016), https://www.nytimes.com/2016/04/19/technology/google-books-case.html.

141. *See* Moran, *supra* note 7 ("Litigation is expensive—no matter how speculative the claims against you nor how worthy your position . . . . With our company ensnared by this legal battle, we have been unable to raise another round of funding to fuel our development and marketing efforts. Our bank account is running out, and we must cease operations in the new year.").

142. *See* Douglas W. Kenyon & Stephen P. Demm, *Supreme Court Denies Cert in Cariou Fair Use Case: What Next?*, INTELL. PROP. MAG., Feb. 2014, at 71, 72 ("And in an unusual step at the district court level, the Andy Warhol Foundation and 30 other arts groups filed an *amicus brief* . . . ."); Hrag Vartanian, *Cariou v. Prince Isn't Over: Orgs Rep'ing 45,000 Creatives File Brief in Support of Cariou*, HYPERALLERGIC (Dec. 20, 2013), https://hyperallergic.com/99668/cariou-v-prince-isnt-over-orgs-reping-45000-creatives-file-brief-in-support-of-cariou ("Amici submit this brief primarily in opposition to the amicus brief filed by the Andy Warhol Foundation

Coupled with the expense of litigation itself, the U.S. remedial regime of copyright law also presents a significant deterrent for new innovators. Statutory damages range from $200 to $150,000 per work, depending on the willfulness of infringement.[143] Training a model from scratch can require millions of data points, and even fine-tuning a machine learning model can require thousands of data points.[144] Because of the large quantity of data needed to develop a machine learning model, a developer could potentially be on the hook for massive fines. This U.S. remedial regime of statutory damages creates the potential for crushing liability, especially when amplified by the massive amounts of data required to train a machine learning model.[145] The legal uncertainty and mere threat of exorbitant fines surrounding the use of copyrighted data would be enough to deter smaller actors from creating and using potentially valuable datasets.

The effect of this legal uncertainty is readily observed in practice. A look at user forums such as Quora and Reddit illustrates caution against using copyrighted data for training machine learning models:

> "Sorry, I can't answer the question. And I doubt anyone can. It probably requires a series of multi-million lawsuits involving herds of lawyers to find out. To be on the safe side don't use copyrighted images, at least for now."—Peter Ned[146]

> "You can find a lawyer to argue for or against this being a case of copyright violation depending on how much you pay him."—farsass[147]

> "It is still derivative work, which is covered under law. When the $$$ you make exceeds the $$$ to sue you, expect to get sued."—bluboxsw[148]

---

and other elite foundations and museums who do not represent the views of working artists." (emphasis omitted)).

143. 17 U.S.C. § 504.

144. Brown et al., *supra* note 9.

145. *See* Menell, *supra* note 40, at 268.

146. Peter Ned, Comment to *In the US, is it Illegal to Train Neural Networks Using Copyrighted Images?*, QUORA (July 1, 2017), https://www.quora.com/In-the-US-is-it-illegal-to-train-neural-networks-using-copyrighted-images.

147. farsass, Comment to *Is it Legal to Use Copyright Material as Training Data?*, REDDIT (July 1, 2016, 8:33 AM), https://www.reddit.com/r/MachineLearning/comments/4qrgh8/is_it_legal_to_use_copyright_material_as_training.

148. bluboxsw, Comment to *[D] Are There Any Legal Issues with Training Machine Learning Models on Copyrighted Content?*, REDDIT (July 12, 2019, 10:50 AM), https://www.reddit.com/r/MachineLearning/comments/cc76us/d_are_there_any_legal_issues_with_training.

A developer seeking to gather a wide expanse of training data could be deterred by such general uncertainty.

Relatedly, questions of access influence how scholars conduct their research. NLP techniques can be used to study literary works.[149] However, researchers may limit research questions based on the availability of texts. Ethan Reed, a digital humanities researcher interested in articulations associated with systemic injustice, laments that copyright plays "an enormous role in determining the initial paths in my scholarly decision-making process."[150] In one example, he limited the scope of an NLP research project to just three books of poetry from 1969 because of copyright concerns.[151] In explaining this decision, he highlights an additional problem of reproducibility of research.[152] Even if scholars can share results of NLP analyses through transformative, non-consumptive use, they cannot share the copyrighted corpora from which the insights came from.[153] Copyright limits the scope of computational humanities research and potentially stymies socially valuable insights that can be derived from contemporary works.

AI practitioner Arjan Wijnveen also describes another copyright problem that developers face—the decay in public datasets used to train models.[154] Public datasets like ImageNet provide huge volumes of annotated data for developers to train on but are based on public image or video hosting sites.[155] ImageNet does not make images publicly available in their original resolutions because they might be subject to copyright, instead providing thumbnails and URLs.[156] When an image or video that is part of an annotated dataset is taken down, developers are "out of luck."[157] An obvious solution would be to store

149. *See, e.g.*, Senja Pollak, Matej Martinc & Katja Mihurko Poniz, *Natural Language Processing for Literary Text Analysis: Word-Embeddings-Based Analysis of Zofka Kveder's Work*, 2607 CEUR WORKSHOP PROC. 33 (2020) (using NLP to gain insights in feminine literary history).

150. Ethan Reed, *First Steps with NLP and a Collection of Amiri Baraka's Poetry*, SCHOLAR'S LAB (Nov. 30, 2017), https://scholarslab.lib.virginia.edu/blog/first-steps-with-nlp-and-a-collection-of-amiri-barakas-poetry ("Though conceptually unglamorous, basic questions of access have played an enormous role in determining the initial paths in my scholarly decision-making process.").

151. *Id.*

152. *Id.*

153. *Id.*

154. Arjan Wijnveen, *How Copyright is Causing a Decay in Public Imagesets*, MEDIUM (Nov. 28, 2016), https://medium.com/@arjanwijnveen/how-copyright-is-causing-a-decay-in-public-datasets-f760c5510418.

155. *Id.*

156. *Download FAQ*, IMAGENET, http://image-net.org/download-faq ("The images in their original resolutions may be subject to copyright, so we do not make them publicly available on our server.").

157. Wijnveen, *supra* note 154.

a cached copy of the images, but Wijnveen believes that this fix is prohibited by copyright law.[158] Thus, vague copyright laws can undermine the usefulness of public datasets.

3. Licensing is an overreach of authors' rights and could propagate bias.

Some commentators argue in favor of licenses for AI training data used in commercial applications.[159] However, requiring the licensing of training data would be burdensome to developers and an overreach of authors' rights. Data for machine learning training sets is usually scraped from the internet, and this process is automated because of the large quantities of data needed.[160] Given the massive size of datasets and the automation of the scraping process, it would be extremely burdensome for developers to go through scraped data, determine what is copyrighted, and request permission from each creator.

Most importantly, because reproductions in data mining are fundamentally not infringement, requiring compensation for these merely technical and non-communicative uses would be an overreach of authors' rights. It does not matter if a market for training data exists or could exist.[161] In *Baker v. Selden*, the fact that a market for accounting forms existed did not change the finding that Baker's use of Selden's accounting forms was an unactionable use under copyright law.[162] Allowing an author to charge for all reproductions of the material form of her work would overextend her claim to rights over her work as a material thing, while copyright applies only to works as communicative acts.[163]

Some may also argue that developers have already adjusted to copyright uncertainty by using licensed or safe datasets. For example, developers can circumvent liability by training on datasets made available under creative commons (CC) licenses[164] or by training on corpora of text from websites, like Wikipedia.org, that allow free access, copying, and distribution.[165] However,

158. *Id.*

159. *See* U.S. Pat. & Trademark Off., *Artificial Intelligence: The Ins and Outs of Copyright and AI*, VBRICK REV, at 43:00 (Jan. 31, 2019), https://rev-vbrick.uspto.gov/#/videos/d6e591c3-64cf-4d74-ab35-9f387a2da4b2 (highlighting the views of Mary Ransenberger, Executive Director of the Authors Guild, who proposes solution of a collective licensing system).

160. *See* Casper Hansen, *Web Scraping For Machine Learning - With SQL Database*, MACH. LEARNING FROM SCRATCH (Dec. 4, 2019), https://mlfromscratch.com/web-scraping-machine-learning/#.

161. *See* DRASSINOWER, *supra* note 10, at 102.

162. *Id.*

163. *Id.* at 99.

164. *See*, *e.g.*, *Updated Dataset*, YOUTUBE-8M, https://research.google.com/youtube8m/download.html ("The dataset is made available by Google LLC. under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.").

165. *See*, *e.g.*, *Wikipedia:Copyrights*, WIKIPEDIA, https://en.wikipedia.org/wiki/Wikipedia:Copyrights ("Wikipedia content can be copied, modified, and redistributed *if and only if* the

Professor Amanda Levendowski points out an important problem that arises from the use of safe datasets—the propagation of AI bias.[166] She argues that copyright law encourages AI creators to use "easily available, legally low-risk works," such as public domain and creative commons-licensed works, for training data.[167] The use of this safe data is problematic because the sources are often biased.[168] For example, most public domain literary works that could be used for NLP applications "were published prior to 1923, back when the 'literary canon' was wealthier, whiter, and more Western than it is today," and thus an AI model trained on these works would reflect the biases of the time.[169] Additionally, training data from CC-licensed websites like Wikipedia can reflect the biases inherent in editors; for example, only 10–20% of Wikipedia editors are women.[170] This gender disparity contributes to male-centric representations of facts, which can be propagated in biased models.[171]

In general, larger and more diverse datasets create better models.[172] And, by clearly broadening the scope of data that can be mined to include copyrighted works, reliance on safe, but biased, data is weakened. Models trained on diverse, but potentially copyrighted, data will be more accurate and representative of modern norms.

IV. A PROPOSED SAFE HARBOR FRAMEWORK

While many legal scholars and practitioners recognize the importance of using copyrighted data for data mining, most argue that fair use is enough to

---

copied version is made available on the same terms to others and acknowledgment of the authors of the Wikipedia article used is included . . . .").

166. *See* Levendowski, *supra* note 132, at 610.

167. *Id.*

168. *Id.*

169. *Id.* at 615.

170. *Id.* at 618; *Wikipedia:Who Writes Wikipedia?*, WIKIPEDIA, https://en.wikipedia.org/wiki/Wikipedia:Who_writes_Wikipedia%3F.

171. *See* Levendowski, *supra* note 132, at 619 ("The English Wikipedia article about New England Patriots tight-end Rob Gronkowski is nearly 4,000 words long and boasts 66 citations. By comparison, Stanleyetta Titus, a revolutionary suffragette and the first woman admitted to the New York state bar, does not even have an article." (footnote omitted)).

172. *See* Enric Junqué de Fortuny, David Martens & Foster Provost, *Predictive Modeling With Big Data: Is Bigger Really Better?*, 1 BIG DATA 215, 215 (2013) ("This study provides a clear illustration that larger data indeed can be more valuable assets for predictive analytics."); John R. Smith, *IBM Research Releases 'Diversity in Faces' Dataset to Advance Study of Fairness in Facial Recognition Systems*, IBM RSCH. BLOG (Feb. 15, 2019), https://www.ibm.com/blogs/research/2019/01/diversity-in-faces ("The AI systems learn what they're taught, and if they are not taught with robust and diverse datasets, accuracy and fairness could be at risk.").

justify this use in the United States.[173] However, as argued in the previous Part, fair use has its limitations, and uncertainty is still apparent among AI creators.[174] A safe harbor for data mining in U.S. copyright law would provide a layer of legal certainty on top of the broad fair use doctrine. This final Part will outline a proposed framework for a safe harbor for data mining in U.S. copyright law.

A. OTHER TEXT AND DATA MINING (TDM) EXCEPTIONS

As the benefits of AI technologies have become apparent, countries have amended their copyright laws to promote innovation and remain competitive in the AI/ML space.[175] The United States already has comparably strong protections for data miners through the fair use doctrine.[176] However, a general safe harbor for data mining would further the spirit of these protections by clarifying the interaction of copyright and AI for smaller innovators.[177] Other jurisdictions have recently enacted text and data mining (TDM) exceptions that can provide guidance here.

Japan was the first country in the world to update its copyright laws to include an exception for text and data mining.[178] Article 47(7) was introduced in 2009 and authorized TDM by all users for all purposes, whether commercial or non-commercial.[179] In line with Prime Minister Shinzo Abe's objective of promoting AI and Big Data industries, the 2018 Amendment to the Copyright Act later introduced three provisions to clarify the law and remove perceived copyright barriers to AI: (1) Article 30-4 authorizes users to "analyse and understand copyrighted works for machine learning;" (2) Article 47-4 "permits electronic incidental copies of works;" and (3) Article 47-5 "allows the use of copyrighted works for data verification."[180] Japan's three new copyright provisions that specifically reference acts relevant to the machine learning process demonstrate a national commitment to the flourishing of AI industries.

---

173. *See, e.g.*, Carroll, *supra* note 78, at 936; Levendowski, *supra* note 132, at 619; U.S. PAT. & TRADEMARK OFF., *supra* note 6, at 26.

174. *See supra* Section III.B.

175. EUR. ALL. FOR RSCH. EXCELLENCE, THE GLOBAL AI RACE 2 (2018), http://eare.eu/assets/uploads/2018/06/Global-AI-Race.pdf.

176. *See* U.S. PAT. & TRADEMARK OFF., *supra* note 6, at iv ("Most commenters found that existing fair use law does not require modification, as fair use is a flexible doctrine and is capable of adapting to the use of copyrighted works in the context of AI.").

177. *See* Menell, *supra* note 40, at 346–48 (discussing fair use discussion).

178. *See Japan Amends its Copyright Legislation to Meet Future Demands in AI and Big Data*, EUR. ALL. FOR RSCH. EXCELLENCE (Sept. 3, 2018), http://eare.eu/japan-amends-tdm-exception-copyright.

179. *Id.*

180. *Id.*

Similarly, the United Kingdom has also enacted an exception for TDM purposes in its copyright laws.[181] However, its exception is markedly narrower than Japan's. Under § 29A of the Copyright, Designs and Patents Act 1988, copies made for TDM analysis do not infringe copyright, provided that the work is carried out "for the sole purpose of research for a non-commercial purpose."[182] The beneficiaries of the exception are also limited to those who have "lawful access" to the work in question.[183] While providing legal certainty for researchers, the U.K. exception leaves out startups and entrepreneurs who aim to commercialize innovative machine learning technologies.

Recognizing the legal uncertainty experienced by researchers who utilize TDM and a need for a harmonized exception among its member states, the European Union recently introduced two exceptions for TDM in Directive 2019/790/EU.[184] Article 3 permits "reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access."[185] Article 4 confers an exception for "reproductions and extractions of lawfully accessible works and other subject matter" for commercial TDM uses but allows rightsholders to opt-out of the exemption.[186] This limitation imposed on commercial TDM uses in Article 4 has been criticized for "effectively creat[ing] and legitimiz[ing] a derivative market for text and data mining, which right holders may wish to control, license[,] or even entirely prohibit."[187] While the TDM exceptions in the Directive recognize the importance of data mining in research and technology, the clear limitations in Article 4 may put commercial AI developers in the European Union at a disadvantage.

Not to be left behind, other countries are also considering changes to their copyright law to provide legal certainty and encourage innovation and research in data mining.[188] Australia is considering several TDM-related exceptions as part of an overall fair use exception.[189] Canadian scholars have advocated for a

181. Copyright, Designs and Patents Act 1988, c. 48, § 29A (UK).

182. *Id.*

183. *Id.*

184. Council Directive 2019/790, arts. 3–4, 2019 O.J. (L 130) 92, 113–14 (EU).

185. *Id.*

186. *Id.*

187. *See* Bernt Hugenholtz, *The New Copyright Directive: Text and Data Mining (Articles 3 and 4)*, KLUWER COPYRIGHT BLOG (July 24, 2019), http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4.

188. EUR. ALL. FOR RSCH. EXCELLENCE, *supra* note 175, at 2.

189. *Id.*

fair use regime or a specific TDM exception for commercial uses.[190] And Singapore is expected to move forward with proposed TDM exceptions enabling broad uses for both commercial and non-commercial contexts.[191] As countries amend and clarify their laws to allow technological flourishing, the United States should also consider revisiting its copyright laws.

B. A U.S. DATA MINING SAFE HARBOR

Lessons can be learned from the implementation of TDM exceptions in other jurisdictions to craft a safe harbor for data mining in the United States. Such a safe harbor should (1) define "data mining" broadly; (2) clearly allow reproductions, derivations, and dataset-sharing for data mining purposes; (3) allow non-commercial and commercial uses to the same extent; and (4) be limited to functional, non-expressive uses of data mining.

A safe harbor should broadly define "data mining" to cover a wide range of pattern extraction techniques, not limited to AI/ML. One potential definition is proposed by Jean-Paul Triaille: "The automated processing of digital materials, which may include texts, data, sounds, images or other elements, or a combination of these, in order to uncover new knowledge or insights" (although he uses the term "data analysis" instead of "data mining").[192] While other jurisdictions refer to "text and data mining," "text" is redundant, as it is encapsulated in "data."[193] In crafting the safe harbor, what matters is not the exact technique through which patterns are extracted, but the automated process of gaining functional insights from expressive data. While this Note specifically considers data mining in a machine learning and artificial intelligence context, using a broad definition of data mining in a copyright safe harbor would not foreclose future technologies that employ other techniques for pattern extraction.

A safe harbor should also be specific enough in the rights that are conferred to data miners to provide legal clarity. Given the intimidating U.S. copyright remedial regime and the unpredictable fair use doctrine, smaller innovators in the United States would benefit from an exception that clearly spells out allowed uses.[194] Here, lessons can be drawn from Japan's recent amendments to clarify its copyright exception, which provide an innovation-friendly

190. *Id.*

191. *Id.*

192. JEAN-PAUL TRIAILLE, JÉRÔME DE MEEÛS D'ARGENTEUIL & AMÉLIE DE FRANCQUEN, STUDY ON THE LEGAL FRAMEWORK OF TEXT AND DATA MINING (TDM) 17 (2014), https://www.fosteropenscience.eu/sites/default/files/pdf/3476.pdf.

193. *See id.* at 8–9.

194. *See supra* Section III.B.

and clear implementation of an exception.[195] It should be clearly enumerated that reproductions and derivations (for labeling and annotation) of works are permitted in a data mining context. Further, the secure storage, retention, and sharing of datasets should also be permitted for verification purposes. Allowing researchers and regulators the ability to scrutinize how data is being used would promote fairer, safer, and higher-quality AI products.[196]

Policymakers should also be careful to not impose limitations on commercial uses, such as opt-out mechanisms, that could unduly stifle innovation contrary to the expressed policy of copyright law in the United States.[197] Unlike the exceptions of the United Kingdom and European Union, commercial uses should be permitted to the same extent that research and non-profit uses are allowed. The U.S. fair use doctrine already includes a copyright exception for research uses and considers nonprofit uses favorably.[198] Thus, the proposed safe harbor would need to specifically free up commercial uses of data mining. Rightsholders should not be allowed to opt-out of uses of their works for commercial data mining purposes because doing so would legitimize a licensing market for machine learning training data that could prompt further legal confusion, put smaller innovators at a competitive disadvantage, and lead to lower quality, biased AI systems.[199]

Finally, and most importantly, the safe harbor should be limited to functional, non-expressive uses of data mining. Some emerging AI technologies use expressive data to generate expressive works and are not covered within the scope of this Note.[200] Because expressive AI-generated works could compete with and effectively replace original works, it would be more difficult to justify a categorical safe harbor for these works without unduly encumbering

---

195. EUR. ALL. FOR RSCH. EXCELLENCE, *supra* note 178.

196. *See* Levendowski, *supra* note 132, at 605.

197. *See* Christophe Geiger, Giancarlo Frosio & Oleksandr Bulayenko, *Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU* 36 (Ctr. for Int'l Intell. Prop. Stud., Research Paper No. 2019-08, 2019); Authors Guild v. Google, Inc., 804 F.3d 202, 212 (2d Cir. 2015) ("The ultimate goal of copyright is to expand public knowledge and understanding . . . . [T]he ultimate, primary intended beneficiary [of copyright] is the public . . . .").

198. 17 U.S.C. § 107.

199. *See* Hugenholtz, *supra* note 187; Levendowski, *supra* note 132, at 579.

200. *See, e.g.*, Chris Baraniuk, *Computer Paints 'New Rembrandt' After Old Works Analysis*, BBC NEWS (Apr. 6, 2016), https://www.bbc.com/news/technology-35977315; Tom May, *AI 'Artist' Creates New Work Inspired by Jean-Michel Basquiat*, CREATIVE BOOM (Sept. 14, 2020), https://www.creativeboom.com/inspiration/ai-artist-resurrects-the-iconic-jean-michel-basquiat-to-mark-32-years-since-his-death-; Steven Poole, *The Rise of Robot Authors: Is the Writing on the Wall for Human Novelists?*, THE GUARDIAN (Mar. 25, 2019, 4:00 AM), https://www.theguardian.com/books/2019/mar/25/the-rise-of-robot-authors-is-the-writing-on-the-wall-for-human-novelists.

authors and creators. Instead, the fair use doctrine should be applied on a case-by-case basis for expressive uses of data mining.

In sum, a data mining safe harbor should (1) define "data mining" broadly; (2) clearly allow reproductions, derivations, and dataset-sharing for data mining purposes; (3) allow non-commercial and commercial uses to the same extent; and (4) be limited to functional, non-expressive uses of data mining.

V. CONCLUSION

As the United States strives to maintain its dominance in artificial intelligence,[201] it must consider how existing laws enable or stifle technological progress. Copyright law presents a potential barrier for AI growth when machine learning models are trained using expressive data.

Fundamentally, data mining is not copyright infringement. While the fair use doctrine provides a degree of flexibility in U.S. copyright law, certainty in the law is desirable for smaller actors. Practically, the fair use doctrine is unpredictable and has been stretched beyond its limits in new technological contexts.

Establishing a clear right to use copyrighted materials to train functional machine learning models is consistent with the goals of copyright law and would ultimately remove barriers to innovation. Rather than rely on courts to rule on important issues of technology policy, Congress should be more proactive in enacting a data mining safe harbor. A safe harbor would make legislatively clear what should already be clear given the scope of copyright law: data mining is not infringement at all.

201. *See* Exec. Order No. 13,859, 84 Fed. Reg. 3967 (Feb. 14, 2019).

GEORGIA V. PUBLIC.RESOURCE.ORG—A MISSED OPPORTUNITY FOR DEMOCRACY

Jeffrey Jacobsen†

I. INTRODUCTION

On April 27, 2020, the U.S. Supreme Court decided *Georgia v. Public.Resource.Org, Inc.* (*Georgia v. PRO*).[1] It held that the state of Georgia was barred from holding a copyright in annotations made to and published alongside its official code because of the government edicts doctrine—the idea that governments cannot copyright certain works.[2]

Copyright "subsists . . . in original works of authorship fixed in any tangible medium . . . from which they can be perceived."[3] This intellectual property right gives the holder the general power to exclude others from copying their work (among other things).[4] Practically speaking, this endows the copyright holder with the ability to set a price. But the power to set a price comes at a cost. When a price is set, there will be a subset of people who wish to access the work but cannot.[5] This is known as deadweight loss.[6]

Not everyone may hold a copyright. The federal government is statutorily barred from doing so.[7] Other government bodies are barred from holding copyright in certain works through a judicially created doctrine—the government edicts doctrine.[8]

Before *Georgia v. PRO*, the exact scope of the government edicts doctrine was unclear. No Supreme Court case had touched the subject in over 130 years, and even then, there were only three major cases.[9] These few cases presented two main ideas. First, publications with the force of law were not eligible for

---

DOI: https://doi.org/10.15779/Z380R9M50K

† J.D., University of California, Berkeley, School of Law 2021

1. *See* Georgia v. Public.Resource.Org, Inc., 140 S. Ct. 1498, 1498 (2020).
2. *See id.* at 1498–1504.
3. 17 U.S.C. § 102.
4. 17 U.S.C. § 106.
5. Oren Bracha & Talha Syed, *Beyond the Incentive-Access Paradigm? Product Differentiation & Copyright Revisited*, 92 TEX. L. REV. 1841, 1843–44 (2014).
6. *Id.*
7. 17 U.S.C. § 105.
8. *See* Georgia v. Public.Resource.Org, Inc., 140 S. Ct. 1498, 1504 (2020).
9. *See id.* at 1498; Callaghan v. Myers, 128 U.S. 617, 619 (1888); Banks v. Manchester, 128 U.S. 244, 244 (1888); Wheaton v. Peters, 33 U.S. 591, 591 (1834).

copyright.[10] Second, legal works adopted by or published under the authority of the government cannot be copyrighted.[11] The Supreme Court clarified the scope of the doctrine in *Georgia v. PRO*, opting for a rule moderately aligned with the second idea: "copyright does not vest in works that are (1) created by judges and legislators (2) in the course of their judicial and legislative duties."[12]

In clarifying the scope of the government edicts doctrine, the Supreme Court faltered twice. First, the rule they promulgated is vague and formless, rendering it almost entirely meaningless. Second, the Court failed to consider democratic policy. This is particularly noteworthy given that the case involved the actions of democratic governments.

This Note seeks to fill the gap left by the Supreme Court by providing a policy framework for evaluating when it is proper to restrict access to a government work through copyright. The framework consists of two parts, each based on distinct policies.

The first policy considered is the incentive-access framework that underlies copyright law.[13] Its main thrust is that although the power to exclude puts limits on access, incentivizing creators with these rights is justified because this creates a system that produces more information than a free market.[14] It follows that when no incentives are needed to produce a work, a restriction in access is unjustified.[15] This constitutes the first part of the framework: the government should not be able to restrict access to one of its works when no incentives are needed for its production.

The second consideration is democracy. A democracy is a form of self-government where the people consent to be governed.[16] Since it is impossible for a large group of people to agree on exactly how to be governed, it is imperative that people identify with the *process* of governing.[17] In this way, an individual in the minority can consent to be governed because they identify

10. *See, e.g.*, Brief for the Petitioners at 19, *Georgia*, 140 S. Ct. 1498 (No. 18-1150).

11. *See, e.g.*, Brief of Respondent at 19, *Georgia*, 140 S. Ct. 1498 (No. 18-1150).

12. *Georgia*, 140 S. Ct. at 1508.

13. David W. Barnes, *The Incentives/Access Tradeoff*, 9 NW. J. TECH. & INTELL. PROP. 96, 98 (2010).

14. *See id.* at 97.

15. *See* Bracha & Syed, *supra* note 5, at 1850 ("However, for goods that are nonrivalrous in consumption—of which informational works are a paradigm example—exclusionary rights may function inefficiently, wastefully preventing uses that would not detract from simultaneous use by others. The justification for incurring this potential inefficiency is, of course, that without it some informational works may fail to be developed in the first place.").

16. *See, e.g.*, THE DECLARATION OF INDEPENDENCE para. 2 (U.S. 1776).

17. Robert Post, *Democracy and Equality*, 603 ANNALS AM. ACAD. POL. & SOC. SCI. 24, 26 (2006).

with the process of governing.[18] In other words, they may disagree with the result but agree with the process by which that result was reached.

Democratic governments must be responsive to the public opinion for consent to be valid.[19] For a government to be truly democratic, there can be no inequalities of opportunity to meaningfully participate in forming public opinion.[20] If there are inequalities, then the minority's consent to be governed would be meaningless; they would not have the same opportunity to form the public opinion, so they cannot identify with the process of governing.[21]

Governmental restriction to access to certain works through copyright is tantamount to limiting the ability of some people to participate in the process of forming the public opinion. Certain government works carry great weight when informing the public opinion because they are official statements from the government itself on how it is governing. Copyrighting these works functions as a paywall. Those who can afford unlimited access to works thus have better information and can shape public opinion more effectively. For those unable to afford access, cost acts a barrier to influencing the public opinion effectively. This systematic tilt in who can meaningfully participate in the political process is undemocratic. Accordingly, the second prong of the framework states that it is improper to restrict access to government works through copyright when it would be undemocratic to do so.

In sum, this Note argues that it is improper to restrict access to a work produced or adopted by the government when copyright is not needed to incentivize the work or when it would be undemocratic to do so. Applying this policy framework, this Note will show that the Court reached the right decision with respect to Georgia's annotations, given the Court's limited options of either allowing copyright or destroying it completely. This Note will then show how restricting access to other government works depends on the nature of the works, rather than which governmental body created them.

This Note proceeds in five parts. Part II provides background on the government edicts doctrine and copyright law. Part III reviews *Georgia v. PRO* and argues that the rule created by the majority is inadequate. Part IV starts by reviewing the policies at play in considering when it is proper to restrict access to a work produced by a government. It then argues for this Note's policy framework and applies the framework to four categories of government works. Part V concludes.

---

18. *See id.* at 26–27, 29.
19. *See id.* at 29.
20. *See id.* at 33.
21. *See id.* at 26–27, 29.

II. BACKGROUND

This Part will explain what copyright law is, situate the government edicts doctrine within copyright, and explain the black letter law behind the government edicts doctrine that the Court relied on in *Georgia v. PRO*.

A. COPYRIGHT LAW

Copyright is a right granted to creators of marginally creative works, which empowers them to prevent others from copying their work.[22] Preventing others from copying is essential to a creator's ability to charge for a work.[23] Without copyright protection few would pay for a creative work, as they could obtain it elsewhere for free.[24] In this way, copyright incentivizes creators to create by guaranteeing them the right to restrict access and thereby the ability to set a price.

B. CONGRESSIONAL (IN)ACTION

By its text, the Copyright Act of 1976 does not prevent states or state governments from holding copyrights,[25] though it does prohibit the federal government from doing so.[26] In so doing, "Congress rejected efforts 'to extend the prohibition' against copyrighting U.S. government works 'to publications of State and local governments.'"[27]

C. THE ACCESS PROBLEM

When a price is set through the exclusionary power granted by copyright, it creates a class of people who wish to access a work but cannot afford it. At certain times, this is impermissible. Take the following example: A state obtained copyright protection over their criminal law and charged for access to it. Notions of fundamental fairness dictate that this is wrong. All citizens are required to know the law and structure their actions accordingly, but many would be unable to access it. The government edicts doctrine remedies this by barring the government from holding copyright in certain works, like the law itself.

---

22. *See* 17 U.S.C. § 106; Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 362 (1991).

23. *See* Barnes, *supra* note 13, at 107.

24. *See id.* at 96–97.

25. 1 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 5.14 (2021).

26. 17 U.S.C. § 105.

27. Brief for the Petitioners at 29, Georgia v. Public.Resource.Org, Inc., 140 S. Ct. 1498 (2020) (No. 18-1150) (citing STAFF OF H. COMM. ON THE JUDICIARY, 87TH CONG., COPYRIGHT LAW REVISION: REPORT OF THE REGISTER OF COPYRIGHTS ON THE GENERAL REVISION OF THE U.S. COPYRIGHT LAW 129–30 (Comm. Print 1961)).

D. THE GOVERNMENT EDICTS DOCTRINE

Congress did not bar state or local governments from holding copyrights in certain works, but the courts have crafted a doctrine that does—the government edicts doctrine.[28] Prior to *Georgia v. PRO*, the exact scope of the doctrine was not clear, but there were two main ideas stemming from three 19th century cases. The first idea was that works with the force of law could not be copyrighted by the government.[29] The second was that legal works adopted by or published under the authority of the government cannot be copyrighted.[30] Under either theory, the government edicts doctrine prevents states from copyrighting the law.

E. THE CASES

1. *Wheaton v. Peters*

Wheaton v. Peters, decided in 1834,[31] was the first copyright case to ever reach the Supreme Court.[32] Wheaton, the Supreme Court's first official reporter, sued Peters, the second official reporter,[33] for copyright infringement.[34] Peters had copied from Wheaton's reports[35] and annotations[36] in preparing condensed reports of previous Supreme Court terms. Wheaton's reports contained no official Supreme Court or U.S. markings, solely "Henry Wheaton, Counsellor at Law,"[37] and were created without any supervision from the Court.[38] Wheaton argued that he had acquired a copyright in the Court's opinions "by judges' gift" after he had been appointed to report the Court's decisions by an act of Congress.[39]

The Court rejected this argument, stating that they were "unanimously of opinion, that no reporter has or can have any copyright in the written opinions delivered by this court; and that the judges thereof cannot confer on any reporter any such right."[40] Interestingly, the Court spent perhaps the least amount of time on this issue, focusing the bulk of its lengthy analysis on

28. *See Georgia*, 140 S. Ct. at 1503–04.
29. *See, e.g.*, Brief for the Petitioners at 19, *Georgia*, 140 S. Ct. 1498 (No. 18-1150).
30. *See, e.g.*, Brief of Respondent at 19, *Georgia*, 140 S. Ct. 1498 (No. 18-1150).
31. Wheaton v. Peters, 33 U.S. 591, 591 (1834).
32. 1 NIMMER ON COPYRIGHT, *supra* note 25, § 5.12.
33. *See, e.g.*, Brief for the Petitioners at 32, *Georgia*, 140 S. Ct. 1498 (No. 18-1150).
34. *Wheaton*, 33 U.S. at 612.
35. *Id.* at 617.
36. *Id.* at 651.
37. *See, e.g.*, Brief of Respondent at 26, *Georgia*, 140 S. Ct. 1498 (No. 18-1150).
38. *See, e.g.*, *id.*
39. *Wheaton*, 33 U.S. at 614.
40. *Id.* at 668.

whether Wheaton could assert a common law claim to avoid compliance with formalities required for federal copyright protection and, if not, whether he had to comply with said formalities.[41] Ultimately, the Court remanded to the circuit court to determine whether Wheaton had complied.[42] By doing this, the Court implicitly held that there was copyrightable expression in Wheaton's reports, such as his annotations; if there was no copyrightable expression in the reports, there would have been no need to remand to determine whether the formalities were complied with.[43]

Thus, the government edicts doctrine was born—the idea that certain works produced by the government cannot be copyrighted.

2. *Banks v. Manchester*

More than fifty years later, the Supreme Court expounded on the government edicts doctrine with two more cases authored by Justice Blatchford. The first, *Banks v. Manchester*, was decided on November 19, 1888, less than one month before the second, *Callaghan v. Myers*.[44] In *Banks*, the state of Ohio gave Derby & Co. the exclusive right to report its highest court's decisions for two years, provided that the reporter would obtain a copyright in the reports on behalf of the state.[45] The reports included "not only the opinion . . . of the court . . . but also the statement of the case, and the syllabus," all of which were created by the judges.[46] Derby & Co. assigned its rights to the contract to Banks & Brothers.[47] After this, Manchester printed two cases nearly verbatim from Banks's reporter in the American Law Journal without Banks's permission, prompting a lawsuit for copyright infringement.[48]

The Court extended *Wheaton* from the federal judiciary to the state judiciary and held that the state could not hold a copyright in judges' work, including the statement of the case or the syllabus.[49] Justice Blatchford came to this conclusion for two reasons. First, he noted that judges are paid for their work, so they have no monetary interest in their works against the public.[50] Second, as a matter of policy, "the products of the labor done by judicial officers in the discharge of their judicial duties" cannot be copyrighted because "[t]he whole

41. *Id.* at 654–67.
42. *Id.* at 667–68.
43. Brief for the Petitioners at 33, *Georgia*, 140 S. Ct. 1498 (No. 18-1150).
44. Banks v. Manchester, 128 U.S. 244, 244 (1888); Callaghan v. Myers, 128 U.S. 617, 617 (1888).
45. *Banks*, 128 U.S. at 246–48.
46. *Id.* at 251.
47. *Id.* at 247.
48. *See id.* at 247–49.
49. *See id.* at 252–53.
50. *Id.* at 253.

work done by the judges constitutes the authentic exposition and interpretation of the law, which, binding every citizen, is free for publication to all, whether it is a declaration of unwritten law, or an interpretation of a constitution or a statute."[51] Justice Blatchford held that the statement of the case was ineligible for copyright despite the fact that no judicial concurrence was required for its publication.[52]

3. *Callaghan v. Myers*

The Supreme Court decided *Callaghan v. Myers*, the last government edicts case (until *Georgia v. PRO*), on December 17, 1888.[53] Myers was the assignee of Norman L. Freeman, the Illinois reporter.[54] Unlike in *Banks*, Freeman himself created the syllabi and headnotes for the cases he reported.[55] But as in *Wheaton*, the reporters simply stated "By Norman L. Freeman, counselor at law."[56] Callaghan copied Myers's work wholesale, including the title of Myers's reports.[57] The Court held that Myers held a valid copyright in the non-judicially created parts of the works.[58] It reasoned,

> [A]lthough there can be no copyright in the opinions of the judges, or in the work done by them in their official capacity as judges, . . . there is no ground of public policy on which a reporter who prepares a volume of law reports, of the character of those in this case, can, in the absence of a prohibitory statute, be debarred from obtaining a copyright for the volume which will cover the matter which is the result of his intellectual labor.[59]

Drawing from his previous decision in *Banks*, Justice Blatchford set forth a rule that ostensibly turned on whether a work was created in the judge's official capacity. The Supreme Court then did not rule on the government edicts doctrine until *Georgia v. PRO*, some 130 years later.[60]

---

51. *Id.*
52. *See id.* at 250, 253–54.
53. Callaghan v. Myers, 128 U.S. 617, 617 (1888).
54. *See id.* at 619.
55. *Id.* at 621.
56. Brief of Respondent at 31, Georgia v. Public.Resource.Org, Inc., 140 S. Ct. 1498 (2020) (No. 18-1150).
57. *See Callaghan*, 128 U.S. at 629.
58. *Id.* at 647.
59. *Id.*
60. *See Georgia*, 140 S. Ct. at 1498; *Callaghan*, 128 U.S. at 629.

III. CASE REVIEW AND CRITIQUE

A. THE CASE

1. *Facts*

Georgia v. Public.Resource.Org,, Inc., started on July 21, 2015 in the Northern District of Georgia.[61] Georgia's Code Revision Commission, on behalf of the State of Georgia, filed a lawsuit against Public.Resource.Org, Inc. (PRO) for infringing Georgia's copyright in the annotations accompanying its one and only official code,[62] the Official Code of Georgia Annotated (hereinafter OCGA).[63] PRO posted the entirety of the OCGA online, ignoring the State's cease-and-desist letters.[64] The district court found that PRO infringed the copyright in the OCGA annotations and that the affirmative defense of fair use did not apply.[65]

The Eleventh Circuit disagreed.[66] It found that the annotations were not copyrightable under the government edicts doctrine and reversed.[67] The Supreme Court granted certiorari and affirmed.[68]

The OCGA, published by Georgia, contains every Georgia statute, along with annotations.[69] The annotations contain, among other things,

> summaries of judicial decisions applying a given provision, summaries of any pertinent opinions of the state attorney general, . . . [and] information about the origins of the statutory text, such as whether it derives from a particular judicial decision or resembles an older provision that has been construed by Georgia courts.[70]

Beyond being a research tool, the annotations contain information important for the citizenry. Interested readers are able to learn from the annotations that various aspects of the Code are unenforceable, though the Georgia legislature

---

61. Complaint for Injunctive Relief at 2, Code Revision Comm'n v. Public.Resource.Org, Inc., 244 F. Supp. 3d 1350 (N.D. Ga. 2017), *rev'd*, 906 F.3d 1229 (11th Cir. 2018), *aff'd*, 140 S. Ct. 1498 (2020) (No. 1:15-cv-2594-MHC), 2015 WL 4999975.

62. *Georgia*, 140 S. Ct. at 1505.

63. *Id.* at 1504.

64. *Id.* at 1505.

65. *Public.Resource.Org*, 244 F. Supp. 3d at 1356–61, *rev'd*, 906 F.3d 1229 (11th Cir. 2018), *aff'd*, 140 S. Ct. 1498 (2020).

66. *Public.Resource.Org*, 906 F.3d at 1243.

67. *Id.*

68. *See Georgia*, 140 S. Ct. at 1506, 1513.

69. *Id.* at 1504.

70. *Id.*

"has not bothered to narrow or repeal" them.[71] For example, the text of § 16-6-2 prohibits certain consensual acts, but the annotations reassure the 21st century reader that this part of the Code was held to violate Georgia's constitutional right to privacy, making it unenforceable.[72]

The annotations are assembled by the Code Revision Commission (the Commission) and published annually by the state legislature.[73] The Commission, established in 1977 by Georgia's legislature, is responsible for "consolidating disparate bills into a single Code for reenactment . . . and [for] contracting with a third party to produce the annotations."[74] It is funded from the pot of money allocated to the legislative branch, and a majority of the fifteen members must be elected Georgia legislators.[75]

Since 1978, the Commission has contracted with a division of LexisNexis to produce the annotations.[76] Per stipulation, any copyright in the OCGA vests in Georgia, but LexisNexis is granted the exclusive right to publish, distribute, and sell the OCGA, provided they limit the price of a hard copy to $412 and provide an unannotated version of the statutes online for free.[77] The Commission supervises Lexis's work and "specifies what the annotations must include in exacting detail."[78]

Once the annotations have been created, the Commission submits the proposed statutory text and annotations to the legislature for approval.[79] The legislature then (1) enacts the statutes, (2) merges the annotations with the statutes, and (3) publishes the final merged product.[80]

2. *Chief Justice Roberts's Majority Opinion*

Chief Justice Roberts's opinion started by reviewing precedent. From *Wheaton*, he gathered that judicial opinions are not copyrightable; from *Banks*, that works done by judges in their capacity as such are not copyrightable; and from *Callaghan*, a "limiting principle" that headnotes and syllabi created by the

---

71. *Id.* at 1512.

72. *See* GA. CODE ANN. § 16-6-2 (2021) (prohibiting acts of sodomy); *see also* GA. CODE ANN. §§ 21-2-131, 16-6-18, 16-15-9 (2021) (noting in the annotations that the statutes are unenforceable in key aspects).

73. *Georgia*, 140 S. Ct. at 1504.

74. *Id.*

75. *Id.*

76. *See id.* at 1505; GA. CODE ANN. § 1-1-1 (2021).

77. *Georgia*, 140 S. Ct. at 1505.

78. *Id.*; *see also* Joint Appendix (Volume 3 of 3) at 269–78, 286–427, *Georgia*, 140 S. Ct. 1498 (No. 18-1150) (specifying what the annotations must contain).

79. *Georgia*, 140 S. Ct. at 1504.

80. *Id.* at 1504–05; *see also* GA. CODE ANN. § 1-1-1.

reporter are copyrightable because the reporter "had no authority to speak with the force of law."[81]

From this, Roberts crafted a rule that focuses on the author: "[u]nder our precedents, therefore, copyright does not vest in works that are (1) created by judges and legislators (2) in the course of their judicial and legislative duties."[82] Roberts also provided a surface-level policy justification: "The animating principle behind this rule is that no one can own the law. 'Every citizen is presumed to know the law,' and 'it needs no argument to show…that all should have free access' to its contents."[83] The majority thus crafted a rule based on the identity of the author and the context in which the work was created, but the majority did not connect it with deeply-held values or substantive policies.[84]

Moving on to application, the Court found that both parts of its test were satisfied. In evaluating the first part, whether the work was created by a legislator, the Court concluded that the Commission was "an adjunct to the legislature" and thus satisfied part one.[85] In reaching this decision, the Court first noted that the Commission was the author of the annotations under copyright law's work-for-hire doctrine.[86] It then relied on six key factors to find that the Commission was a legislator: (1) the Commission consists mainly of legislators, (2) it is funded from the part of the budget allocated to the legislative branch, (3) it was created *by* the legislative branch, (4) it was created *for* the legislative branch, (5) the statutes were merged with the annotations by the legislature, (6) and Georgia's highest court had held that the duties performed by the Commission are within the sphere of legislative authority.[87]

To find that the second part of the test, whether the annotations were part of the Commission's duties, was met, the Court solely relied on the proclamation of Georgia's highest court that "the Commission's preparation of the annotations is under Georgia law an act of 'legislative authority.'"[88]

With its rule created and applied, the Court moved on to rebutting counterarguments. The first was that the annotations should be copyrightable because the Copyright Act explicitly provides that annotations are eligible for

81. *Georgia*, 140 S. Ct. at 1506–07.
82. *Id.* at 1508.
83. *Id.* at 1507 (citing Nash v. Lathrop, 6 N.E. 559, 560–61 (Mass. 1886)).
84. *See id.* at 1508.
85. *Id.* at 1509.
86. *Id.* at 1508.
87. *Id.*
88. *Id.* at 1509 (citing Harrison Co. v. Code Revision Comm'n, 260 S.E.2d 30, 34–35 (Ga. 1979)).

copyright.[89] The Court agreed that annotations are eligible for copyright but said that for copyright to be granted, works must be authored, and the Commission does not qualify as an author under the government edicts doctrine.[90] The second counterargument was that Congress did not intend to abrogate states' ability to hold copyrights, as it explicitly barred the federal government, not state governments, from holding copyright in their works.[91] The Court countered that Congress never intended to abrogate the government edicts doctrine either.[92] The Court then wrestled with the third contention—that the appropriate rule should focus on whether the work carried the force of law.[93] To dismantle this argument, Roberts turned to *Banks*, which withheld copyright from dissenting and concurring opinions, as well as headnotes and syllabi produced by the court, all of which ostensibly lack the force of law.[94] The Court reiterated that the case's holding was directed at works prepared by a judge in their official capacity.[95]

Finally, Roberts noted the high "practical significance" of the OCGA,[96] indicating that there was more going on than a strict textual analysis. As an example, Roberts pointed to the aforementioned constitutional limitations on unrepealed statutes in the annotations.[97] That is, "[i]f everything short of statutes and opinions were copyrightable, then States would be free to offer a whole range of premium legal works for those who can afford the extra benefit."[98] These concerns point to reasoning that is not solely textually-based but is also grounded in certain public policies.

---

89. *Id.*
90. *Id.*
91. *Id.*
92. *Id.* at 1510.
93. *Id.* at 1511.
94. *See id.* at 1511–12.
95. *See id.* at 1511–12.
96. *See id.*
97. *Id.* ("Imagine a Georgia citizen interested in learning his legal rights and duties. If he reads the economy-class version of the Georgia Code available online, he will see laws requiring political candidates to pay hefty qualification fees (with no indigency exception), criminalizing broad categories of consensual sexual conduct, and exempting certain key evidence in criminal trials from standard evidentiary limitations—with no hint that important aspects of those laws have been held unconstitutional by the Georgia Supreme Court. Meanwhile, first-class readers with access to the annotations will be assured that these laws are, in crucial respects, unenforceable relics that the legislature has not bothered to narrow or repeal." (citation omitted)).
98. *Id.* at 1512–13. Roberts also opined that if this were the case, "[t]he less bold among us would have to think twice before using official legal works that illuminate the law we are all presumed to know and understand." *Id.* at 1513.

3. *Justice Thomas's Dissent*

Justice Thomas dissented, along with Justices Alito and Breyer.[99] Justice Thomas first noted his disagreements with the majority's application of their rule, then moved on to criticize the rule itself.[100]

Thomas constructed a rule he believed to align with precedent. Thomas's analysis of *Wheaton* agreed with the majority's in finding that judicial opinions cannot be copyrighted.[101] From *Banks*, Thomas derived that materials prepared by judges in their capacity as such are not authored, and the *Banks* Court did not "categorically prohibit[] [States] from holding copyrights."[102] Moving on to *Callaghan*, Thomas gleaned that a reporter's notes are copyrightable.[103] From this, Thomas derived a rule that focused on whether a work has the force of law: "statutes and regulations cannot be copyrighted, but accompanying notes lacking legal force can be."[104]

Moving on from his textual analysis, Justice Thomas recognized an important policy goal to support his legal force-only stance. In support of his argument that judges cannot be authors because their salary is paid by and represents the will of the people, he noted that "copyright law understands an author to be one whose work will be encouraged by . . . an exclusive right."[105]

In response to the majority's argument that a rule focusing on whether a work has the force of law incorrectly addresses concurring and dissenting opinions, Justice Thomas distinguished annotations from non-binding judicial opinions giving context of the case. He acknowledged that only the judgment has "legal effect" but retorted that concurrences and dissents are necessary for "provid[ing] pivotal insight into how the law will likely be applied in future judicial opinions."[106] Interestingly, this is precisely what the majority argued the annotations do—provide insight into how the law will be applied—but Justice Thomas did not distinguish the annotations from non-binding opinions.[107]

Drawing parallels between the annotations in the OCGA and the "privately created annotations" in *Callaghan*, Justice Thomas ultimately

99. *Id.* at 1513 (Thomas, J., dissenting). Justice Breyer abstained from joining Part II-A of the dissent, which dealt with precedent and the judicial role.

100. *Id.* at 1513–14.

101. *Id.* at 1514.

102. *Id.* at 1514–15.

103. *Id.* at 1515.

104. *Id.*

105. *Id.* at 1516.

106. *Id.* at 1520.

107. *See id.* at 1512–13 (majority opinion).

concluded that the annotations are copyrightable.[108] The annotations are not the law, he argued, because they are not enacted and are instead merely merged with enacted text.[109] As such, they do not embody "the will of the people."[110] Justice Thomas thus argued that the annotations should not be exempt from copyright protection.[111]

4. *Justice Ginsburg's Dissent*

Justice Ginsburg also dissented, along with Justice Breyer, focusing her disagreements entirely on the majority's application of the rule it announced.[112] The nub of her argument was that the annotations were not created in a legislative capacity.[113] Implicit in this is her acceptance of the majority's rule.

To rebut the majority's application of the first part of its test, Ginsburg argued that the scope of a legislature's duties is limited to making laws.[114] "'[T]o the judiciary' we assign 'the duty of interpreting and applying' the law . . . [whereas] the role of the legislature encompasses . . . 'making laws.'"[115] The annotations are interpreting laws, putting them within the province of the judiciary and not the legislature, she argued.[116]

As to the second part of the majority's test, Ginsburg argued that the annotations were not part of the legislative duty of making laws for three reasons. First, annotations are not laws because "annotating begins only after lawmaking ends."[117] Second, they are "descriptive rather than prescriptive."[118] And third, they are "given for the purpose of convenient reference."[119]

Justice Ginsburg only took the time to dismiss one counterargument: the merging of the annotations with the statutes did not make them law, because the Supreme Court of Georgia held that placing the annotations in the official code did not give them any official weight.[120] Thus, Justice Ginsburg's dissent primarily argued that the majority's rule was misapplied.

---

108. *See id.* at 1518 (Thomas, J., dissenting).
109. *Id.* at 1517.
110. *See id.*
111. *See id.* at 1522.
112. *See id.* at 1522–23 (Ginsburg, J., dissenting).
113. *See id.* at 1523.
114. *See id.*
115. *Id.* (citing Massachusetts v. Mellon, 262 U.S. 447, 488 (1923)).
116. *See id.*
117. *Id.*
118. *Id.*
119. *Id.* at 1524 (internal citations omitted).
120. *See id.*

B. CRITIQUE OF THE MAJORITY'S APPLICATION

To the majority's credit, it calculated an ostensibly clear rule from precedent—"copyright does not vest in works that are (1) created by judges and legislators (2) in the course of their judicial and legislative duties"—but its application is impractical and confusing, especially the second step.[121] To see how the rule was misapplied, it is first critical to be clear about the exact nature of the annotations.

1. *Defining the Annotations*

The annotations contain two key sets of information. First, they contain important information on how the law will be applied. They explain constitutional limitations on laws and judge-made exceptions to the laws.[122] As mentioned earlier, they mark certain aspects of unrepealed statutes as unconstitutional.[123] Also included in the annotations are statements by the state attorney general interpreting the law and explaining how it will be enforced.[124] All of this is essential information on how the statutes will be applied—information that the text of the statutes themselves simply does not supply.

Second, and closely related, the annotations also include information on how a statute should be interpreted.[125] For example, the annotations can include selected court interpretations of a statute.[126] Court opinions give precise and grounded examples of a statute's interpretations, providing vivid detail on the meaning of a statute that may not be understood from a plain-text reading. Further, many annotations explain the origin of a statute.[127] The origin of a statute can be key in determining the meaning of the text. Language that comes from common law can, and often does, mean something entirely different than the same language used in common parlance. Another example is the inclusion of law review articles.[128] Law review articles often work through the meanings, implications, and applications of key aspects of statutes. On their own, these do not carry much weight outside of academic circles. However, when cited by the legislature as an official aside to a statute, these articles serve as marginally authoritative interpretations.

Taking these two parts together, the annotations are the legislature's official guidelines on how the laws they have promulgated will be applied and

121. *See id.* at 1508 (majority opinion).
122. *See id.* at 1504, 1512.
123. *Id.* at 1512.
124. *See id.* at 1504.
125. *Id.*
126. *Id.*
127. *Id.*
128. *Id.*

how they should be interpreted. With this characterization of the annotations in mind, the next step in showing that the majority misapplied their rule is to make clear what a legislator's duties are and how the annotations fall within them.

2. *The Duty of a Legislature and the Role of the Annotations*

The legislature's job is to solve problems by creating laws.[129] Legislators introduce bills to address certain problems. These bills become enforceable, binding law only if they pass bicameralism and presentment, two high hurdles.[130]

The annotations in this case are not laws that pass bicameralism and presentment and are therefore not within the duty of the legislature. Rather, the annotations are ex post interpretations and explanations of the law. This undertaking is more aligned to the judiciary's role. As *Marbury v. Madison* put it, "[i[t is emphatically the province and duty of the judicial department to say what the law is."[131] This is not to say that the annotations are not extremely valuable, or that the legislature should not create the annotations. It is just not their duty.

The holding of the Georgia Supreme Court that "the work of the Commission; i.e., selecting a publisher and contracting for and supervising the codification of the laws enacted by the General Assembly, including court interpretations thereof, is within the sphere of legislative authority" does not overturn this conclusion.[132] The nub of this counterargument has a few steps. First, it assumes that the state supreme court has the authority to interpret its own laws and constitution.[133] Second, therefore, if the court says that a task of the Commission is "within the sphere of legislative authority," it is authoritative.[134] The third step is where the error, the false inference, enters: something that is within the sphere of the legislature's authority is therefore their "dut[y]."[135]

To show the falsity of this inference, consider this example. When someone is assigned a job, say a partner at a law firm, their sphere of authority

129. *See, e.g.*, Antonin Scalia, *The Rule of Law as a Law of Rules*, 56 U. CHI. L. REV. 1175, 1176 (1989) ("Executives and judges handle individual cases; the legislature generalizes.").

130. *See, e.g.*, U.S. CONST. art. I, § 7. This happens in Congress and all states except Nebraska, which has a unicameral legislature.

131. Marbury v. Madison, 5 U.S. (1 Cranch) 137, 177 (1803).

132. Harrison Co. v. Code Revision Comm'n, 260 S.E.2d 30, 34–35 (Ga. 1979).

133. *See, e.g.*, U.S. CONST. amend. X.

134. *See* Georgia v. Public.Resource.Org, Inc., 140 S. Ct. 1498, 1508 (2020) (emphasis omitted) (citing *Harrison*, 260 S.E.2d at 34–35).

135. *See id.*

is always as big as or bigger than the sphere of their duties. This must be the case, otherwise the assigned duties would be meaningless.[136] If a law firm partner's role/duty is managing IP cases for large clients, it is certainly within her authority to direct an associate to stay at the office researching a key issue. This exercise of authority corresponds almost exactly with her duties. However, it is also within her authority to instruct maintenance staff to replace a hallway lightbulb. This is not within her duties at the law firm but is rather an auxiliary task that they have been given authorization to delegate.

The Commission's creation of the annotations is similar. It is not within their duties to create the annotations; it is an auxiliary task they have authority over. Accordingly, the statement from Georgia's Supreme Court does not override the fact that annotations are not within the scope of a legislature's duties, as the U.S. Supreme Court's test requires.

C. CRITIQUE OF THE MAJORITY'S RULE

Beyond being misapplied, the rule itself is problematic. When applying the rule, the Court created and utilized a framework that lower courts are likely to follow. Unfortunately, the framework allows the rule to be circumvented, severely blunting its efficacy. Moreover, the Court missed an easy opportunity to craft a broader rule, and this lapse has real-world consequences.

1. The Rule is Easily Circumvented

In deciding that each part was met, the Court relied on a sparse number of factors. In holding that the first part of the test was met, the Court relied entirely on six facts, but this Note will focus on the four that are easily evadable: First, the Commission consisted mostly of legislators.[137] Second, it was funded by the legislative budget.[138] Third, the Commission was created directly by the legislative branch.[139] And fourth, the annotations were merged with the Code under the legislature's authority.[140] To find that the second part was met, the Court only identified one factor: the Georgia Supreme Court had held that the Commission's production of the annotations was "an act of 'legislative authority.'"[141]

136. There can be concurrent duties. It can be two people's duty to ensure a task is completed.

137. *Georgia*, 140 S. Ct. at 1508.

138. *Id.*

139. *Id.*

140. *Id.*

141. *Id.* at 1509 (citing Harrison Co. v. Code Revision Comm'n, 260 S.E.2d 30, 34–35 (Ga. 1979)).

With this framework in mind, states looking to maintain the copyright in their annotations can avoid the government edicts bar. To escape part one of the test, states can undertake a combination of the following: ensure that elected legislators only constitute a minority of that state's commission; move the commission to the executive branch; fund the commission with money from a branch other than the legislative branch; or merge the annotations without the legislature's approval. The Supreme Court's test would likely not be met if four of the six factors were not present.

The second part is seemingly easier to escape. The Court merely relied on a statement from Georgia's Supreme Court that the annotations were an act of legislative authority and inferred that their creation was therefore within the duty of a legislature.[142] It is doubtful that other states have such a holding from their respective supreme courts. Moreover, this inference can be challenged; a task that is within an entity's authority is not necessarily their duty. Thus, through the Supreme Court's precedent-setting application of its rule, it opened many loopholes for states to avoid the government edicts doctrine, rendering the rule ineffective.

2. *The Rule is Not Broad Enough*

More importantly, however, the Supreme Court missed the opportunity to create a broader rule. Under the current rule, no executive materials are covered.[143] Thus, statements from state attorneys general, state executive orders, and directives from state agencies can all be copyrighted. These are not mere speculations; Michigan has copyrighted its constitution, and South Dakota's executive orders are copyrightable by statute.[144] This is flatly incongruous. If annotations, which are not binding law, are not copyrightable, executive orders and state constitutions should also be ineligible for copyright. However, the Supreme Court did not address this issue, so these practices can be expected to continue.

3. *Stakes*

The stakes of these shortcomings are significant. The government produces large quantities of information, much of which has great political and

142. *Id.*

143. *Id.* at 1508.

144. Copyright Registration No. TX0002908667 (registered Apr. 16, 1990) (Michigan Constitution), http://cocatalog.loc.gov; S.D. CODIFIED LAWS § 2-16-6(b), § 2-16-8 (2021) (making South Dakota's executive orders copyrightable); *see also* Brief *Amici Curiae* of the Am. C.L. Union et al. in Support of Respondent at 4–6, *Georgia*, 140 S. Ct. 1498 (No. 18-1150).

economic significance.[145] By copyrighting works, the government excludes part of the citizenry from accessing this information. This exclusion is inconsistent with democratic principles broadly recognized by the United States.[146] As the Second Circuit recognized:

> The evils inherent in allowing government to create a monopoly over the dissemination of public information in any form seem too obvious to require extended discussion. Government may add its own voice to the debate over public issues, but it may not attempt to control or reduce competition from other speakers. … Such actions are an exercise of censorship that allows the government to control the form and content of the information reaching the public.[147]

The rule's shortcomings stem from a failure to examine the policies at play in the government edicts doctrine. The remainder of this Note is dedicated to doing exactly that—grounding this issue in policy and arguing for a framework on which future rules should be based.

IV. A NEW FRAMEWORK

Works produced by the government or adopted by reference should not have their access restricted when no incentives are needed for that work's production or when it would be undemocratic to do so. Section IV.A brings in model codes and engineering standards to provide a more complete picture of works utilized by the government. Section IV.B provides detailed background on the policies at play in this area of law. Section IV.C explains two different metrics for measuring the success of the copyright system. Section IV.D pulls together the policy arguments into a cohesive framework. Finally, Section IV.E applies the framework to five types of works commonly utilized by the government.

A. MODEL CODES AND ENGINEERING STANDARDS: PRIVATE COPYRIGHT OVER BINDING LAW

This Note has already discussed two common types of government works: annotations, which have been explained, and legal documents, such as judicial opinions and statutes, which do not require explanation. But there is another important tool of lawmaking in the government's belt: utilizing privately developed model codes and standards. In short, when a private entity has

145. Robert M. Gellman, *Twin Evils: Government Copyright and Copyright-Like Controls over Government Information*, 45 SYRACUSE L. REV. 999, 1003 (1995).

146. *Id.* at 1007.

147. Legi-Tech, Inc. v. Keiper, 766 F.2d 728, 733 (2d Cir. 1985) (citations omitted).

developed a set of rules that the government wishes to use, the government will incorporate it into binding law.[148] These private works are often copyrighted, and thus a private entity often holds a copyright on binding law.[149] Since this is similar to when a government holds a copyright over a work, this Note will extend its analysis to cover standards and model codes. However, to do so requires a more detailed explanation of model codes and standards and how they are used.

Model codes and standards with binding power are developed by private organizations and later incorporated into the law.[150] Often they are copyrighted.[151] For example, an engineering standard detailing the design specifications of pressure vessels might be developed privately among boiler manufacturers, and the government might later enact it as law. Enacting that standard or model code is done in one of two ways. First, the government might "copy and paste" it into the law.[152] Second, the government might reference the code or standard as binding law—incorporation by reference.[153] In either case, the outcome is the same. A private entity holds the copyright to a piece of binding law.

Despite due process and democratic concerns, the government adopts model codes and standards because it conserves resources.[154] Developing the expertise necessary to competently rule a given area is extremely expensive and time-consuming.[155] It is wasteful for governments to redevelop codes or standards that already exist.[156]

Moreover, it is extremely important that standards and model codes are promulgated. Every time an email is sent, over 200 standards are utilized.[157] But standards are not limited to the internet. The federal government has incorporated about 9,500 standards by reference, covering fields like consumer safety and home sprinkler systems.[158] To further illustrate the importance of standards, consider the following example.

---

148. Nina A. Mendelson, *Private Control over Access to the Law: The Perplexing Federal Regulatory Use of Private Standards*, 112 MICH. L. REV. 737, 739, 743 (2014).

149. *Id.* at 743.

150. *Id.* at 739.

151. *Id.* at 743.

152. Jim Sergent, Matt Wynn & Aamer Madhani, *They copied bills from corporations. These lawmakers say that's OK.*, USA TODAY (Dec. 17, 2019), https://www.usatoday.com/in-depth/news/investigations/2019/07/17/lawmakers-ok-with-copying-model-bills/3424549002.

153. Mendelson, *supra* note148, at 739.

154. *See id.* at 776.

155. *See id.*

156. *See id.* at 750.

157. Pamela Samuelson, *Questioning Copyrights in Standards*, 48 B.C. L. REV. 193, 193 (2007).

158. Mendelson, *supra* note 148, at 739, 750, 752.

In 1904, a fire started in Baltimore and quickly grew to a size beyond the local fire department's ability to control.[159] Baltimore telegraphed for help, and many neighboring precincts responded.[160] However, upon arriving they found that their hoses did not fit Baltimore's hydrants.[161] They watched a large portion of the city burn to the ground, unable to help.[162] If the size of the hydrant and hose fitting were standardized, this would not have happened. By creating certainty, standards are both beneficial and necessary in today's market of technically complex goods.[163]

Yet standards create situations in which private entities can restrict access to binding law through copyright. To account for the full range of binding and explanatory works utilized by the government whose access might be restricted through copyright, the question thus becomes: when does copyright impermissibly restrict access to works in which copyright is held by the government or to works utilized by the government in which copyright is held privately?

B. POLICY BACKGROUND

Answering this requires examining policies relevant to the government edicts doctrine: the economic framework that underlies copyright law and democratic considerations.

1. *The Incentive-Access Framework*

The Framers of the Constitution authorized Congress "[t]o 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."[164] This is the constitutional basis for Congress's creation of copyright and patent law. Implicit in this is a statement of the incentive-access framework—the paradigm by which copyright and patent law are justified.

Copyright and patent law ultimately protect ideas and information.[165] The complex system of laws guides this information through the market with rules on what information qualifies for protection and when the information can be

159. Emily S. Bremer, *On the Cost of Private Standards in Public Law*, 63 KAN. L. REV. 279, 300 (2015).

160. *Id.*

161. *Id.*

162. *See id.* at 300–01.

163. Peter L. Strauss, *Private Standards Organizations and Public Law*, 22 WM. & MARY BILL RTS. J. 497, 500 (2013).

164. U.S. CONST. art. I, § 8, cl. 8.

165. *See* Barnes, *supra* note 13, at 104.

used without incurring liability. But why does information need this complex set of rules for it to be bought and sold in the market?

Information is a unique good.[166] Unlike a tangible normal product bought and sold in commerce, information is "non-rivalrous in consumption" and "non-excludable in production."[167] Non-rivalrous means that information is costless for multiple people to enjoy once it has been produced, and non-excludable means that it is extremely difficult and often prohibitively expensive for producers to prevent consumers from using information once it is produced.[168] Each characteristic represents the view of a good from either the consumer's (non-rivalrousness) or the producer's (non-excludability) perspective.

Because of these characteristics, the free market has little incentive to produce information. Furthermore, consumers will not pay for information, as they can obtain it elsewhere—because it is non-excludable—for free—because it is non-rivalrous.[169] Because of this, in a truly free market producers are unable to recoup their costs unless they can find a purchaser willing to pay a large amount for the information upfront. But what customer would do that? Very few. True to the tragic nature of the commons, people are incentivized to underinvest in public goods—here, information.[170] Thus, consumers will wait for someone else to make the purchase and then find the information for free, or at least cheaper. Indeed, a group of consumers could coalesce to purchase the information at a lower price per person, but the costs of finding such a group and negotiating an agreement are often prohibitive. Based on these problems alone, there is little incentive for producers to invest into creating information for the public, so a truly free market produces very little intellectual property or information.[171]

This is problematic because information is extremely valuable to society.[172] It makes people more productive through technology, happier through artistic expression, and more informed as voters. The correct amount of information that should be produced is highly debated among IP scholars, but for this

166. *See* Bracha & Syed, *supra* note 5, at 1848.
167. Barnes, *supra* note 13, at 102.
168. *Id.* at 98–99, 102–03.
169. *See* Bracha & Syed, *supra* note 5, at 1848–50; Barnes, *supra* note 13, at 107.
170. *See generally* Garrett Hardin, *The Tragedy of the Commons*, 162 SCI. 1243 (1968) (describing the tragedy of the commons).
171. Barnes, *supra* note 13, at 99.
172. *See id.*

Note's purpose, it suffices to say that a free market does not produce nearly enough.[173] Thus, a system is needed to produce information.[174]

The copyright and patent systems remedy this problem. They do so by balancing the adverse interests of the producer and the consumer. Each brings persuasive arguments to the table. The consumer argues that morally, information should be free because once it has been produced, sharing it imposes no new cost on anyone else while having tremendous value.[175] Drawing on a more fundamental economic principle, the consumer also posits that in a market, the cost of a good should approach the cost of producing one additional unit of that good.[176] In the case of information, the cost of producing an additional unit of a particular piece of information is zero because it is non-rivalrous.[177] Thus, the consumer wants free *access* to information.

On the other hand, the producer desires *incentives* to make such information. Morally, the producer could argue that since they made it, they should be able to control who has access to it. Additionally, at the very least they need to have enough exclusion power over her information to recoup the costs of creation.[178] The producer also calls upon a bedrock economic principle to support her argument: "[i]f [information was] excludable, private markets could efficiently allocate resources to their production."[179] In other words, they argue that if information were excludable, the private markets could allocate it in a way that would maximize the value bestowed upon society—i.e., distribute it efficiently.[180] And therefore producers should have the right to exclude consumers from the information the producer created.[181] Thus, the consumer wants free *access* to information, while the producer wants *incentives*, in the form of rights to exclude, to create information.

These two competing interests are what incentive-access theory balances. The balancing is tricky—"[i]ncreased incentives may create more beneficial information, while increased access allows more people to benefit from it"[182]—but the reward is clear: "[o]nce the law identifies the goods for which

173. *Id.* at 97 n.7, 99.
174. *See id.*
175. *Id.* at 97–98, 106.
176. *Id.* at 100.
177. *Id.* at 106.
178. *Id.* at 97.
179. *Id.* at 100.
180. *See, e.g.*, Peter S. Menell, *Tailoring Legal Protection for Computer Software*, 39 STAN. L. REV. 1329, 1336 (1987).
181. *See* Barnes, *supra* note 13, at 116.
182. *Id.* at 121.

people must pay, who must pay, and for how long they must pay, markets do their work."[183] Determining the appropriate balance can be accomplished through the "Net Benefit Principle,"[184] which submits that "[a]n increase in exclusive rights to intellectual property is justified only when the value of increased creative activity resulting from increased incentives is greater than the value of the benefits lost from reduced access" and that "[a]n increase in access to intellectual property is justified only when the value of the benefits resulting from increased access is greater than the value of decreased creative activity resulting from decreased incentives."[185] Utilizing some version of this principle, Congress and the courts have crafted an extremely complex body of copyright law that is meant to reconcile the two competing interests, access and incentives, in a way that is most beneficial to society.[186] This is why copyright law exists, and it justifies copyright's mechanics. Economic theory, however, is not the only policy at play here.

2. Democracy

As a basic premise, democratic values should underlie any discussion of actions taken by a democratic government. But what exactly does it mean to be a democracy?

As theorized, a democratic government's power flows directly from the people.[187] For the government only has power because the people consent to be governed.[188] And "the laws are made by the same people to whom they apply"; there is no exempt upper class.[189] But even more essentially, a democratic government is one that normatively values autonomy and self-determination and allows for people to engage in the practice of self-determination as a whole.[190] "Self-government is about the authorship of decisions, not the making of decisions."[191] Thus, the hallmark of democratic

183. *Id.* at 110.
184. *Id.* at 122.
185. *Id.*
186. *Id.* at 97.
187. Brief of the R Street Inst. et al. as *Amici Curiae* in Support of Respondent's Acquiescence in the Petition at 12, Georgia v. Public.Resource.Org, Inc., 140 S. Ct. 1498 (2020) (No. 18-1150) ("It is fundamental to the constitution (and Constitution) of this country that sovereignty derives from 'we the people.'" (quoting U.S. CONST. pmbl.)).
188. *Id.*; JOHN LOCKE, TWO TREATISES OF GOVERNMENT §§ 95-99 (1690).
189. Post, *supra* note 17, at 25; *See also* Thomas Paine, *Common Sense*, in Nelson F. Adkins, ed, *Common Sense and Other Political Writings* 3, 32 (Liberal Arts, 1953).
190. *See id.* at 25–26.
191. *Id.* at 26.

governance is one where the people truly feel engaged in the process of governing themselves—in the process of authoring decisions.[192]

This is not the same thing as majoritarian rule.[193] Take, for example, a country that votes every morning on what to eat for dinner, yet bans public discussion.[194] Assuming the people do not have homogeneous tastes, every day there will be some people in the minority who must eat a dinner other than the one they selected.[195] The people in the minority are not engaged in the process of governing themselves because there was no public discourse, so they have had no opportunity to shape the public opinion and thereby affect the results of the daily dinner election.[196] Rather, they would feel that they are subject to the whims of the majority.[197]

Returning to the dinner example above, assume that the daily dinner election remains, but now public discourse is allowed. Dissenters can now voice their opinions on the mashed potatoes and green beans. From this, some minds will be instantly changed. Others will have rebuttals. "No, the mashed potatoes are better than the green beans because the green beans always have too much salt." Despite the discourse, some faction must eat a dinner they did not choose or desire. However, in this case, consent to be governed is not defeated because those in the minority feel they are engaged in the process of governing themselves by participating in the public debate.[198] Thus, it is of the utmost importance that individuals who disagree with the outcome of a government decision can identify with the process of self-government and still feel they are governing themselves, thereby giving meaningful consent to be governed.[199]

Democratic governance occurs in response to public opinion. This must be, because for a people to be self-governed, government must be responsive to the ideas of the people.[200] However, since the people rarely agree on one thing, the government must be responsive to the public opinion, the opinion that most accurately represents the singular will of the people.[201] Thus, because it is essential that minority political factions feel they are engaged in the process of governing themselves, it becomes essential that individuals feel that they are

192. *Id. See*, Locke, *supra* note 188 at § 95.
193. Post, *supra* note 17, at 25.
194. *See id.* at 26.
195. *See id.* at 26–27.
196. *See id.*
197. *See id.*
198. *See*, Locke, *supra* note 188 at §§ 96–97.
199. *See* Post, *supra* note 17, at 26–27, 29.
200. *See id.* at 29.
201. *See id.*

engaged in the process of forming the public opinion.[202] If this were not the case, the idea of consent to be governed would be defeated.[203]

The public opinion begins to form when individuals express their independent views.[204] From this, a dialogue is formed, and the public opinion is born.[205] Since it is vital that individuals participate in the formation of the public opinion, it is vital that individuals are able to express themselves freely.[206] This is a major reason the First Amendment exists—to protect the individual right of expression in order for individuals to engage in the process of forming the public opinion.[207]

A democratic government must therefore value individual self-determination—the ability to decide for yourself what your values are and what you want to do—highly.[208] In order for individuals to meaningfully participate in the formation of the public opinion and thereby identify with the process of self-government, they must first have the opportunity to reflectively and meaningfully determine their own opinion. Otherwise, participation in democratic processes has only token value, and consent to be governed is meaningless.

In sum, a democratic government is one that values the ability of individuals to be autonomous so that they can meaningfully engage in the process of forming the public opinion and thus truly be self-governed.[209]

A postulate of this definition is that a democratic government cannot systematically restrict the ability to engage in democratic processes like voting or the formation of public opinion.[210] When this happens (and it does in the real world),[211] there is a minority of similarly situated individuals that is not engaged in the process of governing themselves, thereby avoiding any consent to be governed. Of course, as a practical matter, there are individual instances

---

202. *See id.* at 27–28.

203. *See id.*

204. *See id.* at 29.

205. *See, e.g.*, JOHN STUART MILL, ON LIBERTY, 50 (1859) (describing how a "collision of adverse opinions" is how society finds "truth").

206. *See* Post, *supra* note 17, 29–30.

207. *Id.* at 27–28.

208. *See id.* at 27.

209. *Id.* at 25–26.

210. *See id.* at 33 ("Democracy requires only that inequities that undermine democratic legitimacy be ameliorated. It does not require this for reasons of fairness or distributive justice, or because of any philosophic commitments that stand outside of democratic debate and decision making, but simply because such inequities undermine democratic legitimacy.").

211. *See, e.g.*, *Block the Vote: How Politicians are Trying to Block Voters from the Ballot Box.*, ACLU (Aug. 17, 2021), https://www.aclu.org/news/civil-liberties/block-the-vote-voter-suppression-in-2020.

where someone's ability to engage in democratic self-government may be restricted, but this must be because it is simply impracticable to ensure that every individual has the exact same rights to engage in democratic processes.[212] What is clear, however, is that it is undemocratic for there to be a minority (or majority) that is unable to meaningfully engage in democratic processes. This flows directly from the very definition of democratic self-government.[213]

Having discussed the main policies at play, this Note now moves into a brief discussion of two of the metrics used to determine whether the copyright system's incentive-access framework is balanced appropriately.

C. METRICS OF THE COPYRIGHT SYSTEM

1. *The Democratic Theory of Copyright*

The idea of seeing copyright as a means to perpetuate democratic values states that "copyright . . . [should] enhance the democratic character of civil society."[214] Its metric is therefore whether the democratic culture of civil society is being enhanced by granting a copyright. This does not mean that every work should be free because there would be some deadweight loss. The democratic theory recognizes that most works would not be created absent artificial scarcity.[215] Thus, the theory recognizes that copyright is something that is in, but not of, the market.[216]

From this theory, at the very least, it is clear that it is proper to consider democratic values when evaluating the copyright system. However, this is not the most prominent theory.

2. *The Economic Theory of Copyright*

Classic economic theory is another metric by which copyright's efficacy is measured.[217] This theory has been prevalent for some time and is taken seriously by courts and scholars alike. Economists seek to maximize wealth and increase allocative efficiency.[218] In other words, these theorists are primarily concerned with creating the most amount of dollars—maximizing wealth—and putting works to their highest and best use—increasing allocative efficiency.

---

212. *See*, Locke, *supra* note 188 at § 98.

213. Post, *supra* note 17, at 32.

214. Neil Weinstock Netanel, *Copyright and a Democratic Civil Society*, 106 YALE L.J. 283, 288 (1996).

215. *See id.*

216. *Id.*

217. *See generally* William M. Landes & Richard A. Posner, *An Economic Analysis of Copyright Law*, 18 J. LEGAL STUD. 325 (1989).

218. *See* Barnes, *supra* note 13, at 111–12, 122.

This theory has the benefit of being seemingly easy to apply. Good policies should maximize wealth at a low cost to society. However, in practice this is extremely difficult to do, especially ex ante. Even ex post, "the data [can be] maddeningly inconclusive."[219] Though difficult to apply, economic theory is thus also appropriate to consider when evaluating the copyright system.

Having reviewed the policies at play and common metrics of the copyright system's efficacy, this Note now moves into answering the question at hand by showing how the policies interact. When does copyright impermissibly restrict access to works whose copyrights are held by the government or to works whose copyrights are held privately but are nevertheless utilized by the government? The answer is: when there are no copyright incentives needed to produce the work or when it would be undemocratic to restrict access using copyright.

D. IDENTIFYING THE POLICIES AT PLAY AND HOW THEY INTERACT

This Note's answer is justified by three necessary principles. First, government works and government-utilized private works that would be produced regardless of any copyright incentive should not have their access restricted by copyright. This is because no incentive is needed and the incentive-access framework does not apply. Second, economics submits that allowing copyright in certain works, even works that utilized and created by the government maximizes efficiency. Third, democratic principles dictate that access to these works should not be restricted when it is undemocratic to do so.

For the sake of brevity, going forward this Note will refer to works that are copyrighted by the government or copyrighted by a private entity and utilized by the government as CGCPUG works (a loose abbreviation for Copyrighted by the Government or Copyrighted by a Private entity and Utilized by the Government).

1. Step One: When the Incentive-Access Framework Does Not Apply

First, access to CGCPUG works should not be restricted by copyright when the works do not fit into the incentive-access framework.

As discussed above a main objective of copyright law is to incentivize works that would not otherwise be made. By incentivizing these works, society ultimately benefits. But this benefit comes at the cost of access for a limited time.[220] The House Report accompanying the Copyright Act noted as much:

219. ROBERT P. MERGES, JUSTIFYING INTELLECTUAL PROPERTY 3 (2011).
220. *See* 17 U.S.C. § 302 (limiting copyright duration).

> Under the U.S. Constitution, the primary objective of copyright law is not to reward the author, but rather to secure for the public the benefits derived from the authors' labors. By giving authors an incentive to create, the public benefits in two ways: when the original expression is created and . . . when the limited term . . . expires and the creation is added to the public domain.[221]

However, when no incentive is needed for a work to be created, restricted access is without justification. Thus, CGCPUG works that need no incentive for creation should not have their access restricted via copyright.

2. Step Two: Maximizing Efficiency

Second, works that would not be created without copyright are extremely valuable to society. While they impose an access cost, they also fulfill a consumer need and create surplus wealth.[222] This is an extremely significant policy justification for copyright to many judges and scholars.[223] For example, in *Georgia v. PRO* the government was able to create first-class annotations that were then sold at economy rates.[224] By economic standards, society was made better because wealth was created when these annotations were sold at a low price. This could not have been done without copyright. Thus, CGCPUG works that are incentivized by copyright should be allowed to receive copyright.

3. Step Three: Democratic Principles Save the Day

Third, notwithstanding that a CGCPUG work needs incentives for creation, access to that work should not be restricted via copyright when it would be undemocratic to do so.

Democracy is of utmost importance in the United States. This is grounded in the First Amendment and the Supreme Court jurisprudence that follows, as described at length in Professor Bhagwat's article, *The Democratic First Amendment*.[225] In it, he details how each non-religious clause was intended to support democratic self-governance.[226] This has not been lost upon the Supreme Court's First Amendment jurisprudence; in fact, it is something that

---

221. H.R. REP. NO. 100-609, at 17 (1988).

222. *See, e.g.*, Landes & Posner, *supra* note 217, at 343.

223. *See, e.g.*, Georgia v. Public.Resource.Org, Inc., 140 S. Ct. 1498, 1517 (2020) (Thomas, J., dissenting).

224. *Id.* at 1513 (pointing out that the OCGA costs $404, while comparable privately-produced annotations cost $2,750).

225. *See generally* Ashutosh Bhagwat, *The Democratic First Amendment*, 110 NW. U. L. REV. 1097, 1111 (2016).

226. *Id.* ("What [the five clauses] have in common, however, is that each of the rights has as its primary goal the advancement of democratic self-governance.").

has been specifically articulated. For example, in *Walker v. Texas Division, Sons of Confederate Veterans*, the Court stated that "the Free Speech Clause helps produce informed opinions among members of the public, who are then able to influence the choices of a government."[227] Indeed, Professor Bhagwat recognizes that "[o]ver the past several decades, the Supreme Court has repeatedly taken the position that the primary—albeit not necessarily the only—reason why the First Amendment protects freedom of speech is to advance democratic self-governance."[228]

Since democracy is so important to this country's laws and values, it follows that the democratic government should not take actions opposed to democratic values. Unfortunately, however, allowing copyright in CGCPUG works can create an undemocratic systematic tilt in the citizenry's ability to participate in forming public opinion. To show this, it is necessary to be clear how CGCPUG works can relate to the formation of public opinion.

For individuals to participate in forming public opinion as it relates to governance, they first must have the ability to know how the government is governing, so that they can determine how their values align. The individuals can then determine and voice their opinion, thereby shaping public opinion. In this way, individuals can feel that they are engaged in the process of self-governance.

CGCPUG works can provide rich and unique insight into how the government is governing because they provide first-hand evidence of government actions (e.g., binding criminal law). Whether the government is promulgating a law, explaining its own statutes in annotations, or adopting a technical standard, it is explaining to the public how it is governing. As such, CGCPUG works can have high democratic value.

Given this, it can be undemocratic to restrict access to certain CGCPUG works via copyright. Since copyright creates deadweight loss, when copyright is granted in CGCPUG works that have high democratic value, like official annotations, a minority is created that is interested in the work but cannot access it. When this happens, there are two classes of citizens: one who can afford the CGCPUG work and one who cannot. The first class is better informed and able to participate in the formation of public opinion, while the second class is disadvantaged. This is undemocratic because it creates a systematic skewing in the availability of participation in democratic processes.

The final question is: how big must a group of citizens be to implicate democratic concerns? There will likely always be one individual who is unable

227. Walker v. Tex. Div., Sons of Confederate Veterans, Inc., 576 U.S. 200, 207 (2015).
228. Bhagwat, *supra* note 225, at 1098.

to equally participate in democratic processes. This does not raise any concerns. Two individuals who are unable to equally participate in democratic processes will also not raise concerns. The transformation from individuals to a group comes when there is a set of similarly situated individuals, a cognizable minority. Thus, restricting access is undemocratic when there is a cognizable minority that cannot equally participate in forming the public opinion.

Indeed, defining precisely when a cognizable minority exists is a difficult task and perhaps is almost always arbitrary. But this difficulty does not overturn the conclusion that it is undemocratic for a set of similarly situated individuals to have unequal access to democratic processes. Moreover, the line-drawing should be done by those better situated to decide this issue: Congress and the courts.

4. *Summing the Policy Arguments Up*

The government is well positioned to maximize efficiency by creating or utilizing copyrighted works. But copyright should not restrict access to CGCPUG works where copyright does not serve as an incentive to produce such works or where it would be undemocratic to do so—when it would restrict a cognizable minority's ability to equally engage in a democratic process.

E. ANALYSIS USING THE POLICY FRAMEWORK

To illustrate the need for further refinement in this approach, this Note will now use it to analyze four common types of works utilized by the government: core governing works, the OCGA, model housing codes, and engineering standards. Specifically, this Note will show that it may not always be undemocratic to restrict access to certain CGCPUG works.

It is worth noting that there is no distinction between whether a government body adopts one of the types of works above by reference, copies it in its entirety or manages its creation and retains the copyright. In each case, copyright stands as a barrier to access.

1. *Core Governing Works*

Certain works made by the government do not fit into the incentive-access framework. On the incentive side, the government does not need incentives to create them. The government's job is to rule: to make, interpret, and enforce the law.[229] In carrying out its job, the government is going to produce certain works to accomplish its task of ruling, regardless of whether it can obtain

229. Brief *Amici Curiae* of the Am. C.L. Union et al. in Support of Respondent at 7, *Georgia*, 140 S. Ct. 1498 (No. 18-1150).

copyright protection for them or not. This is also partly because the government has already been paid to create these works via taxes and thus does not need any further incentives.[230] These core governing works include works such as executive orders, constitutions, statutes, and opinions with their headnotes. Because core governing works do not need copyright to incentivize their creation, it is improper to have their access restricted via copyright under the first part of this Note's test. Accordingly, it is unnecessary to reach the democratic question.

2. Model Housing Codes

a) Is Copyright Needed to Incentivize Creation?

A thin form of copyright is likely needed to incentivize creation. Professor Goldstein notes:

> [I]t is difficult to imagine an area of creative endeavor in which the copyright incentive is needed less. Trade organizations have powerful reasons stemming from industry standardization, quality control and self-regulation to produce these codes; it is unlikely that without copyright they will cease producing them. A rule that confines protection to the literal text of these codes, and forecloses protection for their enacted versions, is consistent with the limited need for incentives in this field.[231]

Trade organizations that create model housing codes have strong incentives to standardize the building process.[232] However, there are costs associated with production that businesses may need to recoup. Thus, it is entirely appropriate that they receive protection that is "confine[d] . . . to the literal text of these codes."[233] However, what this statement overlooks is that the amount of copyright protection corresponds with the amount of creative input. The closer something is to a fact, the less protection it receives. If a copyrighted model code became binding and thus the law, it would become very close to a fact and would receive little protection. Thus, these works likely already receive the thin form of copyright needed to incentivize production.

b) Would it be Undemocratic to Restrict Access Via Copyright?

It would, however, be undemocratic to restrict access to an adopted model housing code via copyright. A cognizable minority is easily imaginable: parents who want to ensure that the future homes they buy will be safe for their

230. *See, e.g.*, Banks v. Manchester, 128 U.S. 244, 253 (1888).
231. 1 PAUL GOLDSTEIN, GOLDSTEIN ON COPYRIGHT § 2.5.2, at 2:59 (3d ed. 2020).
232. *See id.*
233. *Id.*

children. These parents would wish to engage in shaping the rules but cannot afford to buy the model housing codes to understand the current law. Thus, it would be impermissible to restrict access to model housing codes using copyright.[234]

c) Conclusions

In sum, though a model housing code likely needs some form of copyright protection to be produced privately, its access should not be restricted because it would be undemocratic to do so. In other words, because of democratic concerns, governments should not utilize copyrighted model housing codes; they should bear the burden of producing the codes themselves.

3. Engineering Standards

a) Is Copyright Needed to Incentivize Creation?

The literature is sharply divided on whether copyright is needed to produce engineering standards.[235] Those who are against copyright bring strong arguments to the table.[236] For example, Professor Samuelson notes, among other things, that:

> First, SSOs [standard setting organizations] generally have ample incentives to develop standards for use by professionals in their fields. It is simply not credible to claim that organizations like the AMA [American Medical Association] and ADA [American Dental Association] would stop developing standard nomenclature without copyright protection. The fields they serve need these standards for effective communication with other health care providers, insurers, and government agencies.
>
> Second, SSOs generally do not actually develop the standards in which they claim copyrights. Rather, they typically rely upon volunteer service by experts in the field to develop standards and require volunteers to assign any copyright interests to the SSOs. The community development of a standard is a reason to treat the standard itself as a shared resource.
>
> Third, SSOs generally use the revenues they derive from selling or licensing the standards to subsidize other activities of their

234. It is worth noting that the Fifth Circuit in *Veeck v. Southern Building Code Congress International, Inc.* held that a model housing code should not be copyrighted. 293 F.3d 791, 793 (5th Cir. 2002) (en banc).

235. *See* Samuelson, *supra* note 157 (arguing against copyright in standards); Emily S. Bremer, *Incorporation by Reference in an Open-Government Age*, 36 HARV. J.L. & PUB. POL'Y 131, 136–37 (2013) (arguing that abrogating copyright in standards destroys value so a different route should be pursued).

236. *See, e.g.*, Samuelson, *supra* note 157, at 193.

> organizations, rather than to recoup investments in making the standards. Even without copyright in the standards, SSOs can derive revenues from sales of print materials embodying the standard and value-added products or services.[237]

However, there are rebuttals to each point. Some SSOs, like the American Society for Testing and Materials (ASTM), develop some of their own standards (countering Professor Samuelson's second assertion).[238] Since these SSOs are organizations with dedicated employees, it is doubtful that they have the same incentives as a business in the industry to create engineering standards (countering point 1). Indeed, they take input from industry professionals, but there are still costs associated with gathering the information, organizing the information, holding committee meetings, and keeping the lights on (countering Professor Samuelson's third assertion).[239] All these tasks, which cost real world dollars, are necessary to produce a standard in some cases. To further rebut the third point, it is unlikely that an SSO that sells the official version of a standard will retain their revenues should copyright be removed. Why would an engineering firm pay for something when it can find an equal quality replacement for free or for cheaper? It follows that some SSOs need the ability to exclude to recoup their development costs.

A case study illustrates that SSOs likely need protection to incentivize creation.[240] In 2012, Congress mandated that government agencies ensure that free access to all of the codes they adopted by reference is provided.[241] Since government agencies cannot infringe on copyright, negotiations ensued.[242] The Pipeline and Hazardous Materials Safety Administration (PHMSA) of the Department of Transportation (DOT) was not successful in obtaining free access for every incorporated code.[243] Likely, many SSOs simply were not willing to lose the revenue.[244]

It is likely that SSOs did not provide free access because they needed the revenue, not because they were engaged in monopoly pricing.[245] As the same study showed, SSOs routinely charge equal or more for current editions of a

---

237. *Id.* at 222 (footnotes omitted).

238. *Standards Development in ASTM*, ASTM INT'L, https://www.astm.org/studentmember/StandardsProcess.html.

239. *See id.*

240. *See* Bremer, *supra* note 159, at 281 (noting that the PHMSA was unable to negotiate free access to all of its incorporated standards).

241. *Id.*

242. *See id.* at 282.

243. *Id.*

244. *See id.*

245. *See id.*

standard than for an older edition of the same standard that is adopted by reference.[246] This rebuts the contention that SSOs are engaged in monopoly pricing.[247] If an SSO was engaged in monopoly pricing, would it not charge more for a binding version than a newer version?

Thus, on balance, it is foreseeable that standards need some form of protection to incentivize their creation.

b) Would it be Undemocratic to Restrict Access via Copyright?

It would probably not be undemocratic, generally speaking, to restrict access to engineering standards via copyright, but this is a nuanced question. Take, for example, a standard regulating the thickness of pressure vessels. There are two groups imaginable. First, consider an engineering company. Suppose they want access to the work to engage in public discussion. This is certainly a valid democratic goal. But they likely already purchase the work as a cost of doing business. So, in a sense, they already have (or at least should have) access. This raises the question as to whether it is appropriate to have a standard as a cost of doing business. A complete discussion of this question is beyond the scope of this Note, but it suffices to say that allowing standards to remain a cost of doing business keeps the cost of regulation, at least in part, on those who are profiting from the industry and off the citizenry at large, which is normatively desirable.

Second, imagine a group of citizens who have technical knowledge and are interested in the standards for democratic purposes. It is hard to imagine a group such as this that does not already have access from being in the industry.

However, imagine a boiler explodes, killing dozens. It is then foreseeable that a group of people who have technical knowledge and are interested in viewing the standard for democratic purposes coalesces.

Thus, it is not clear that democratic concerns are always implicated by allowing copyright in a standard. This is a fact-heavy and time-sensitive inquiry. Therefore, allowing copyright in standards is not automatically undemocratic.

a) Conclusions

Engineering standards likely require some form of copyright protection. Unlike model housing codes, most engineering standards are much further removed from public interaction, so it is less likely that it would be undemocratic to restrict access to them with copyright. But as noted, there

246. *Id.* at 318.
247. *See id.*

may be instances where copyright should not be granted, and more refinement is thus needed as a rule is formed from this framework.

4. OCGA

a) Is Copyright Needed to Incentivize Creation?

Without a doubt, copyright is needed to incentivize creation of Georgia's official annotations. Without copyright protection, the state would not be able to grant Lexis the ability to be the sole producer of the annotations. Lexis would not produce these annotations without the ability to exclude others, as others could obtain the annotations for free, depriving Georgia of the ability to produce the annotations at no cost to the taxpayers. Thus, copyright is needed to produce the OCGA.

b) Would it be Undemocratic to Award Copyright?

As mentioned before, annotations are very important to the democratic process. They explain how the laws will be applied, and they provide more intimate information on the actions of the government.[248] Given the content and breadth of the annotations, there will always be an interested minority in this work who cannot afford it. Thus, it is impermissible to restrict access to the OCGA through copyright.

c) Conclusions

Removing copyright in these works would completely destroy their value. However, a democratic government perpetuating this systematic tilt in the opportunity for people to participate in democratic processes would go against its very nature. Accordingly, copyright should not restrict access.

d) The Supreme Court's Decision

The average citizen will likely not be searching the internet for the official code or its annotations, but democratic principles do not always cater to the average citizen. What is clearly undemocratic is a situation where a cognizable minority of people who wish to better engage in democratic process cannot do so. It is quite imaginable that there is an interested group of similarly situated citizens who wish to use the OCGA to better inform themselves but cannot.

Thus, given the binary the Supreme Court was presented with in *Georgia v. PRO*, allowing copyright or barring it entirely, the Court made the correct decision. Though the rule it laid down was vague and hard to apply, it managed to use that rule to reach the result that promotes democracy.

248. *See* Georgia v. Public.Resource.Org, Inc., 140 S. Ct. 1498, 1504, 1512 (2020).

V. CONCLUSION

The next problem to be resolved by Congress is remedying impermissible restrictions on access. This is beyond the scope of this Note. But it is worth noting that there is a range of options beyond the Court's binary of copyright versus no copyright. For example, one could require free, read-only online access to a work.[249] Large companies would likely purchase the official versions of the work, complete with better searching tools and other functionalities for the online versions. Yet lay persons would still be able to access the work for democratic purposes.

The policy framework for the proper rule entails economic and democratic considerations that extend beyond a mere examination of century-old cases. First, CGCPUG works should not have their access restricted when they do not fit in the incentive-access framework. Access should not be restricted when there are no incentives needed to produce. Second, CGCPUG works should not have their access restricted when it would be undemocratic to do so. Democratic governments should not take undemocratic actions.

249. Bremer, *supra* note 235, at 153–83.

BOOKING.COM AND THE SHRINKING HYPERLINGUISTIC COMMONS

Rachel Paige Thompson†

I. INTRODUCTION

How much of the cultural lexicon should corporations control, especially as that lexicon moves online? In *United States Patent & Trademark Office v. Booking.com B.V.*,[1] the U.S. Supreme Court allowed a company to cordon off a large swath of both our shared linguistic commons and our hyperlinguistic commons. In *Booking.com*, the Supreme Court rejected the U.S. Patent and Trademark Office's (USPTO) per se rule that a generic term, such as "booking," plus a generic top-level domain name, such as ".com," necessarily yields an unprotectable generic mark.[2] The decision aligns with precedent, but in doing so, it ignores how new technology and new media operate outside the usual linguistic structures underlying American trademark law. It also ignores the protections already provided under contract law and the domain name system, affording a company double assurance that a generic word may belong to it.

In Part II, this Note begins by situating the purpose of trademark law—consumer protection—in the abstract functioning of the distinctiveness spectrum. When determining eligibility for trademark protection, trademark law classifies marks into several categories of distinctiveness,[3] and the incentive structure of trademark law becomes clear through an analysis of these categories. Creative terms should be easy for companies to protect because those terms do not deplete the linguistic commons, and it should be difficult, if not impossible, for companies to claim a linguistically valuable term for themselves. Commonplace terms that many competitors would want or even need to use in marketing their products and services, like "grocery" or

---

DOI: https://doi.org/10.15779/Z38T43J405.

† J.D. 2022, University of California, Berkeley, School of Law. Sincere thanks to Professor Talha Syed, my fellow students in the 2020 Law & Technology Writing Workshop at Berkeley Law, and the Berkeley Technology Law Journal editors.

1. 140 S. Ct. 2298 (2020). Interestingly, this was the first case in United States Supreme Court history where the Justices conducted oral argument remotely. Adam Liptak, *Supreme Court Hears First Arguments via Phone*, N.Y. TIMES (May 4, 2020), https://www.nytimes.com/2020/05/04/us/politics/supreme-court-coronavirus-call.html [https://perma.cc/UUQ4-L768].
2. 140 S. Ct. at 2301.
3. *See* Abercrombie & Fitch, Co. v. Hunting World, Inc., 537 F.2d 4, 9 (2d Cir. 1976).

"sandwich" or "shoes," should not fall under the control of any single company. To determine distinctiveness and thereby trademark eligibility, factfinders use the primary significance test,[4] but this test's imprecision allows wealthy companies to gain protection for highly useful words and phrases.

Part II moves on to explain domain names, the subject of contention in *Booking.com*, including the overarching framework and processes for managing them. How someone acquires a domain name, how that system interacts with existing trademarks, and how the market for domain names has matured and expanded—all of this groundwork helps show that domain names are unlike traditional word marks. These differences create inefficiencies and redundancies when the trademark system is overlaid on the domain name system.

With this background, this Note moves on in Part III to provide an overview of *Booking.com* as it moved through the USPTO and the court system. The confusion at every step about whether the BOOKING.COM mark was generic or descriptive is a prime example of the shortcomings inherent in the primary significance test. In applying this test, courts ask what the relevant consuming public thinks of a term that is being considered for trademark protection to determine whether a mark is distinctive and may receive trademark protection.[5] As this Note explains, the Supreme Court issued a predictable decision, holding that a domain name made up of generic parts could receive protection under trademark law.[6] But in so doing, the Court elided the more fundamental question of why the precedential primary significance test should apply here, to the new commercial context introduced by online services, in the first place. And even beyond *Booking.com* and domain names, the primary significance test has faced criticism for decades.[7] This case presented a missed opportunity for the Supreme Court to revisit the degree to which public perception matters, and how it can be determined accurately.

Part IV attempts to answer how an accurate distinctiveness determination should be made for marks including generic top-level domain names, always keeping consumer protection as the animating principle undergirding trademark law. In doing so, this Note builds on Justice Breyer's dissent,[8] as

4. *See* Bayer Co. v. United Drug Co., 272 F. 505, 509 (S.D.N.Y. 1921).

5. *See id.*

6. *Booking.com*, 140 S. Ct. at 2301.

7. *See, e.g.*, Ralph H. Folsom & Larry L. Teply, *Trademarked Generic Words*, 89 YALE L.J. 1323 (1980); John F. Coverdale, Note, *Trademarks and Generic Words: An Effect-on-Competition Test*, 51 U. CHI. L. REV. 868 (1984); Stephen R. Baird, Note, *Putting the Cart Before the Horse in Assessing Trademark Validity—Toward Redefining the Inherently Generic Term*, 14 J. CORP. L. 925, 927 (1989).

8. *Booking.com*, 140 S. Ct. at 2309–16 (Breyer, J., dissenting).

well as scholarly suggestions for how to adapt the distinctiveness spectrum.[9] Public perception is malleable, especially because it is measured by consumer surveys paid for by large companies with strong marketing and advertising power. Instead of focusing on consumer comprehension, this Note suggests that examining attorneys at the USPTO could ask how protecting the applied-for mark and its component terms would impact competitors acting in good faith. If others would have a viable fair use defense, then the mark should perhaps not enjoy protection in the first place. But this proposal could vitiate the current trademark system, returning it to the era where only inherently distinctive terms could receive protection.

Although replacing the focus on the consumer with a question about impact on competitors is desirable, this Note nonetheless offers an alternative and more palatable method for preventing the overprotection of domain names. Building off the work of new media theorists, this Note proposes a new category for factfinders to consider during the eligibility stages of a trademark application: executable matter. This category recognizes that parts of a mark might look like natural language and sound like natural language but, in reality, are coded elements that, when used for their executable purpose, should not receive trademark protection. In other words, ".com" operates more as code, appended onto other terms in order to trigger an operation within a system, than as communicative language. These are hyperlinguistic elements, and the domain name system regulates our hyperlinguistic commons.

This proposition, that executable elements of a trademark are ineligible for protection, leads to a discussion about how and why the *Booking.com* decision was wrong on even narrower grounds. It evaluated a domain name using the same framework factfinders use for words—but generic.com domains do not fit neatly into a framework developed around natural language. Arguably, the domain name system already provides greater protections than the trademark system. A generic.com domain owner is the only one who can own that domain, and nefarious lookalikes can be sued under unfair competition laws.[10] Trademarks, on the other hand, can have multiple meanings and connotations. Some terms have to be disclaimed in the registration process, and some of the protections are weak. Seeking principal register trademark status for a generic.com domain, then, speaks to the tendency toward monopolization of language. Part IV concludes by outlining exactly how the distinctiveness

9. *See supra* note 7.

10. *See* J. Thomas McCarthy, *Trademarks, Cybersquatters and Domain Names*, 10 DEPAUL J. ART TECH. & INTELL. PROP. L. 231, 231 (2000); *see also* Lisa Carroll, *A Better Way to Skin the Cat: Resolving Domain Name Disputes Using State Unfair Competition Law*, 18 IPL NEWSL. 1 (2000).

spectrum could be modified to exclude executable elements from trademark protection, either by examining attorneys at the USPTO or by Congress.

II. BACKGROUND

Booking.com was wrongly decided on two fronts. First, in uncritically following precedent, it moved modern trademark law even further away from its original consumer-protection purposes. Second, it overlayed an already-strained intellectual property framework on top of an existing infrastructure and market. This Part provides the background necessary for grounding this dual analysis, beginning with a discussion of the distinctiveness spectrum and moving on to an examination of the domain name system and its relevant statutes and actors.

A. PURPOSES OF TRADEMARK LAW AS EMBODIED BY THE DISTINCTIVENESS SPECTRUM

At its core, trademark law protects consumers.[11] It allows buyers to create associations with brands and products and to trust these associations. Producers, in turn, enjoy a monopoly over certain expressions that, unlike other intellectual property protections, can "last in perpetuity."[12] Restrictions on language and symbols are balanced with stamping out malicious deception and keeping consumer search costs low.[13] One foundational example of this balancing is the distinctiveness requirement for trademark eligibility, found in the Lanham Act:

11. *See* Glynn S. Lunney, *Trademark Monopolies*, 48 EMORY L.J. 367, 417 (1999) ("Originally, trademark law was justified on grounds of preventing consumer deception."); Mark A. Lemley, *The Modern Lanham Act and the Death of Common Sense*, 108 YALE L.J. 1687, 1695 (1999) ("We give protection to trademarks for one basic reason: to enable the public to identify easily a particular product from a particular source."). *But see* Mark P. McKenna, *The Normative Foundations of Trademark Law*, 82 NOTRE DAME L. REV. 1839, 1840–41 (2013) ("[T]rademark law was not traditionally intended to protect consumers. Instead, trademark law, like all unfair competition law, sought to protect producers from illegitimate diversions of their trade by competitors."); KONSTANTINOS KOMAITIS, THE CURRENT STATE OF DOMAIN NAME REGULATION: DOMAIN NAMES AS SECOND-CLASS CITIZENS IN A MARK-DOMINATED WORLD 25 (2010) ("A trademark serves multiple and multi-dimensional purposes that extent [sic] from consumer protection to market efficiency to the institution of values, such as responsibility and obligation to others.").

12. *See* Folsom & Teply, *supra* note 7, at 1358.

13. *See* Wayne F. Osoba, Note, *The Legislative Response to* Anti-Monopoly*: A Missed Opportunity to Clarify the Genericness Doctrine*, 1985 U. ILL. L. REV. 197, 198 ("The law . . . must balance the protection of trademarks against the public policy of free competition by preventing the monopolization of common terminology.").

> No trademark by which the goods of the applicant may be distinguished from the goods of others shall be refused registration on the principal register on account of its nature unless it—
>
> . . .
>
> (e) Consists of a mark which (1) when used on or in connection with the goods of the applicant is merely descriptive or deceptively misdescriptive of them . . . .
>
> (f) Except as expressly excluded in [certain] subsections . . . of this section, nothing in this chapter shall prevent the registration of a mark used by the applicant which has become distinctive of the applicant's goods in commerce.[14]

Distinctive marks, those that can distinguish goods in the marketplace, are presumed eligible for strong trademark protection; nondistinctive marks, those that merely describe products, may acquire distinctiveness and become eligible for weak trademark protection.[15]

In *Abercrombie & Fitch, Co. v. Hunting World, Inc.*,[16] Judge Friendly provided a seminal explanation of the four categories of distinctiveness and their relationship to trademark eligibility and protection. Inherently distinctive[17] marks, divided into arbitrary/fanciful and suggestive,[18] occupy one end of the

---

14. 15 U.S.C. § 1052.

15. It is "conventional wisdom" that "stronger marks always receive broader protection." Barton Beebe & C. Scott Hemphill, *The Scope of Strong Marks: Should Trademark Law Protect the Strong More Than the Weak?*, 92 N.Y.U. L. REV. 1339, 1344–45 (2017). In trademark law, the scope of protection speaks to the "range of marks that if employed by a defendant . . . would be held to be infringing." *Id.* at 1347. An inherently distinctive mark receives stronger and broader protection than a mark that has acquired distinctiveness. *See* U.S. PAT. & TRADEMARK OFF., TRADEMARK MANUAL OF EXAMINING PROCEDURE § 1207.01(b)(ix) (Oct. 2018), http://tmep.uspto.gov/ [https://perma.cc/5NLP-A6KG] ("[M]erely descriptive and weak designations may be entitled to a narrower scope of protection than an entirely arbitrary or coined word.").

16. 537 F.2d 4 (2d Cir. 1976).

17. Inherent distinctiveness could simply be presumed secondary meaning by another name. Under this interpretation, arbitrary and suggestive marks are not in and of themselves distinct. Rather, they can be presumed to carry secondary meaning, and this presumption keeps administrative costs low. *See* Robert G. Bone, *Enforcement Costs and Trademark Puzzles*, 90 VA. L. REV. 2099, 2130 (2004).

18. It is not self-evident that suggestive marks should be presumed distinctive. Some scholars have argued that suggestive marks are more akin to descriptive than arbitrary marks and that applicants should be required to show secondary meaning when attempting to register them. *See* Jake Linford, *The False Dichotomy Between Suggestive and Descriptive Trademarks*, 76 OHIO ST. L.J. 1367, 1373–74 (2015) ("[T]he inherent distinctiveness line in trademark law should be redrawn, and a suggestive mark should be protected only if the mark owner can provide evidence of source significance.").

spectrum. These marks are thought to require an imaginative leap on the part of the consumer. APPLE is an arbitrary mark when used for technology products because it carries little to no immediate connotation; it does not tell consumers anything more than which company produced the item at hand. COPPERTONE, when used for sunscreen, and JAGUAR, when used for sports cars, are examples of suggestive marks. They trade on the consumer connecting some quality of the thing named with some quality of the product or service sold. Both arbitrary/fanciful and suggestive marks are presumed distinctive and are automatically protected because it is unlikely that other producers will be constricted by such protection. Whether or not "jaguar" is trademarked, an automobile manufacturer can advertise its cars effectively and attractively, and the boundaries of the protection are defined by the class of products or services for which it is trademarked. Although another automaker could not use "jaguar," a footwear designer conceivably could.

On the other end of the spectrum, there are descriptive and generic marks. As the name implies, descriptive marks describe the product or service, like HOLIDAY INN for a hotel or QUIK-DRI for a hairdryer.[19] In contrast, generic marks are described as "the genus of which the particular product is a species."[20] These terms are the general name of the product, like "computer store" or "brick oven pizza." At common law, neither category was eligible for trademark protection, but the Lanham Act carved out an exception for merely descriptive marks that gain some secondary meaning in commerce.[21] Secondary meaning refers to an additional association with a mark that consumers develop based on encounters with the mark in the market.

While descriptive marks may now acquire distinctiveness and become eligible for weak protection, as opposed to the strong protection afforded

19. Factfinders often struggle with determining the line between suggestive and descriptive marks. *See* Edward J. Heath & John M. Tanski, *Drawing the Line Between Descriptive and Suggestive Trademarks*, 12 COM. & BUS. LITIG. 11, 11 (2010) ("In theory, the distinction between 'descriptive' and 'suggestive' marks is clear. In practice, however, the line is a blurry one.").

20. *Hunting World*, 537 F.2d at 9. Courts began affording trademark protection to certain descriptive marks which had acquired secondary meaning around 1938, and this exception was later codified in the Lanham Act. Elizabeth A. Horky, *1-800-I-AM-VAIN: Should Telephone Mnemonics Be Protected as Trademarks?*, 3 J. INTELL. PROP. L. 213, 226 n.61 (1995). In 1989, Congress amended the Lanham Act to change "common descriptive name" to "generic name." 2 J. THOMAS MCCARTHY, MCCARTHY ON TRADEMARKS AND UNFAIR COMPETITION § 12:21 (5th ed. 2017). Before and even after this change, courts were often imprecise in their use of "descriptive" and "generic." *See id.*

21. *Hunting World*, 537 F.2d at 9.

inherently distinctive marks,[22] generic marks may never be trademarked.[23] Indeed, if arbitrary or fanciful marks become generic, they are stripped of trademark protection—a process referred to as genericide. This categorical prohibition is rooted in the consumer-centric purposes of trademark law: buyers benefit when sellers may freely use common language to describe and explain their products and services.[24]

In determining where a mark falls on the spectrum, USPTO examining attorneys consider evidence such as "the dictionary, puffery, competitors' need, competitors' use, and imagination tests."[25] They also look to rhetorical devices like "incongruity, musicality, or double entendre . . . as a proxy for distinctiveness."[26] These tests have been described as "ill-defined and haphazardly-applied," leading factfinders to be accused of "overprotecting" descriptive marks.[27] Sometimes, like in *Booking.com*,[28] consumer survey evidence will be provided, although this can be cost-prohibitive because the applicant covers the price of creating and executing the survey.[29]

Specifically, to determine whether a mark is or has become generic, factfinders apply the primary significance test.[30] As Judge Learned Hand put it

---

22. *See* Beebe & Hemphill, *supra* note 15, at 1344–45.

23. *Hunting World*, 537 F.2d at 9 ("[N]o matter how much money and effort the user of a generic term has poured into promoting the sale of its merchandise and what success it has achieved in securing public identification, it cannot deprive competing manufacturers of the product of the right to call an article by its name."). *But see* Rebecca Tushnet, *Registering Disagreement: Registration in Modern American Trademark Law*, 130 HARV. L. REV. 867, 878 (2017) ("There is no ontologically 'nontrademarkable' class of symbols. Anything that is capable of distinguishing the source of goods or services—known as being 'distinctive'—can serve as a mark." (internal citations omitted)).

24. *See* Hans Zeisel, *The Surveys that Broke Monopoly*, 50 U. CHI. L. REV. 896, 897 (1983) ("Giving a competitor trademark protection for whatever word he chooses . . . might impoverish the language of commerce."); Osoba, *supra* note 13, at 198 ("The law . . . must balance the protection of trademarks against the public policy of free competition by preventing the monopolization of common terminology.").

25. Alexandra J. Roberts, *How to Do Things with Word Marks: A Speech-Act Theory of Distinctiveness*, 65 ALA. L. REV. 1035, 1038 (2014).

26. *Id.*

27. *Id.* at 1039, 1085–86.

28. U.S. Pat. & Trademark Off. v. Booking.com B.V., 140 S. Ct. 2298, 2313–14 (2020) (Breyer, J., dissenting).

29. Robert C. Bird & Joel H. Steckel, *The Role of Consumer Surveys in Trademark Infringement: Empirical Evidence from the Federal Courts*, 14 U. PA. J. BUS. L. 1013, 1017 (2012) ("Consumer surveys are expensive to create and time consuming to administer.").

30. U.S. PAT. & TRADEMARK OFF., *supra* note 15, § 1209.01(c)(i). Following *Booking.com*, the USPTO issued special instructions for generic.com domain names. U.S. PAT. & TRADEMARK OFF., EXAMINATION GUIDE 3-20: GENERIC.COM TERMS AFTER *USPTO V. BOOKING.COM* (Oct. 2020), https://www.uspto.gov/sites/default/files/documents/TM-ExamGuide-3-20.pdf [https://perma.cc/9JEB-8L73].

a century ago, "What do the buyers understand by the word for whose use the parties are contending?"[31] Phrased another way, in the "minds of the consuming public," does the primary significance of the term correspond to the product or the producer?[32] If the mark primarily signifies the product, then the mark is generic and unprotectable. In 1979, the Ninth Circuit infamously deviated from this test, resulting in an amendment to the Lanham Act, which codified the primary significance test as to genericide.[33]

The primary significance test focuses mainly on public perception of a mark, but its application is muddy. How is public perception determined? Who is the relevant public? How does the stark binary between genus and species map onto the realities of language? The shortcomings of the test have been exposed by contradictory decisions: some circuits find a term to be generic, while others find the same term to be descriptive.[34] To remedy this, scholars have long called for a new test, with the form of their suggestions depending on their interpretations of the purposes and policies underlying trademark law.[35]

B. DOMAIN NAMES

Booking.com suffers both theoretically and functionally. It miscarries the foundational goals of trademark law, and in doing so it overlooks the protections provided by a preexisting system for allocating space in the virtual marketplace. To understand the latter, it is essential to explain some of the structure and economics of domain names, as well as the relevant statutes and actors.

In the early years of the internet, as more sites proliferated and the online landscape grew more complex, the domain name system (DNS) was developed as a scaling solution, allowing "hierarchical host names" (e.g.,

31. Bayer Co. v. United Drug Co., 272 F. 505, 509 (S.D.N.Y. 1921).

32. Kellogg Co. v. Nat'l Biscuit Co., 305 U.S. 111, 111–12, 118 (1938).

33. *See* Osoba, *supra* note 13, at 199. In *Anti-Monopoly*, the Ninth Circuit used the consumer motivation test rather than the primary significance test. The motivation test asked whether consumers provided product-related reasons or source-related reasons for buying the product. The Lanham Act does not codify the primary significance test as to determining whether an as yet unprotected mark should be categorized as generic or descriptive in the first instance. *See* 15 U.S.C. § 1052 (lacking the primary significance test in distinguishing between descriptive and generic marks). Courts, and even the Trademark Manual of Examining Procedure, nonetheless lift the test from the genericide section and apply it to the initial determination of genericness for a trademark application. *See Booking.com*, 140 S. Ct. 2298, 2304 n.3 (listing cases and a section from the Trademark Manual of Examining Procedure that apply the primary significance test from genericide to determine genericness in the first instance).

34. *See* Folsom & Teply, *supra* note 7, at 1351 (explaining that the Seventh Circuit found "alo" to be generic for cosmetics, while the Fifth Circuit found it to be descriptive).

35. *See supra* note 7.

"www.google.com") to lead users directly to an internet address.[36] This replaced a number-based system, similar to phone numbers, which had required users to memorize IP addresses (e.g., "128.0.0.1").[37] This simple switch implicated trademark law and its semiotics.[38]

Today, domain names consist of a top-level domain (.com) and a second-level domain (booking).[39] The Internet Corporation for Assigned Names and Numbers (ICANN), founded in 1998, manages the "allocation and assignment of names in the root zone of the Domain Name System . . . and coordinates the development and implementation of policies concerning the registration of second-level domain names in generic top-level domains."[40] ICANN maintains agreements with the organizations that operate current top-level domains, and it coordinates the release of new top-level domains.[41]

To register a domain name, a user first visits a domain name registrar, like GoDaddy.com or Google Domains, and searches for an available domain name.[42] Not all registrars are licensed to sell all top-level domains, especially country-specific endings like .in for India or .be for Belgium.[43] Some registrars offer top-level domains like .buzz and .health, but .com remains the most used top-level domain on the internet.[44] Once the user finds a domain name that is

---

36. BARRY M. LEINER, VINTON G. CERF, DAVID D. CLARK, ROBERT E. KAHN, LEONARD KLEINROCK, DANIEL C. LYNCH, JON POSTEL, LARRY G. ROBERTS & STEPHEN WOLF, INTERNET SOC'Y, A BRIEF HISTORY OF THE INTERNET 8–9 (1997), https://www.internetsociety.org/wp-content/uploads/2017/09/ISOC-History-of-the-Internet_1997.pdf [https://perma.cc/22QT-EUWN].

37. *Id.*

38. *See* Barton Beebe, *The Semiotic Analysis of Trademark Law*, 51 UCLA L. REV. 621, 703 (2004) (describing the trademark system as "an agent in its own struggle to establish itself as the preeminent system of classification").

39. For a more complete explanation of the DNS, see KOMAITIS, *supra* note 11, at 41–45.

40. *Bylaws for Internet Corporation for Assigned Names and Numbers*, art. 1 § 1.1(a)(i), ICANN (Nov. 28, 2019), https://www.icann.org/resources/pages/governance/bylaws-en [https://perma.cc/M6KX-8VB4].

41. *See generally New Generic Top-Level Domains*, ICANN, https://newgtlds.icann.org/en/ (last visited Apr. 21, 2022) (providing information about newly delegated generic top-level domains).

42. *See General Questions*, ICANN, https://www.icann.org/resources/pages/faqs-84-2012-02-25-en#2 (last visited Apr. 21, 2022); *List of Accredited Registrars*, ICANN, https://www.icann.org/registrar-reports/accreditation-qualified-list.html (last visited Apr. 21, 2022).

43. *See Root Zone Database*, IANA, https://www.iana.org/domains/root/db (last visited Apr. 21, 2022); *see also ccTLDs – What Are Country Domain Names?*, IONOS (Sept. 16, 2020), https://www.ionos.com/digitalguide/domains/domain-extensions/cctlds-all-you-need-to-know-about-country-code-domains/ [https://perma.cc/DN6T-4TB4].

44. *See What Are the Most Popular TLDs/Domain Extensions in 2020?*, ICDSOFT (Feb. 10, 2020), https://www.icdsoft.com/blog/what-are-the-most-popular-tlds-domain-extensions/ [https://perma.cc/8AYJ-8GEG].

eligible for registration, the user pays a fee to register it and an annual fee to renew it.

The inevitable interaction between a language-based internet addressing system and trademark law was recognized from the outset, and trademark lobbyists used this interaction to push for the creation of regulatory schemes and bodies.[45] A statement of policy appearing in the *Federal Register* and providing the basis for ICANN's establishment addressed the "trademark dilemma," which centered on the fear that cyberspace could not function as an effective market if businesses could not protect their trademarks.[46] It called for the creation of a dispute resolution process aimed at preventing "cybersquatting and cyberpiracy," while leaving legitimate disputes to be settled in the "appropriate court."[47] Such legitimate disputes include cases where "two parties [have] legitimate competing interests in a particular mark."[48]

By December of 1999, the Uniform Dispute Resolution Policy (UDRP) was in place and guiding proceedings, the first of which involved worldwrestlingfederation.com.[49] Complainants can bring disputes to a dispute resolution provider approved by ICANN, such as the World Intellectual Property Organization or the National Arbitration Forum.[50] The UDRP sets out three elements that must be presented in every complaint: confusion, lack of legitimate interest, and bad faith. Complainants must show how the allegedly infringing domain is confusingly similar to a trademark; that the alleged infringer lacks a right to or legitimate interest in the domain; and in what way the domain was registered and used in bad faith.[51] The only remedies

45. *See* KOMAITIS, *supra* note 11, at 40 ("The current environment concerning the regulation of domain names is a direct consequence of the political pressure and economic rationalization pursued by trademark lobbying.").

46. Management of Internet Names and Addresses, 63 Fed. Reg. 31,741, 31,746 (June 10, 1998).

47. *Id.* at 31,747.

48. *Id.*

49. *Timeline for the Formulation and Implementation of the Uniform Domain-Name Dispute-Resolution Policy*, ICANN (Feb. 25, 2012), https://www.icann.org/resources/pages/schedule-2012-02-25-en [https://perma.cc/SFG9-MBBE].

50. *List of Approved Dispute Resolution Service Providers*, ICANN (Feb. 25, 2012), https://www.icann.org/resources/pages/providers-6d-2012-02-25-en [https://perma.cc/Y779-CT3X]; Elizabeth G. Thornburg, *Going Private: Technology, Due Process, and Internet Dispute Resolution*, 34 U.C. DAVIS L. REV. 151, 161 (2000).

51. *Rules for Uniform Domain Name Dispute Resolution Policy*, § 3(b)(ix), ICANN (Mar. 11, 2015), https://www.icann.org/resources/pages/udrp-rules-2015-03-11-en [https://perma.cc/R9C6-64HV].

are cancellation or transfer of the disputed domain name.[52] When registering a domain, owners agree to participate in the UDRP if an issue arises, much like agreeing to arbitration.[53]

For example, in January of 2020, an individual registered the domain name "tommybahamaboomer.com."[54] Tommy Bahama is a fashion brand that sells "island-inspired apparel," and its mark TOMMY BAHAMA is on the principal register.[55] Tommy Bahama filed a complaint against the individual registrant in March of 2020, alleging that the domain had been registered in bad faith and was confusingly similar to its own trademark.[56] The registrant failed to respond to the complaint, and ultimately the domain was transferred to Tommy Bahama in April of 2020.[57]

Along with the UDRP, the Anticybersquatting Consumer Protection Act (ACPA), codified at 15 U.S.C. § 1125(d) and enacted in 1999, offers trademark owners recourse in the federal courts. Broadly, the ACPA requires a showing of (1) bad faith intent to profit from the mark; and (2) that the owner registers, traffics in, or uses a domain name that is identical or confusingly similar to a distinctive or famous mark.[58] ACPA litigation is more expensive and time-consuming than UDRP proceedings, but the courts are better equipped to handle complex cases and have the ability to offer more remedies.[59] To clarify: the UDRP is an arbitration-like proceeding, whereas ACPA proceedings are federal civil lawsuits.

Before the ACPA and the creation of this centralized allocation and resolution system, web "speculators" had preemptively purchased desirable, non-trademarked domain names.[60] Currently, the UDRP only covers

---

52. *See Uniform Domain Name Dispute Resolution Policy*, § 4(i), ICANN (Aug. 26, 1999), https://www.icann.org/resources/pages/policy-2012-02-25-en [https://perma.cc/3XEG-WX3C].

53. *See, e.g.*, *Domain Registration Agreement*, NAMECHEAP, https://www.namecheap.com/legal/domains/registration-agreement/ (last visited Apr. 21, 2022).

54. Tommy Bahama Grp., Inc. v. Russell Pike, Case No. D2020-0502, Administrative Panel Decision (WIPO Arb. & Mediation Ctr. 2020).

55. *Id.*

56. *Id.*

57. *Id.*

58. *See* 15 U.S.C. § 1125(d)(1)(A); *see also* Tamara Kurtzman, *CYBER CENTER: The Continued Hijacking and Ransoming of the Domain Name System by Modern-Day Corporate Privateers*, BUS. L. TODAY (June 20, 2016), https://www.americanbar.org/groups/business_law/publications/blt/2016/06/cyber_center_kurtzman/ [https://perma.cc/K5GU-5WWN].

59. Jordan A. Arnot, *Navigating Cybersquatting Enforcement in the Expanding Internet*, 13 J. MARSHALL REV. INTELL. PROP. L. 321, 332–34 (2014).

60. Jamie Murphy & Brian L. Massey, *Domain-Name Speculators Buy Up Web Real Estate and Toss the Dice*, N.Y. TIMES (Feb. 21, 1997), https://archive.nytimes.com/

trademark infringement cases, such as someone unaffiliated with McDonald's purchasing mcdonalds.com and holding out for an inflated price.[61] It does not cover an individual or company purchasing a generic domain name, like hotels.com or candy.com, and attempting to sell that domain to Hyatt or Hershey's.[62]

This latter category accounts for several of the most expensive published domain name sales.[63] More than a decade ago, Toys.com sold at auction for $5.1 million to ToysRUs.[64] Typing Toys.com into the browser today redirects to Toysrus.com. Before receiving trademark protection, the president of Hotels.com spent $11 million to purchase the domain name.[65] The infamous Pets.com domain from the dot-com boom sold to its rival, PetSmart, upon bankruptcy.[66] Typing Pets.com into the browser redirects to Petsmart.com. In

---

www.nytimes.com/library/cyber/week/022197domain.html [https://perma.cc/9P6D-BXV4].

61. *See* ICANN, *supra* note 52; *see also* Laurence R. Helfer & Graeme B. Dinwoodie, *Designing Non-National Systems: The Case of the Uniform Domain Name Dispute Resolution Policy*, 43 WM. & MARY L. REV. 141, 152–53 (2001) ("[T]he UDRP covers only a narrow segment of the myriad intellectual property disputes currently percolating in cyberspace. In essence, the UDRP establishes fast and inexpensive online procedures that allow trademark owners to recapture domain names held by persons who register and use domain names containing those marks in bad faith.").

62. *See* Julia Hörnle, *The Uniform Domain Name Dispute Resolution Procedure: Is Too Much of a Good Thing a Bad Thing?*, 11 SMU SCI. TECH. L. REV. 253, 254 (2008) ("The UDRP was drafted narrowly for the purpose of combating the Internet phenomenon known as cybersquatting. Cybersquatting is the registration of a domain name similar to a trademark for an illegitimate purpose . . . ."); Karl Maersch, *ICANN't Use My Domain Name? The Real World Application of ICANN's Uniform Domain-Name Dispute Resolution Policy*, 34 J. MARSHALL L. REV. 1027, 1036 (2001) (explaining that the UDRP was developed to address disputes where a "person with no rights in the trade or service mark registers the mark" as a second-level domain); *see also* Doug Isenberg, *Why 'Generic' Domain Names Are Inappropriate for UDRP Complaints*, GIGALAW BLOG (June 9, 2015), https://giga.law/blog/2015/06/09/why-generic-domain-names-are-inappropriate-for-udrp-complaints [https://perma.cc/4NW7-W2CK]. *But see* Thornburg, *supra* note 50, at 167 (citing a domain name resolution decided by the World Intellectual Property Organization Arbitration and Mediation Center involving cybersquatting and a generic second-level domain name).

63. *See* KOMAITIS, *supra* note 11, at 48 ("The domain <loans.com> generated the extravagant amount of $3 million because of its genericness and its ability to encapsulate the sector of providing services for loans.").

64. Chris Irvine, *Top 10 Most Expensive Domain Names*, THE TELEGRAPH (Mar. 10, 2010), https://www.telegraph.co.uk/technology/news/7412544/Top-10-most-expensive-domain-names.html [https://perma.cc/292P-MRLF].

65. *'I Bought Hotels.com Name for $11m'*, BBC (Nov. 2, 2012), http://news.bbc.co.uk/today/hi/today/newsid_9765000/9765923.stm [https://perma.cc/D3ZU-ZNY3]; Irvine, *supra* note 64.

66. Melinda Patterson Grenier, *Pets.com Transfers Domain Name to Former Rival Petsmart.com*, WALL ST. J. (Dec. 5, 2000), https://www.wsj.com/articles/SB975999845454193456 [https://

March 2020, "domain investor" Jon Schultz purchased about a dozen coronavirus-related domain names, hoping to turn a profit.[67]

The value of these generic domain names speaks to the competitive advantage that comes with such a web address. As with property, some online locations are more valuable and more desirable than others. Their value also illustrates which bodies of law operate effectively in this sphere; under the current system, owning a domain name is like renting property or entering into a contract.[68] And, although trademark law is applied in many domain name cases, intellectual property and its underlying principles may fail to map onto the domain name system.[69]

III. BOOKING.COM

Booking.com was an 8-1 affirmance of the Fourth Circuit with Justice Ginsburg writing for the majority, Justice Sotomayor writing a concurrence, and Justice Breyer writing alone in dissent. The procedural history of this case illustrates the quiet erosion of the distinctiveness spectrum, which is a symptom of the more fundamental shift away from consumer protection as the core of trademark law.

A. PRE-SUPREME COURT

Beginning in 2011, Booking.com, a company, filed four separate trademark applications for the mark "booking.com."[70] The USPTO rejected them all,

---

perma.cc/28YH-8ZCS]; Robert McMillan, *Turns Out the Dot-Com Bust's Worst Flops Were Actually Fantastic Ideas*, WIRED (Dec. 8, 2014), https://www.wired.com/2014/12/da-bom/ [https://perma.cc/W9M3-TK2J].

67. Brandy Zadrozny, *Who Owns Coronavirus.com? Behind the Lucrative Profiteering on Website Domains*, NBC NEWS (Mar. 27, 2020), https://www.nbcnews.com/news/amp/ncna1170411 [https://perma.cc/3BUZ-QWS2].

68. *See* Emily Hurtt, *What's in a Name: The Rights to Domain Names Under Various Principles of Law*, 85 U. CIN. L. REV. 839, 855 (2017) (explaining how contract and property law can map onto domain name rights, but that "intellectual property law is largely developed around the need to incentivize individuals to create . . . [T]his underlying purpose is not served in the context of domain names. Domain names are not registered to incentivize internet users to create; rather, they serve a functional purpose"); *see also* KOMAITIS, *supra* note 11, at 13 ("[D]omain names do not always perform the same functions as trademarks and ought not to be associated with the goodwill of a product or a service. For these and other considerable differences domain names should find their own place and individuality within legal theory.").

69. *See* Hurtt, *supra* note 68, at 855–56.

70. Booking.com B.V. v. Matal, 278 F. Supp. 3d 891, 896 (E.D. Va. 2017), *amended*, No. 116CV425LMBIDD, 2017 WL 4853755 (E.D. Va. Oct. 26, 2017), *aff'd sub nom.* Booking.com B.V. v. U.S. Pat. & Trademark Off., 915 F.3d 171 (4th Cir. 2019), *as amended* (Feb. 27, 2019), *aff'd*, 140 S. Ct. 2298 (2020), and *cert. granted, judgment vacated*, 141 S. Ct. 187 (2020), and *aff'd sub nom.* Booking.com B.V. v. U.S. Pat. & Trademark Off., 915 F.3d 171 (4th Cir. 2019), *as amended*

finding that the mark was "merely descriptive of plaintiff's services."[71] After Booking.com objected, the examining attorney found that acquired distinctiveness had not been proven and, in the alternative, that the mark was generic.[72] Booking.com appealed, and the Trademark Trial and Appeal Board (TTAB) affirmed, specifically concluding that the addition of ".com" failed to transform the generic "booking."[73]

Booking.com then filed a civil action in the Eastern District of Virginia.[74] The district court found generally that "a mark composed of a generic [second-level domain] and a [top-level domain] is usually a descriptive mark eligible for protection upon a showing of secondary meaning."[75] This conclusion contradicted the Federal and Ninth Circuits, which had considered similar terms, such as LAWYERS.COM, MATTRESS.COM, and ADVERTISING.COM, and held that the addition of a generic top-level domain failed to presumptively transform a generic term into a protectable term.[76] Nonetheless, based on the facts of this case, including survey evidence that had not been submitted with the trademark application, the court found specifically that Booking.com had "carried its burden of demonstrating the mark's secondary meaning" and so was entitled to have it registered by the USPTO.[77]

The USPTO appealed this decision to the Fourth Circuit.[78] Although the court of appeals considered the district court's general rule to be incorrect, it affirmed the ruling based on the facts.[79] It held, "[A]dding '.com' to [a second-level domain] can result in a non-generic, descriptive mark upon a showing of

---

(Feb. 27, 2019), and *aff'd*, 140 S. Ct. 2298 (2020), and *cert. granted, judgment vacated*, 141 S. Ct. 187 (2020).

71. *Id.* at 897.

72. *Id.*

73. *Id.* at 898.

74. *Id.* at 895.

75. *Id.* at 923.

76. *See, e.g.*, *In re* Reed Elsevier Props., 482 F.3d 1376, 1378 (Fed. Cir. 2007) ("[I]t is abundantly clear that LAWYERS.COM is generic for the services identified."); *In re* 1800Mattress.com IP, LLC, 586 F.3d 1359, 1364 (Fed. Cir. 2009) ("[B]ecause Dial-A-Mattress presented no evidence that '.com' evoked anything but a commercial internet domain, this is not a case in which the addition of '.com' affects the genericness of the mark."); Advertise.com, Inc. v. AOL Advert., Inc., 616 F.3d 974, 980 (9th Cir. 2010) (rejecting the proposed per se rule that the addition of a top-level-domain to a generic term resulted in a protectable mark); *see also* KOMAITIS, *supra* note 11, at 49–51.

77. *Booking.com*, 278 F. Supp. 3d at 923.

78. Booking.com B.V. v. U.S. Pat. & Trademark Off., 915 F.3d 171, 185 n.9 (4th Cir. 2019), *as amended* (Feb. 27, 2019), *cert. granted*, 140 S. Ct. 489 (2019), and *aff'd*, 140 S. Ct. 2298 (2020), and *cert. granted, judgment vacated*, 141 S. Ct. 187 (2020).

79. *Id.*

primary significance to the relevant public."[80] Judge Wynn dissented, arguing in part that companies should be required to choose between an unprotected but valuable generic domain name and a protected yet less valuable non-generic domain name; they should not enjoy the benefits of both.[81]

At each stage the presumption shifted. The USPTO and TTAB argued for a per se rule against protecting certain domain names, depending on the distinctiveness categorization of the second-level domain. The district court, based on the same facts, the same precedent, and the same statute, came to the opposite conclusion. It devised a new rule, providing that the addition of a top-level domain to a generic second-level domain created a new descriptive term. Although the court of appeals held this rule to be legal error, it considered that error harmless.

B. SUPREME COURT

In an 8–1 decision authored by Justice Ginsburg, the Supreme Court affirmed the Fourth Circuit, holding that "[b]ecause 'Booking.com' is not a generic name to consumers, it is not generic."[82]

The Court faithfully applied the primary significance test in its reasoning. According to the Court's statutory interpretation, 15 U.S.C. § 1064(3), which "govern[s] cancellation of registration," evidences the primacy of consumer perception.[83] Under this section, Congress requires the factfinder to consider "the relevant public" when determining whether a registered mark has become generic.[84] In other words, if the relevant public perceives a once-distinct mark as generic, then that mark is no longer eligible for protection. The central question in *Booking.com*, however, was not genericide; *Booking.com* centered around the initial distinctiveness determination.[85] The sections that do address registration eligibility focus instead on the kinds of evidence that can be shown to prove secondary meaning and overcome the exclusion of "merely descriptive" marks.[86]

Regardless, in assessing the primary significance of a mark in the minds of consumers, the Court first dispensed with the USPTO's argument that the

80. *Id.* at 187.
81. *Id.* at 189 (Wynn, J., dissenting).
82. U.S. Pat. & Trademark Off. v. Booking.com B.V., 140 S. Ct. 2298, 2305 (2020).
83. *Id.* at 2304.
84. 15 U.S.C. § 1064(3).
85. The Court dispensed with this differentiation, which the USPTO raised in its briefing, explaining that the courts and the USPTO's own training manual regularly apply the primary significance test to both an initial determination of genericness and to a determination of genericide. 140 S. Ct. at 2304 n.3.
86. 15 U.S.C. § 1052(e)–(f).

addition of a generic top-level domain to a generic second-level domain resulted in nothing more than a generic term.[87] As the Court explained, appending ".com" to a generic term could "convey to consumers a source-identifying characteristic: an association with a particular website."[88] The USPTO, on the other hand, suggested a more comprehensive per se rule against protecting generic.com terms, which would have rendered some of its own determinations invalid.[89]

Having addressed the USPTO's proposed rule, the Court moved on to a discussion of the distinctiveness determination, affirming the court of appeals's finding of harmless error as to the district court's finding of distinctiveness based on the consumer survey evidence provided by Booking.com.[90] According to Booking.com's expert's interpretation of these "Teflon survey" results, 74.8% of respondents identified "booking.com" to be a brand name, while only 30% identified the control, "washingmachine.com," to be a brand name.[91] The USPTO did not challenge these findings. According to the Eastern District of Virginia, the Fourth Circuit, and the Supreme Court, these results indicated that the primary significance of "booking.com" was source-identifying, not identifying a class of products.

Along with considering the survey evidence, the Court also engaged in a linguistic thought experiment: would a consumer consider another booking company, like Travelocity or Expedia, to be "a Booking.com"?[92] If so, the term is generic. If not, the term is capable of distinguishing goods in the marketplace and serving a source-identifying function. This resembles the "who-are-you/ what-are-you" test, which posits that protectable marks can answer questions like, "Who are you? Where do you come from? Who do you vouch for?" while the generic names can answer only, "What are you?"[93]

The Court also addressed concerns about the "animating policies" of trademark law, highlighting the weak protections descriptive terms receive and the limited anticompetitive impacts that may follow the Court's holding.[94] The Court did not consider the domain name context to be meaningfully distinct

87. *Booking.com*, 140 S. Ct. at 2305.

88. *Id.* at 2306.

89. *Id.* at 2305 ("The PTO's own past practice appears to reflect no such comprehensive rule.").

90. 15 U.S.C. § 1064(3).

91. Brief for Respondent at 12, *Booking.com*, 140 S. Ct. 2298 (No. 19-46), 2020 WL 833244.

92. *Booking.com*, 140 S. Ct. at 2305.

93. Filipino Yellow Pages, Inc. v. Asian J. Publ'ns, Inc., 198 F.3d 1143, 1147 (9th Cir. 1999) (internal quotation marks omitted).

94. *Booking.com*, 140 S. Ct. at 2307.

from any other context; all of the anticompetitive ramifications that might stem from protecting BOOKING.COM would also stem from protecting any other descriptive mark in any other commercial situation.[95] In the Court's view, so long as other descriptive marks could be protected in other contexts, BOOKING.COM should be protectable as well. Similarly, the availability of other avenues for relief, such as unfair competition law, failed to alter the Court's interpretation of protectability under the Lanham Act.[96]

In her concurrence, Justice Sotomayor questioned the probative value of consumer surveys but ultimately sided with the majority in rejecting the USPTO's per se rule.[97] She pointed to "[f]laws in a specific survey design, or weaknesses inherent in consumer surveys generally" to explain that consumer surveys were not the "be-all and end-all."[98] Further, she explained that the USPTO could have found booking.com to be generic based on "dictionary and usage evidence," which would have changed the underlying question of the case.[99] But as it stood, the USPTO did not contest the consumer perception determination as supported by the survey results; it argued that looking to consumer perception was unnecessary because BOOKING.COM was a generic compound.

Justice Breyer dissented, taking issue with the majority's willingness to offer protection to generic terms that have acquired secondary meaning. As he explained, "[The Lanham Act] extended trademark protection to descriptive marks that have acquired secondary meaning. But it did not disturb the basic principle that *generic* terms are ineligible for trademark protection."[100] He considered a generic top-level domain to function like "Inc." or "Co." at the end of a business entity's name. Under *Goodyear Co. v. Goodyear Rubber Co.*,[101] a trademark case from 1888, such generic commercial designations, when combined with other generic terms, created nothing more than a generic compound. While the majority held that the Lanham Act liberalized this common law principle, Breyer found *Goodyear* to be controlling.[102]

---

95. *Id.* at 2308 ("All descriptive marks are intuitively linked to the product or service and thus might be easy for consumers to find using a search engine or telephone directory. The Lanham Act permits registration nonetheless.").

96. *Id.* ("We have no cause to deny Booking.com the same benefits Congress accorded other marks qualifying as nongeneric.").

97. *Id.* at 2309 (Sotomayor, J., concurring).

98. *Id.*

99. *Id.*

100. *Id.* at 2311 (Breyer, J., dissenting) (internal citations omitted).

101. 128 U.S. 598 (1888).

102. 140 S. Ct. at 2311 (Breyer, J., dissenting) ("I cannot agree with respondent that the 1946 Lanham Act 'repudiate[d] *Goodyear* and its ilk.'").

In addition to differing on doctrine, Justice Breyer also departed from the majority as to the efficacy of consumer surveys, especially when the term at issue is exclusive by definition.[103] Linguistic markers like "the" and "co." signal single entities, much like ".com," which might give survey respondents a misleading sense of source identification.[104] "The Shoe Company" might read like a brand because it sounds like an official place that sells shoes.[105] In other words, survey results can be misleading—a third of the survey respondents believed washingmachine.com was a brand. As Justice Breyer pointed out, the difference is not due to the innate distinctiveness of BOOKING.COM but rather the advertising surrounding it.[106]

Justice Breyer closed his dissent by considering the anticompetitive ramifications of the majority's holding.[107] Like Judge Wynn argued in his Fourth Circuit dissent, Justice Breyer emphasized the "valuable piece of online real estate" that generic.com domain owners enjoy.[108] This competitive advantage is not tied to the goods or services being offered but instead to the luck and resources of someone operating in a first-come, first-serve system. For generic domain names, trademark law duplicates existing protections while stifling competition.

IV. ANALYSIS

This Part explains that, although the majority predictably aligned with precedent settled by cases like *Bayer Co. v. United Drug Co.*,[109] it did so without interrogating whether those precedents served the purposes of trademark law, especially as applied in new contexts. The anticompetitive effects that come with protecting merely descriptive marks are amplified in the domain name sphere. Companies have the ability to capture a piece of not only the linguistic commons, but also the online commons. With these purposes and this context in mind, this Part offers a new distinctiveness category that would improve the validity analysis under trademark law. This new category would encompass executable matter, such as top-level domain names, and explicitly exclude that category from trademark protection.

---

103. *See id.* at 2313 ("[S]urvey evidence has limited probative value in this context.").
104. *See id.* at 2312.
105. *See id.* ("'Wine Company' expresses the generic concept of a company that deals in wine . . . .").
106. *Id.* at 2314.
107. *Id.*
108. *See id.* at 2315.
109. 272 F. 505, 509 (S.D.N.Y. 1921).

A. BOOKING.COM WAS A MISSED OPPORTUNITY TO REVIVE GENERICNESS AND REAFFIRM CONSUMER PROTECTION AS THE CORE OF TRADEMARK LAW.

Under the majority's reasoning, any term could be eligible for trademark protection, so long as evidence is presented to show that the term is source-identifying to consumers.[110] Such evidence would push a term past the generic threshold and into the descriptive category. Once in that category, the applicant would need to prove acquired distinctiveness, an analysis that closely resembles the primary significance test.[111]

Whether or not a mark is eligible for registration depends on whether it is generic, and according to the Court, this status is determined by consumers.[112] Paradoxically, this deference to consumers is antithetical to consumer protection; it defeats the very purpose of the distinctiveness requirement. Any term is eligible regardless of where it falls on the spectrum, and the only question left is the scope of protection afforded.

The majority's reliance on the genericide section of the Lanham Act fails to acknowledge the underlying purpose of that section: to render an entire category of terms ineligible for trademark protection.[113] This section indicates that Congress intended for there to be a black hole of genericity. For determining whether a once-distinct mark had been swallowed by this black hole, Congress explicitly referred factfinders to the perceptions of the relevant public.[114] Indeed, where else would the factfinder look?

On the other hand, for unregistered terms the factfinder proceeds from a different starting point: into which distinctiveness category does this proposed mark fall? Public perception can render an inherently distinctive mark generic, but public perception cannot render a generic term distinct. This axiom reveals the consumer protection purposes justifying the existence of the

---

110. Other commentators have recognized this reduction of the distinctiveness spectrum to a stark binary as the essential question at the heart of a trademark validity determination. *See* Beebe, *supra* note 38, at 671; Laura A. Heymann, *The Grammar of Trademarks*, 14 LEWIS & CLARK L. REV. 1313, 1350 (2010) (describing the fundamental trademark validity test as whether "consumers understand the word in question to be the proper name of a brand of product or service or the common name of a product or service").

111. *See* Beebe, *supra* note 38, at 671 (explaining that eligibility is a binary distinction, making a non-generic determination equivalent to a descriptive-with-secondary-meaning determination).

112. *Booking.com*, 140 S. Ct. at 2305.

113. *See* 15 U.S.C. § 1064(3).

114. *See id.* ("The primary significance of the registered mark to the relevant public rather than purchaser motivation shall be the test for determining whether the registered mark has become the generic name of goods or services on or in connection with which it has been used.").

distinctiveness spectrum and the generic marks category. No matter how much a company invests, a generic term remains a generic term for trademark purposes.[115] Under the primary significance test as conceived by the Supreme Court and lower courts, however, the reasoning becomes circular: generic terms cannot be protected, but whether or not something is a generic term in the first place depends on consumer perception.

Moreover, trademark law has not always kept pace with linguistic evolutions, expressive changes, and semantic shifts.[116] This disconnect has led scholars to reconsider how factfinders should evaluate distinctiveness. Some believe the prohibition against protecting generic terms should be lifted,[117] reasoning within a traditional trademark context:

> [P]roducers of goods or services typically do not choose obviously generic names as trademarks. They have no incentive to do so, for (1) the name can and will be duplicated by competitors (with no legal redress); and (2) for that reason, it serves no source-identifying purpose.[118]

But *Booking.com* located the weaknesses in this reasoning. It found a system that dissuades duplication by competitors and always implies, to some degree, a source. In this way trademark law lags behind developments in linguistics,[119] and, as discussed below, it may also lag behind technological advancements.

---

115. Abercrombie & Fitch, Co. v. Hunting World, Inc., 537 F.2d 4, 9 (2d Cir. 1976) ("[N]o matter how much money and effort the user of a generic term has poured into promoting the sale of its merchandise and what success it has achieved in securing public identification, it cannot deprive competing manufacturers of the product of the right to call an article by its name.").

116. *See* Jake Linford, *A Linguistic Justification for Protecting "Generic" Trademarks*, 17 YALE J.L. & TECH. 110, 117 (2015) ("While trademark law reflects many regular forms of semantic shift, the doctrine of trademark incapacity drives a wedge between the law of trademark acquisition and the reality of semantic shift."); Rochelle Cooper Dreyfuss, *Expressive Genericity: Trademarks as Language in the Pepsi Generation*, 65 NOTRE DAME L. REV. 397, 398 (1990) ("[A]s trademark owners have begun to capitalize on the salience of these symbols in the culture, the justifications that formerly delineated the scope of the law have lost significance.").

117. *See, e.g.*, Linford, *supra* note 116; Deven R. Desai & Sandra L. Rierson, *Confronting the Genericism Conundrum*, 28 CARDOZO L. REV. 1789 (2007); Sandra L. Rierson, *Toward a More Coherent Doctrine of Trademark Genericism and Functionality: Focusing on Fair Competition*, 27 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 691 (2017). In these articles, the authors locate the purpose of trademark law in the common law tradition of unfair competition. On the other hand, this Note's interpretation of the goals of trademark law subordinates unfair competition to the broader, more fundamental purpose of consumer protection. As this Note argues, unfair competition is not prevented for its own sake, but rather because unfair competition disserves the consuming public.

118. Desai & Rierson, *supra* note 117, at 1810–11.

119. For further writing on how trademark law could benefit from linguistics, see Graeme B. Dinwoodie, *What Linguistics Can Do for Trademark Law*, *in* TRADE MARKS AND BRANDS: AN

The Supreme Court could have used this as an opportunity to develop a new test for determining whether a mark is generic or descriptive. Justice Breyer flirted with such a shift by referring to the generic "nature" of certain terms, which suggests that, just as terms can be inherently distinctive, other terms can be inherently generic.[120] If a term is by nature generic, then no shift in public perception could change that nature. What Justice Breyer failed to address is how examining attorneys and courts should determine which terms are by nature generic. This determination might be simple for single words—"shoes," "laptops," and "blouses" are all clear examples of generic product categories. But the analysis gets trickier when generic terms are strung together and the result is considered as a whole. In *Goodyear*, the Supreme Court devised a shortcut for a specific situation: generic term + generic company designation = unprotectable mark.[121] But, because the majority in *Booking.com* was unconvinced that ".com" and "Co." operated equivalently, the discussion defaulted back to the primary significance test and, with it, manipulable public perception.[122]

Despite the primary significance test being the default for factfinders, a variety of new and modified tests have been put forth by scholars over the past four decades.[123] Some tests ask factfinders to ignore non-commercial evidence altogether, focusing solely on evidence about how consumers understand a proposed mark to function in the marketplace.[124] This would foreclose the use

---

INTERDISCIPLINARY CRITIQUE 140 (Lionel Bently, Jennifer Davis & Jane C. Ginsburg eds. 2008).

120. *See* U.S. Pat. & Trademark Off. v. Booking.com B.V., 140 S. Ct. 2298, 2310 (2020) (Breyer, J., dissenting).

121. *See* Goodyear Co. v. Goodyear Rubber Co., 128 U.S. 598, 604 (1888) ("'Goodyear Rubber' are terms descriptive of well-known classes of goods . . . . Names which are thus descriptive of a class of goods cannot be exclusively appropriated by any one. The addition of the word 'Company' only indicates that parties have formed an association or partnership to deal in such goods, either to produce or to sell them . . . . The designation 'Goodyear Rubber Company' not being subject to exclusive appropriation, any use of terms of similar import, or any abbreviation of them, must be alike free to all persons.").

122. *See* 140 S. Ct. at 2305–06 ("According to the PTO, adding '.com' to a generic term—like adding 'Company'—'conveys no additional meaning that would distinguish [one provider's] services from those of other providers.' . . . That premise is faulty.").

123. *See, e.g.*, *supra* note 7; *see also* Jason K. Levine, *Contesting the Incontestable: Reforming Trademark's Descriptive Mark Protection Scheme*, 41 GONZ. L. REV. 29, 71 (2005).

124. *See* Desai & Rierson, *supra* note 117, at 1854 ("Under our proposed approach, however, . . . [t]he court would first examine whether [the term] had any source identifying qualities. If that inquiry failed to demonstrate that [the term] retained any level of source identification, the inquiry would end because, absent source-identification, the term is not functioning as a trademark. If, however, source identification were found, the court would examine whether the competitors required the term to enter and compete in the market. The defendant would bear the burden of showing that there was no other commercially viable way

of dictionary definitions and media uses.[125] Others believe that generic terms should be protectable in certain circumstances, again hinging on consumer comprehension.[126] These proposals are a distilled version of the primary significance test; they give exclusive power to public perception in commercial contexts.[127] In doing so, these tests isolate and intensify the central weakness of the primary significance test: relying on impressionable public understanding to determine which marks receive protection.

Rather than focusing on consumers, another option for factfinders would be to inquire about competitors. This would put "fair use first,"[128] which some scholars have argued for in other areas of the distinctiveness inquiry. For example, Alexandra Roberts offers, "[The speech-act test] simply asks whether a hypothetical competitor could use one or more of the trademark's component terms in good faith to describe its own product if the term is defined in accordance with contemporary usage."[129] Roberts introduced this test to address the problem of overprotecting descriptive terms. By inquiring about competitors' speech, she posits, consumers are better protected because more descriptive terms will be available for common use.

This fair use first test could be modified to replace the primary significance test altogether, serving as the initial hurdle applicants would need to clear in order to obtain a trademark. If applied to the genericness determination, this fair use first test would categorically exempt generic terms from eligibility—as the Lanham Act codified. For *Booking.com* in particular, competitors would have good faith reasons to use both "booking" and ".com" separately, as well as "booking.com" as part of a larger term.[130] This approach, making fair use an analytical tool in deciding initial eligibility rather than an after-the-fact

---

in which to refer to a given product."); Heymann, *supra* note 110, at 1333 ("[O]nce it has been determined that a word functions as a mark—that it can and does function as a proper name of a X product or service—its degree of metaphor is irrelevant to its validity.").

125. *See* Desai & Rierson, *supra* note 117, at 1836.

126. *See* Rierson, *supra* note 117, at 756 ("A court dispensing equitable remedies (which are typical in trademark infringement cases) should shape such remedies in a manner sufficient to prevent the consumer fraud and unfair competition that may result when words or product features have acquired distinctiveness, but nonetheless cannot be protected from trademark infringement as a result of the genericism or functionality doctrine.").

127. A question beyond the scope of this Note is how intellectual property scholars using semiotic frameworks define the relevant consuming public for purposes of the primary significance test.

128. Roberts, *supra* note 25, at 1081–85.

129. *Id.* at 1078.

130. The Court rejected an argument that ".com" was functional and, like with trade dress, could not be protected. U.S. Pat. & Trademark Off. v. Booking.com B.V., 140 S. Ct. 2298, 2306 n.5 (2020). Looking at it from the hypothetical competitor's perspective, however, the generic top-level domain does seem more like a necessity and less like a choice.

defense would better protect potential junior mark holders and later entrants. Their language needs would be considered at the outset of the trademark eligibility decision, rather than asserted only as a shield when a senior mark holder brings an enforcement action.

Under this Note's conception of this new test, however, focusing on fair use first could unravel the current trademark system and vitiate the Lanham Act's express protection of descriptive terms that have acquired secondary meaning. Although this solution could bring the trademark system closer to its consumer protection roots, it cuts against Supreme Court jurisprudence and USPTO practices.

B. A NEW DISTINCTIVENESS CATEGORY WOULD ALLOW THE CURRENT LEGISLATIVE SCHEME TO PERSIST, WHILE GUARDING AGAINST THE MONOPOLIZATION OF LANGUAGE.

Rather than overhaul the current trademark eligibility determination by discarding the primary significance test, factfinders could instead recognize a new category along the distinctiveness spectrum: executable matter. Executable matter refers to the parts of a term that operate outside of linguistic communication. The "1-800" in a phone number and the ".com" in a web address, for instance, trigger an operation in a system—they ring a phone or take a user to a site, and without this function they would lose their value. While the basic question in *Booking.com* centered on whether a top-level domain was generic or descriptive, another solution could have been to acknowledge that it, as executable matter, was essentially neither.

In order to rethink the proper categorization of a top-level domain, this Note borrows from the fields of new media theory and literary criticism. This is in keeping with the intersection of trademark law and linguistics/semiotics, an overlap pioneered by the intellectual property scholars whose arguments are discussed in the previous Section. Some of these commentators applied their cross-field frameworks to *Booking.com*,[131] but in doing so they may have overlooked the unique functions of quasi-linguistic matter like ".com." As new media theorists have explained, "Code that runs on a machine is performative

131. *See, e.g.*, Rebecca Tushnet, *A Few Thoughts on the* Booking.com *Argument*, TUSHNET.COM (May 7, 2020), https://tushnet.com/2020/05/07/a-few-thoughts-on-the-booking-com-argument/ [https://perma.cc/MV6X-4K3F]; Jake Linford, USPTO v. Booking.com – *A Linguistic Justification for 'Generic Trademarks'*, WRITTEN DESCRIPTION (June 16, 2020), https://writtendescription.blogspot.com/2020/06/uspto-v-bookingcom-linguistic.html. Linford also contributed to a brief arguing that the trademark incapacity doctrine, which treats the generic nature of a trademark as immutable, is misguided. *See* Brief of Trademark and Internet Law Professors as Amici Curiae in Support of Respondent, *Booking.com*, 140 S. Ct. at 2298 (No. 19-46), 2020 WL 1028360.

in a much stronger sense than that attributed to language . . . . [C]ode running in a digital computer causes changes in machine behavior . . . ."[132] Or, put differently, code is "*hyper*linguistic. Code is a language, but a very special kind of language. *Code is the only language that is executable*."[133] Instead of treating top-level domain names like language, this Note proposes treating them like code. The domain names do something—they take an internet user to a particular website. Without this function, the trademark application for BOOKING.COM likely would never have been filed.

Indeed, looking backward, this category of executable matter may have existed before the concept of code as performative language had been developed. It may have justified the outcomes in the cases discussed in *Booking.com*. For example, in *Goodyear* company designations were considered generic pieces of language—but what if instead they were viewed as being purely legally operative, with at most a shadow of linguistic meaning? Put differently, when appended to a company name, designations like "inc." and "co." *do* something. They flag to consumers the type of organizational category a business falls into, while also speaking that category into existence. They describe the legal positioning of an entity, and through this description reify the legal framework. When articulated with this legal function, "inc." and "co." are not linguistic but rather executive, and they should be considered non-distinctive.

Or, even more similar to the function of ".com" in a domain name, take the case of mnemonic telephone numbers like 1-800-FLOWERS.[134] Can these numbers receive trademark protection? Courts struggled with this problem three decades ago, sometimes treating mnemonic phone numbers as run-of-the-mill word marks requiring a comparative distinctiveness analysis, and other times protecting an admittedly generic mnemonic.[135] With the new category of executable matter, those telephone digits, 1-800-, would be excluded from receiving protection. The trademark applicants include the numbers because the numbers do something—they indicate to consumers what buttons to press on their telephone to connect with the applicant—not because the numbers are creative or catchy. By design, the digits are destination-identifying, not source-identifying.

---

132. KATHERINE N. HAYLES, MY MOTHER WAS A COMPUTER: DIGITAL SUBJECTS AND LITERARY TEXTS 50 (2005).

133. ALEXANDER R. GALLOWAY, PROTOCOL: HOW CONTROL EXISTS AFTER DECENTRALIZATION 165 (2004).

134. Horky, *supra* note 20, at 213.

135. *Id.* at 230–37 ("The courts have forced mnemonics into the traditional [trademark] categories even though their technological anomalies require special consideration.").

With the introduction of new technological platforms, the executable nature of these terms may be disarticulated. Under this Note's theory, 1-800-FLOWERS could receive protection today if applicants showed that it was being used in commerce not as a phone number, but as something else, like a website address. Similarly, company designations could be disarticulated in ways that would allow for trademark protection, like using "inc." as the name of a magazine. In the future, the internet as we now understand it may shift. URLs may fall out of use, and then, if Booking.com elects to keep the nostalgic appendage of a top-level domain name, it should finally be allowed trademark protection. This further protects consumers by promoting fair competition on popular platforms; as long as companies are vying for domains, the executable piece of a domain name lies beyond the reach of trademark.

For *Booking.com*, applying this new category post-hoc results in a simple analysis. Before looking at the compound term, a factfinder would ask the threshold question of whether any part of the mark constituted executable matter. If so, then that part of the mark is ignored for the distinctiveness analysis. For BOOKING.COM, this means that the mark's eligibility would have hinged on just the word "booking," which is easily categorized as a generic term.

C. TRADEMARK LAW IS REDUNDANT AND INEFFICIENT WHEN APPLIED IN THE DOMAIN NAME CONTEXT.

Consumer protection, semiotics, and executability aside, the domain name system infrastructure and market both provide ample protection for companies and their investments. Providing trademark protection for full domain names in this context is redundant and inefficient.

Because of regulatory bodies like ICANN, consumers can trust that a domain name resolves to the correct website.[136] As discussed above, according to the domain name system only one entity can be recognized as the owner of a particular domain name at a particular time. This could be offered as evidence of source identification for trademark purposes, as the majority relied on it.[137] But when considered in the context of the domain name system, this exclusivity function of websites reveals how unnecessary trademark protection really is.[138] Because a domain name corresponds to one and only one website,

136. Thornburg, *supra* note 50, at 159 ("[T]here can be no more than one amazon.com.").

137. U.S. Pat. & Trademark Off. v. Booking.com B.V., 140 S. Ct. 2298, 2306 (2020).

138. *See* Brief for Elec. Frontier Found. as Amicus Curiae in Support of Petitioners at 17, *Booking.com*, 140 S. Ct. at 2298 (No. 19-46), 2020 WL 256707 ("Trademark law should re-enforce the organizational structure of the DNS by limiting protection to domain names that combine [generic top-level domains] with non-generic [second-level domains] that can actually serve a unique distinguishing function."); *see also* Thornburg, *supra* note 50, at 159 n.28 ("The

competitors are incentivized to *avoid* associations with a generic web address; confusion runs the risk of a consumer ending up at the competitor's website. If generic.com domains necessarily signal a single entity, as the Supreme Court found, then such domains do not require the protection trademark law provides.

Furthermore, the measures Congress put in place for regulating trademarks and domain names reveal the extent to which it expected the systems to meaningfully interact. ICANN, UDRP, and ACPA all protect domain names that incorporate a registered trademark from being held hostage by squatters. This organization, policy, and statute could each evolve in purpose and in scope, but in their current form they conceive of only the second-level domain, such as "booking," as the trademarkable part of a domain name.[139] Trademark applications for generic.com domains were being filed before ICANN was formally established—this is not a new issue to the trademark-domain intersection.[140]

Now that the Supreme Court has injected intellectual property protection into an area more akin to real property or contract rights, new questions will arise that regulatory bodies are not equipped to resolve. For example, if Booking.com were to lose ownership of its domain, could another entity purchase it and use it? And would the trademark protection transfer to the new company? Could a company trademark a generic.com domain that it uses but doesn't own?[141] If a company did receive trademark protection for a domain name it did not currently own, could it then request a proceeding pursuant to the UDRP? As mentioned above, trademark rights last in perpetuity, so even when the domain registration ends, the trademark protection continues.

---

uniqueness of second-level domain names means that domain name holders will have broader rights than those granted by trademark law, where multiple companies can have the same trademark so long as there is no danger of confusion.").

139. Indeed, many commentators argue that the regulatory regime *should* evolve. *See* KOMAITIS, *supra* note 11, at 172–84 (cataloging reforms at the regulatory and techno-legal levels that would provide a more balanced approach and end the abuse of the DRP system); Hörnle, *supra* note 62, at 289 (detailing changes that would address the UDRP's "several serious procedural deficiencies"); David A. Simon, *An Empirical Analysis of Fair Use Decisions Under the Uniform Domain-Name Dispute-Resolution Policy*, 53 B.C. L. REV. 65, 129 (2012) (explaining how ICANN should reform the UDRP to eliminate "U.S. favoritism").

140. *See, e.g.*, U.S. Trademark Application Serial No. 75/408570 (filed Dec. 19, 1997) (cars.com); U.S. Trademark Application Serial No. 74/625099 (filed Jan. 24, 1995) (casino.com).

141. *See* Laura A. Heymann, United States Patent and Trademark Office v. Booking.com B.V.*: How Do We Know When Something Is a Name?*, GEO. WASH. L. REV.: ON THE DOCKET (July 2, 2020), https://www.gwlr.org/booking-com-b-v-how-do-we-know-when-something-is-a-name/ [https://perma.cc/UK4M-2H7E].

Furthermore, offering trademark protection to a generic.com domain will cause changes in the market beyond those addressed by Justice Breyer's dissent. As he explained, this decision will likely have a chilling effect, dissuading "others from using variants on the registered mark."[142] In addition, trademarking of a generic.com domain will devalue all slightly similar domain names. Because "ebooking.com" or "booked-up.com" or other neighboring domain names could infringe on the registered trademark, they necessarily carry some risk for any potential buyer. *Booking.com* will also impact the worth of any generic.com domain because there is now the possibility of trademark protection. A currently non-protected generic.com domain could, with the right consumer survey and the right advertising, become eligible for trademark protection.

At the same time, the domain name market is not a perfect system. Just as pets.com and toys.com now redirect to well-known brand websites, Booking.com could conceivably have purchased the same domain names that will be chilled or devalued by the Court's decision. Put differently, it's possible that Booking.com could have achieved an internet monopoly by buying up domains somewhat similar to its own domain. This alternative means to the same end raises a fundamental question: why did Booking.com pursue trademark protection in the first place? It alleged at oral argument that it would not enforce its protection against similar domain names.[143] Trademark litigation is expensive, consumer surveys are expensive, and, according to Booking.com, the active benefits of having a trademark on the principal register will not be enjoyed.

One explanation is that Booking.com does not need to formally initiate litigation to ward off potential competition. If potential competitors search for BOOKING.COM in the trademark database today, they will find it in the principal register. In reality, Booking.com holds very weak protection over the mark,[144] but business owners who lack a legal background or legal counsel may not be aware of the scope of protection. Indeed, trademark disputes are often

142. 140 S. Ct. at 2315.

143. Transcript of Oral Argument at 55–56, *Booking.com*, 140 S. Ct. 2298 (No. 19-46) ("JUSTICE ALITO: You would not—your client would not object to the registration of any trademark that simply made a slight variation in Booking.com? That would be fine? All of those companies could register their trademarks? . . . Would your client object to that? . . . MS. BLATT: They don't and have not and would not. Now, if there was fraud and somebody ripping off the goodwill based on Booking.com, I'm sure they'd want to sue, but it–it would be very hard to bring that lawsuit. Very hard.").

144. *Booking.com*, 140 S. Ct. at 2308 ("Booking.com concedes that 'Booking.com' would be a 'weak' mark.").

resolved in the pre-litigation stages.[145] Companies often send cease-and-desist letters, reportedly even when the arguments in those letters lack merit, and the threat of suit is often enough to push the alleged infringer to choose a different mark.[146] Booking.com could also initiate complaints under the UDRP now that it possesses the trademark.

By design, the domain name system protects companies and their investments. There is an exclusivity inherent in the one-owner-one-domain structure, which incentivizes competitors to avoid confusingly similar web addresses. And if an entity did try to cause confusion, then unfair competition laws are always available for recourse. Other laws and regulatory bodies also safeguard valid trademarks in the domain space. Adding traditional trademark protection atop all of this frustrates the system, creating redundancies and inefficiencies that work against the underlying purposes of trademark law.

D. IMPLEMENTING POSSIBLE SOLUTIONS

The once-clear poles of the distinctiveness spectrum have been pulled together. At this point, what purpose does the spectrum serve? Generic and descriptive have collapsed into one category: not inherently distinctive but can acquire secondary meaning. Suggestive, arbitrary, and fanciful make up the other category: inherently distinctive.

Allowing consumers to determine genericness fails to protect consumers, but there are options for strengthening the distinctiveness requirement. In the wake of *Booking.com*, like in the wake of *Anti-Monopoly*, Congress could amend the Lanham Act and strengthen the spectrum. It could require that both descriptive *and* suggestive marks require a showing of secondary meaning to be placed on the principal register.[147] It could also clarify how initial genericness is determined, making the primary significance test applicable only to cases of genericide. Under this Note's interpretation, factfinders should not look to consumer opinions to determine whether an applied-for mark is generic. Instead, they should make the threshold distinctiveness determination with consumer protection as their prime motivation, focusing on what component terms should remain available for collective use. The registration process needs reforming,[148] and this is one place to start.

---

145. William Gallagher, *Trademark and Copyright Enforcement in the Shadow of IP Law*, 28 SANTA CLARA HIGH TECH. L.J. 453, 456 (2012).

146. *Id.* at 485–87.

147. *See* Linford, *supra* note 18, at 1416.

148. *See* Tushnet, *supra* note 23, at 871–72 ("Registration provides opportunities to limit trademark's current structurelessness.").

Congress could also leave the existing legal framework and decades of precedent alone, while nonetheless undoing *Booking.com*. 15 U.S.C. §§ 1052(a)–(e) enumerate categorical exemptions from trademark eligibility, including insignia of the United States and the portrait of a deceased President of the United States during the life of his widow. Congress could add a new subsection excluding generic second-level and top-level domain compounds. This would still leave the examining attorneys, and potentially the courts, to determine genericness as to the domain name parts, but it would create an initial presumption that domain names like booking.com are non-trademarkable when attached to services that match the generic term. This maps onto the nearly per se rule that the USPTO advanced in *Booking.com*. It would undo existing trademark decisions,[149] but it would also create more predictability and administrability. And, regardless of the trademark decision, those domains would continue to belong to their owners.

Furthermore, as Justice Breyer pointed out in his dissent, even without trademark protection, unfair competition law protects companies against malicious, deceptive advertising.[150] Being denied a position on the principal register does not leave a company stranded.

Ultimately, there are many options to reform the distinctiveness spectrum. Factfinders could take Congress at its word and only apply the primary significance test to questions of genericide. For determining whether an applied-for mark is generic, then, examining attorneys and courts could justifiably employ a different test, though this would contradict existing practice and the dicta of the Supreme Court. Congress could amend the Lanham Act to nullify this precedent and clear the way for factfinders. In the alternative, Congress could carve out an exception solely for generic.com domain names, making them presumptively unprotectable, but potentially causing the disorder the Court warned of when dismissing the USPTO's per se rule.

Finally, and most preferably, courts and USPTO examining attorneys could begin drawing upon this Note's proposed new category, executable matter, in making distinctiveness determinations. This option leaves the overarching legal framework untouched but evolves one area within that framework to better align with the current structures and incentives of the domain name system.

149. *See* Brief for Respondent at 1a–11a, *Booking.com*, 140 S. Ct. 2298 (No. 14-96), 2020 WL 833244.

150. *Booking.com*, 140 S. Ct. at 2316 (Breyer, J., dissenting) ("[U]nfair-competition law will often separately protect businesses from passing off and false advertising.").

V. CONCLUSION

In writing about trademarks, scholars, courts, and even this Note often describe trademark terms as being created by companies. Bayer gave us "aspirin."[151] Kellogg's gave us "shredded wheat."[152] This description is only half true; the public "authors" trademarks, too.[153] Without consumers associating a symbol with a product and then purchasing that product, the words and marks would not carry any significance. This is why, at first blush, it seems intuitive to ground any trademark inquiry in relevant consumer understandings. Trademark terms are our terms, and so we should decide their capacity for distinguishing goods and sources in the marketplace.

But this is not the question that consumer surveys generally ask. They ask how consumers think a term operates based on controls and comparisons, but they fail to inquire about how consumers think a term *should* operate. If trademark power devolves in the public, then the tests should go beyond assessing consumer comprehension or public perception. They should evaluate whether the respondents believe language restrictions should be allowed for the words at issue. Put differently, do consumers know what's at stake when they indicate that WASHINGMACHINE.COM is a brand?

I doubt that there will be a people's PTO anytime soon. In the meantime, it bears repeating that trademark rights and domain name contract rights operate differently. Trademarks are context dependent. So long as there is not confusion or dilution, two identical trademarks can exist to identify different goods or services.[154] JAGUAR is a live trademark for the vehicle manufacturer, as well as for a plumbing and bathroom fixture provider.[155] The same word carries multiple meanings and associations simultaneously. In this way, domain names only approximate language—they still operate like IP addresses and lead to only one place. Even if a domain name carries multiple associations, its purpose is to operate within the internet system. Because of this, the well-established linguistic hypotheticals and consumer surveys fall short when applied in this new context without accounting for the fundamental differences. Domain names, indicated by the appendage of ".com," answer

151. Bayer Co. v. United Drug Co., 272 F. 505, 509 (S.D.N.Y. 1921).

152. Kellogg Co. v. Nat'l Biscuit Co., 305 U.S. 111, 111–12 (1938).

153. *See* Steven Wilf, *Who Authors Trademarks?*, 17 CARDOZO ARTS & ENT. L.J. 1 *passim* (1999).

154. *See* Thornburg, *supra* note 50, at 159.

155. PF Waterworks LP succeeded in registering JAGUAR over Jaguar Land Rover Ltd.'s opposition. Jaguar Land Rover Ltd. v. PF Waterworks LP, Opposition No. 91240200 (T.T.A.B. 2019).

neither "Who are you?" nor "What are you?" They answer—and execute—an entirely new question: Where do I go to find you?

The American trademark system has arguably allowed for the "depletion" and "congestion" of useful, viable candidates for emerging brands to use as trademarks.[156] The same logic applies even more forcefully to the by-design scarcity of domain names.[157] Perhaps without realizing it, we use our language in the "shadow of trademark rights,"[158] despite the promise that our public perception is what ultimately invests and imbues a mark with legal significance.

This Note used *Booking.com* to illustrate how this overcrowding has unfolded. Courts' uncritical applications of longstanding precedents have eroded certain safeguards for the linguistic commons[159] and expanded trademark protection for mark owners.[160] As this Note cataloged, there are a variety of potential solutions that would restore consumer protection to the heart of the federal trademark analysis. Although the Supreme Court avoided reworking the system in *Booking.com*, changes could be implemented by examining attorneys, lower courts, and even Congress.

The distinctiveness spectrum is a foundational piece of American trademark law. In theory, it incentivizes creativity and protects our shared linguistic commons, all in the name of promoting fair competition and protecting consumer interests. But language and technology change; meanwhile, the distinctiveness analysis has failed to evolve in tandem. Until URLs fall out of fashion, domain names should be controlled solely under contract rights and the existing domain name infrastructure, safely beyond the reach of trademark law.

---

156. *See* Barton Beebe & Jeanne C. Fromer, *Are We Running Out of Trademarks? An Empirical Study of Trademark Depletion and Congestion*, 131 HARV. L. REV. 945 (2018).

157. *See id.*

158. *See id.* at 1025.

159. *See* Daniel J. Hemel & Lisa Larrimore Ouellette, *Trademark Law Pluralism*, 88 U. CHI. L. REV. 1025, 1026 (2021) ("Scholars and judges of trademark law have come to recognize in recent years that the pool of words, sounds, and symbols that are readily available to signify new goods and services is not unlimited . . . . Earlier arrivals already have devoured the most memorable marks, leaving slim pickings for latecomers. The resulting pressure is pushing trademarks into regions once considered part of the 'linguistic commons'—most recently, in a 2020 Supreme Court decision that a generic name with '.com' appended to it, like 'Booking.com,' was potentially eligible for federal trademark registration.").

160. *See* McKenna, *supra* note 11, at 1899, 1915 (explaining how "courts steadily expanded the scope of trademark law during the twentieth century" such that "[v]irtually every significant doctrinal development in the last century has given mark owners greater control over the use and meaning of their marks").

LUCKY BRAND V. MARCEL: RETHINKING TRADEMARK REMEDIES IN A WORLD WITHOUT DEFENSE PRECLUSION

Yegina Whang†

I. INTRODUCTION

Born out of the ever-elusive desire to optimize the use of judicial resources, discourage repeat litigation, and promise finality for the parties, preclusion has failed to assume the mantle.[1] Preclusion, a product of federal common law, was designed to be a straightforward doctrine allowing courts to quickly throw out repeat cases implicating claims or issues that have already been adjudicated. However, its application has proven to be anything but simple, raising more questions (and, ironically, more litigation) than answers.[2] One such question is whether preclusion can bar a defense that could have been asserted in a prior suit. In other words, does a separate doctrine of "defense preclusion" exist?

The Supreme Court attempted to answer this question in *Lucky Brand v. Marcel*, a civil procedure case couched in a trademark infringement suit.[3] In a unanimous decision authored by Justice Sotomayor, the Supreme Court announced that there was no principle of defense preclusion "unmoored" from the traditional doctrines of claim preclusion and issue preclusion.[4] Instead, the Court held that any preclusion of defenses that could have been raised in a prior suit must first meet the elements of claim preclusion: (1)

DOI: https://doi.org/10.15779/Z38PG1HP7N

† J.D. Candidate, 2022, University of California, Berkeley School of Law. My sincerest gratitude to Professor Talha Syed for supervising this Note. I am also thankful to Kevin Yang and my fellow participants in the Law and Technology Writing Workshop for invaluable support and helpful feedback.

1. For example, litigants continued to file lawsuits challenging President Obama's eligibility for presidency based on his citizenship, even after a court had already decided the issue. Alexandra Bursak, Note, *Preclusions*, 91 N.Y.U. L. REV. 1651, 1652 (2016). The public also continued to file FOIA requests for President Obama's birth certificate, overwhelming the Hawaii Department of Health and demanding considerable time and resources. *Id.*

2. *See* B&B Hardware, Inc. v. Hargis Indus., 575 U.S. 138, 148 (2015) ("Although the idea of issue preclusion is straightforward, it can be challenging to implement."). In the seven years between 1979 and 1986, the Supreme Court handed down twenty-one decisions regarding the application of claim preclusion and issue preclusion. William V. Luneburg, *The Opportunity to be Heard and the Doctrines of Preclusion: Federal Limits on State Law*, 31 VILL. L. REV. 81, 82–83 (1986).

3. Lucky Brand Dungarees, Inc. v. Marcel Fashions Grp., 140 S. Ct. 1589, 1592 (2020).

4. *Id.* at 1595.

identity of claims between the prior and subsequent suits; (2) a final judgment on the merits; and (3) identity or privity between the parties.[5] In rejecting the argument that preclusion should apply to the defenses raised in *Lucky Brand*, the Court emphasized that it was not resolving "when (if ever) applying claim preclusion to defenses may be appropriate."[6] Yet in the same breath, the Court noted in dicta that "there may be good reasons to question *any* application of claim preclusion to defenses,"[7] thereby leaving open the future possibility of rejecting defense preclusion in its entirety.

At first blush, the Court seemed to persuasively make the case for a *per se* rule against defense preclusion. The Court reasoned that defendants often make strategic decisions when forgoing a defense—a decision that preclusion law, out of fairness for the defendant, should not seek to punish. Further, the Court suggested that a doctrine of defense preclusion would be superfluous, since previously successful plaintiffs could avoid the issue altogether by seeking judgment enforcement actions instead of filing new suits. This latter argument assumes that prevailing plaintiffs have a forward-looking remedy, like an injunction, to enforce. In the context of trademark law, however, prevailing plaintiffs often do not have such a remedy at hand to protect their marks against repeat infringers ever since the Supreme Court's decision in *eBay v. MercExchange* rejected the automatic granting of injunctions upon a showing of infringement.[8]

In place of the old automatic injunction rule,[9] *eBay* required plaintiffs to affirmatively show the traditional four factors prior to obtaining an injunction, making it far more difficult for trademark owners to enjoin a competitor's infringement even after winning in court.[10] And while *eBay* addressed a rule

5. *See* Media Rts. Techs., Inc. v. Microsoft Corp., 922 F.3d 1014, 1020–21 (9th Cir. 2019) (identifying the elements of claim preclusion).

6. *Lucky Brand*, 140 S. Ct. at 1595 n.2.

7. *Id.* (emphasis added).

8. *See* eBay Inc. v. MercExchange, L.L.C., 547 U.S. 388, 392–93 (2006) ("[T]his Court has consistently rejected invitations to replace traditional equitable considerations with a rule that an injunction automatically follows a determination [of infringement.]"); *see also* Mark A. Lemley, *Did* eBay *Irreparably Injure Trademark Law?*, 92 NOTRE DAME L. REV. 1795, 1806–09 (2017).

9. The "automatic injunction rule" is a bit of a misnomer for two reasons. First, it was never a formalized rule articulated by any court. Second, injunctions were presumed, but not automatically granted, once a trademark plaintiff showed infringement. *See* Lemley, *supra* note 8, at 1797. In fact, "[b]etween 1984 and 2006 . . . the Federal Circuit never once denied an injunction a prevailing patentee." *See id.* Nevertheless, I refer to this presumption as the "automatic injunction rule" throughout this Note.

10. Lex Machina's database shows that, prior to the Ninth Circuit's *Herb Reed* decision extending *eBay* to trademark infringement suits, district courts within the Ninth Circuit granted 117 (86%) and denied 19 (14%) requests for permanent injunctions after a judgment on the

specific to patent law, the lower courts have since extended *eBay* to the trademark context as well.[11] As a result, *Lucky Brand*, combined with *eBay* and its progeny, could have troubling consequences for trademark plaintiffs by forcing them to make a Sophie's choice: either avoid filing a new case in the face of continuing infringement or risk exposing themselves to new defenses that may undo a prior favorable judgment.[12]

This Note argues that the Court's skepticism of defense preclusion in *Lucky Brand v. Marcel* was misplaced for two reasons. First, the Court's reasoning relied heavily on an emphasis on the potential fairness issues raised by applying preclusion to defenses—a significant departure from the Court's prior preclusion cases. Second, the Court's dicta is problematic in light of *eBay*'s application to trademark law, which has made it more difficult for prevailing trademark owners to obtain an injunction. The Court's suggestion of a categorical rule against defense preclusion underscores that courts should not have extended *eBay* to trademark law in light of its distinct quirks from patent and copyright law, such as the underlying purpose of reducing consumer transaction costs (as opposed to incentivizing creation), the challenges in calculating accurate monetary damages, the difficulty in showing the "likelihood of irreparable harm," and the absence of a parallel "patent troll" problem. By ignoring these differences, the Court overlooked the finality issues implicated by a *per se* rule against defense preclusion in the trademark context. Notwithstanding these flaws in the Court's reasoning, this Note argues that rethinking *eBay*'s applicability to trademarks would help alleviate the potential harmful impact of *Lucky Brand*'s dicta on trademark owners in the future. Specifically, Congress should pass an automatic injunction rule for trademark infringement cases to address the issues identified in this Note.

---

merits. This data was obtained using Lex Machina, with the tag for "Trademark" for "Case Types" and limited to cases pending before courts in the Ninth Circuit from January 1, 2009 to December 1, 2013. After courts began applying the *eBay* factors to trademark cases, district courts granted 89 permanent injunctions (75%) and denied 29 (25%) requests for permanent injunctions following a judgment on the merits. This data was obtained using Lex machina, with the tag for "Trademark" for "Case Types" and limited to cases pending before courts within the Ninth Circuit from December 2, 2013 to March 13, 2021. The data shows that, since the *Herb Reed* decision, over 50% of the 179 trademark plaintiffs who prevailed on the merits did not obtain a permanent injunction.

11. *See, e.g.*, Herb Reed Enters., LLC v. Fla. Ent. Mgmt., 736 F.3d 1239, 1249 (9th Cir. 2013) (finding that *eBay*'s rejection of the automatic injunction rule also "applies to trademark infringement under the Lanham Act"); North American Med. Corp. v. Axiom Worldwide, Inc., 522 F.3d 1211, 1288 (11th Cir. 2008) (finding *eBay* applicable in a trademark case).

12. This Note will only focus on confusion-based trademark infringement cases. Dilution-based infringement suits are outside the scope of this Note.

Part I contextualizes the Court's holding in *Lucky Brand* by providing background on the federal common law doctrine of preclusion. Part II proceeds by discussing *Lucky Brand* and its procedural history. Part III follows by explaining why the Court's analysis of the fairness issues raised by defense preclusion was unwarranted given the policy justifications driving preclusion doctrine. Part IV discusses why *Lucky Brand*, along with the unique features of trademarks more generally, demands reforming trademark law's injunction standard. Finally, Part V recommends that Congress restore the automatic injunction rule for trademark infringement suits to avoid some of the issues potentially caused by *Lucky Brand*'s dicta.

II. BACKGROUND

Courts have struggled to articulate a clear framework for and application of preclusion. This Part attempts to dispel this confusion by first providing a doctrinal background on issue preclusion and claim preclusion and, second, discussing the treatment of defense preclusion prior to *Lucky Brand*.

A. THE DOCTRINES OF ISSUE PRECLUSION AND CLAIM PRECLUSION

A creation of federal procedural common law, the doctrine of preclusion forecloses subsequent litigation where a court has already adjudicated a similar claim or issue.[13] The effects of preclusion are often harsh, as courts will apply it even if the prior judgment was wrong or rested on a legal principle that was later overruled.[14] Under modern law, preclusion comes in two varieties: claim preclusion (or res judicata) and issue preclusion (or collateral estoppel).

Though the two concepts are similar in their effect, they are distinct in their uses. For instance, where there is a prior judgment, claim preclusion forecloses "successive litigation of the very same *claim*, whether or not relitigation of the claim raises the same issues as the earlier suit."[15] Issue preclusion, on the other hand, bars "'successive litigation of an *issue* of fact or law actually litigated and resolved in a valid court determination essential to the prior judgment,' even if the issue recurs in the context of a different claim."[16] In sum, claim preclusion prevents litigation of claims that *were or could*

13. Alan M. Trammell, *Precedent and Preclusion*, 93 NOTRE DAME L. REV. 565, 567 (2017). Preclusion doctrine has deep historical roots. *See, e.g.*, Allan D. Vestal, *Rationale of Preclusion*, 9 ST. LOUIS U. L.J. 29, 30 n.7 (1964); Kevin M. Clermont, *Res Judicata as Requisite for Justice*, 68 RUTGERS U. L. REV. 1067, 1071–73 (2016).

14. Federated Dep't Stores v. Moitie, 452 U.S. 394, 398 (1981).

15. New Hampshire v. Maine, 532 U.S. 742, 748 (2001) (emphasis added).

16. Taylor v. Sturgell, 553 U.S. 880, 892 (2008) (emphasis added) (quoting *New Hampshire*, 532 U.S. at 748–49).

have been litigated in a prior action, whereas issue preclusion prevents re-litigation of issues that were *actually* litigated.[17]

With increasingly overcrowded dockets and limited resources, the American judicial system decided to adopt preclusion for two reasons: to promote finality and to preserve judicial resources.[18] First, by preventing disgruntled litigants from taking a second bite of the proverbial apple, preclusion protects parties from duplicative and vexatious litigation and lends confidence to a court's decision as final.[19] Second, while finality focuses on the parties directly involved in the suit, judicial economy focuses on society's desire for efficiency.[20] With the ever-increasing rise in litigation, society has a great interest in "seeing that cases are tried just once" so that judges can adjudicate more cases fairly, accurately, and quickly.[21] This desire for judicial efficiency can be traced back to the phrase *interest reipublicae ut sit finis litium*, which translates to "it is in the interest of the State that there be an end to litigation."[22] Efficiency is also an important policy goal that underlies civil procedure generally, manifesting itself in other procedural rules regarding compulsory counterclaims,[23] joinder,[24] and summary judgment.[25]

Notwithstanding these goals of finality and efficiency, procedural due process rights impose an outer limit on preclusion. In applying claim preclusion, the Supreme Court has required mutuality, or identity of the parties between the two suits, to afford everyone "a full and fair opportunity to litigate the claims and defenses or the issues in the prior proceeding."[26] While courts

17. Claim preclusion requires the following: (1) identity of claims between the prior and subsequent suits; (2) a final judgment on the merits; and (3) identity or privity between the parties. Media Rts. Techs., Inc. v. Microsoft Corp., 922 F.3d 1014, 1020–21 (9th Cir. 2019). On the other hand, issue preclusion requires that "(1) the issue at stake was identical in both proceedings; (2) the issue was actually litigated and decided in the prior proceedings; (3) there was a full and fair opportunity to litigate the issue; and (4) the issue was necessary to decide the merits." Janjua v. Neufeld, 933 F.3d 1061, 1065 (9th Cir. 2019).

18. *See* Parklane Hosiery Co. v. Shore, 439 U.S. 322, 326 (1979) (emphasizing that preclusion has the "dual purpose of protecting litigants from the burden of relitigating an identical issue with the same party or his privy and of promoting judicial economy by preventing needless litigation").

19. *See id.*

20. *See id.*

21. Vestal, *supra* note 13, at 32.

22. Clermont, *supra* note 13, at 1069.

23. *See* FED. R. CIV. P. 13(a).

24. *See id.* at 18(a).

25. *See id.* at 56.

26. Clermont, *supra* note 13, at 1089.

"largely have abandoned . . . mutuality with respect to issue preclusion,"[27] they have made clear that "the person *against whom* preclusion is being applied must have been a party to the initial litigation."[28] Because preclusion requires mutuality or, at the very least, requires that the person against whom preclusion is asserted was represented in the prior suit, courts largely dismiss any due process concerns when applying the doctrine.[29]

Despite the Supreme Court's best efforts to clarify preclusion, lower courts and scholars have criticized the doctrine as inconsistent and incoherent. The source of this problem is two-fold. First, parties often "conflate . . . claim preclusion with . . . issue preclusion, and vice versa, when it comes time to employ them in a specific scenario."[30] Second, and relatedly, courts and litigants have not been doctrinally clear about the meanings and uses of claim preclusion and issue preclusion.[31] This confusion has led courts to apply incorrect elements in their analysis or weigh interests that are immaterial to the decision on whether to bar a claim or issue.[32]

Preclusion presents its own unique set of challenges in the trademark context, as the Supreme Court noted in *Lucky Brand*.[33] Trademark infringement suits are especially fact-dependent, since they rely on the twin showings of a mark's distinctiveness and a likelihood of confusion between marks.[34] To make matters worse, a mark's distinctiveness may change over time. For instance, protectible marks, like aspirin, escalator, and chapstick, may become generic over time (a phenomenon referred to as "genericide").[35] By the same token, an

27. Trammell, *supra* note 13, at 574 n.41. For instance, a stranger may use issue preclusion defensively to prevent re-litigation by a party from a prior suit. *See* Bernhard v. Bank of Am. Nat'l Tr. & Sav. Ass'n, 122 P.2d 892, 894 (Cal. 1942) (en banc).

28. Trammell, *supra* note 13, at 574–75.

29. *See Bernhard*, 122 P.2d at 894 (explaining how res judicata conforms "to the mandate of due process of law").

30. Bursak, *supra* note 1, at 1656.

31. *Id.* at 1656–57.

32. *See, e.g.*, *id.* at 1656 (citing a Federal Circuit case where the court erroneously used the term "res judicata" when it was actually applying collateral estoppel).

33. Lucky Brand Dungarees, Inc. v. Marcel Fashions Grp., 140 S. Ct. 1589, 1596 (2020) (affirming that preclusion "generally 'does not bar claims that are predicated on events that postdate the filing of the initial complaint[,]'" a principle that "takes on particular force in the trademark context" (citation omitted)).

34. *See* 3 ANNE GILSON LALONDE, GILSON ON TRADEMARKS § 13.20 (2020) ("Application of the doctrines in trademark cases presents special problems, however, because constantly changing trademark use and shifting public perceptions do not easily lend themselves to final adjudication.").

35. Whitson Gordon, *How a Brand Name Becomes Generic*, N.Y. TIMES (June 24, 2019), https://www.nytimes.com/2019/06/24/smarter-living/how-a-brand-name-becomes-generic.html [https://perma.cc/VLQ8-NX86].

unprotectable descriptive mark may become protectable by acquiring a secondary meaning.[36] Thus, changes in marketplace realities and other operative facts may transform the same claim or issue from a prior suit to the point of rendering preclusion inapplicable.[37] Adding to this challenge, the Supreme Court has previously held that, so long as all the other elements of issue preclusion are met, the Trademark Trial and Appeal Board's determination of likelihood of confusion will bar later determinations of the same issue in federal court infringement actions.[38]

Applying the traditional doctrines of claim preclusion and issue preclusion consistently has been no easy feat for courts. However, preclusion is further complicated by the fact that some courts have applied it to bar *defenses* that could have been raised in a prior suit.

B. "DEFENSE PRECLUSION" BEFORE LUCKY BRAND

Even before the Supreme Court's holding in *Lucky Brand*, there was debate about the applicability of preclusion to defenses. On one hand, the Supreme Court has seemingly made clear in at least two cases that courts can preclude defendants from raising defenses. In *Beloit v. Morgan*, the Supreme Court suggested that the city of Beloit could not raise a validity defense against the bonds it had issued to Morgan because a prior lawsuit had already determined that the city had to pay earlier bonds from the same set.[39] Even though the bonds were nominally different, the Court determined that the two suits were between "the same parties touching the same subject matter."[40] Thus, the previous suit had necessarily already determined the bonds' validity.[41] In so holding, the Court reasoned that "[a] party can no more split up defences than indivisible demands, and present them by piecemeal in successive suits growing out of the same transaction."[42]

The Supreme Court reaffirmed *Beloit* in *Cromwell v. County of Sacramento*, where it ruled that defenses that were not raised in a prior action could not be heard in "the prosecution of a second action upon the same claim or demand."[43] According to the *Cromwell* Court, a judgment "estops not only as

36. *See* Lanham Act § 2(f), 15 U.S.C. § 1052(f) (permitting registration of a mark "which has become distinctive"); *see also* Qualitex Co. v. Jacobson Prods. Co., 514 U.S. 159, 166, 174 (1995) (finding that the green-gold color of dry-cleaning press pads at issue acquired a secondary meaning and were thus protectable).

37. *Lucky Brand*, 140 S. Ct. at 1596.

38. B&B Hardware, Inc. v. Hargis Indus., 575 U.S. 138, 160 (2015).

39. Beloit v. Morgan, 74 U.S. (7 Wall.) 619, 621–22 (1869).

40. *Id.* at 622.

41. *Id.* at 623.

42. *Id.*

43. Cromwell v. County of Sacramento, 94 U.S. 351, 352–53 (1877).

to every ground of recovery or defence actually presented in the action, but also as to every ground *which might have been* presented."[44] Restatements also provide that "when a valid and final personal judgment is rendered in favor of the plaintiff . . . the defendant cannot avail himself of defenses he might have interposed, or did interpose, in the first action."[45] Prior to *Lucky Brand*, however, the Supreme Court had never made clear whether defense preclusion was merely an extension of the traditional doctrines of preclusion, thereby requiring both mutuality of parties and identity of claims, or an entirely new doctrine in and of itself.[46]

On the other hand, scholars and litigants have pushed back against preclusion's applicability to defenses. For instance, some have argued that "[c]laim preclusion generally applies only to those asserting claims, not to those defending against them."[47] Under this theory, courts should not punish defendants for forgoing certain defenses since such decisions are not based on the actual merits of the argument. Rather, the decision to forgo a defense is usually based on other practical considerations, like the amount in controversy, the difficulty of obtaining certain evidence, and the expense of litigation.[48] While plaintiffs may have the same considerations when filing suit, they still have the advantage of affirmatively bringing the suit and framing the case by choosing which claims to pursue and selecting the forum. Thus, scholars against defense preclusion believe that perhaps only plaintiffs should be limited in the types of claims they are able to bring forward.

It is no surprise that this debate has also confused courts about if and when preclusion should apply to defenses. For example, the Eleventh Circuit suggested that a plaintiff could bar a defendant from raising a defense on claim preclusion grounds if "the same cause of action [was] involved in both cases."[49] Similarly, the Federal Circuit has ruled that a defendant is precluded from a validity defense in a patent infringement case "only if the accused device was 'essentially the same' as the previous device admitted to infringe, or [if] any changes were merely 'colorable.'"[50] The Second Circuit also created its own

44. *Id.* at 353 (emphasis added).

45. Restatement (Second) of Judgments § 18 (AM. L. INST. 1982).

46. In *Lucky Brand*, the Supreme Court ultimately held that defense preclusion is an extension of traditional preclusion doctrines, ruling against Marcel because the two cases lacked an identity of claims. Lucky Brand Dungarees, Inc. v. Marcel Fashions Grp., 140 S. Ct. 1589, 1595 (2020).

47. Amy Coney Barrett, *Procedural Common Law*, 94 VA. L. REV. 813, 830 (2008).

48. *Lucky Brand*, 140 S. Ct. at 1595 n.2.

49. McKinnon v. Blue Cross & Blue Shield, 935 F.2d 1187, 1192 (11th Cir. 1991) (rejecting plaintiff's argument that claim preclusion barred defendant from raising a new defense, because the first and second cases did not involve the same cause of action).

50. Hallco Mfg. Co. v. Foster, 256 F.3d 1290, 1295 (Fed. Cir. 2001).

novel defense preclusion test that permitted preclusion of defenses if certain conditions were met.[51] The Supreme Court decided to address the Second Circuit's approach in *Lucky Brand*.

III. LUCKY BRAND V. MARCEL

Lucky Brand v. Marcel reached the Supreme Court in 2019, but the history between the two parties began nearly two decades prior. Lucky Brand Dungarees ("Lucky Brand") and Marcel Fashion Groups ("Marcel") are fashion retailers known for selling jeans and other apparel.[52] The two parties became adversarial after Marcel sued Lucky Brand for copying its "Get Lucky" mark. This Part provides an overview of the three successive lawsuits that led to the case before the Supreme Court and analyzes the Court's unanimous decision striking down the Second Circuit's defense preclusion test.

A. THE 2001 ACTION

In 1986, Marcel federally registered the mark "Get Lucky."[53] In 1990, Lucky Brand started selling clothes under its "Lucky Brand" mark and other marks containing the word "Lucky," including "Get Lucky."[54] Eventually, Marcel sued Lucky Brand in 2001, alleging unfair competition and trademark infringement.[55] The parties reached a settlement agreement in 2003 wherein Lucky Brand agreed to cease using Marcel's "Get Lucky" mark.[56] In return, Marcel agreed to release any future claims arising out of Lucky Brand's use of its "Lucky Brand" trademark and any other marks "owned, registered and/or used by Lucky Brand . . . as of the date of this Agreement."[57]

As demonstrated by the subsequent lawsuits, the parties had conflicting understandings of the scope of the settlement agreement ("the Release").[58] Marcel believed that the settlement agreement had "only released claims as to *infringement* that occurred *prior* to the 2003 execution of the agreement."[59] Lucky Brand, on the other hand, believed that the settlement agreement had "releas[ed] any claim Marcel may have in the *future* in relation to any *trademark*

51. Marcel Fashions Grp. v. Lucky Brand Dungarees, Inc. (*Marcel II*), 898 F.3d 232, 241 (2d Cir. 2018), *rev'd*, 140 S. Ct. at 1589.
52. *Lucky Brand*, 140 S. Ct. at 1592.
53. *Id.*
54. *Id.*
55. *Id.* at 1592 n.2.
56. *Marcel II*, 898 F.3d at 234.
57. *Id.*
58. *Id.*
59. *Id.* (emphasis added).

registered prior to the execution of the agreement."[60] In other words, Marcel believed that the settlement waived claims of infringement that occurred prior to the agreement, whereas Lucky Brand interpreted the agreement more broadly as applying to any and all infringement claims on marks registered prior to the agreement.

B. THE 2005 ACTION

The amity between the parties was short-lived, as Lucky Brand sued Marcel and its licensees shortly thereafter in 2005 for violating its trademarks.[61] In response, Marcel counterclaimed under two theories. First, Marcel alleged that Lucky Brand had continued to infringe on Marcel's "Get Lucky" mark in violation of the Release.[62] Second, Marcel asserted that Lucky Brand's use of "Get Lucky" and Lucky Brand marks together was confusingly similar to, and therefore likely to cause confusion about the origins of, Marcel's "Get Lucky" mark.[63] Lucky Brand moved to dismiss both of Marcel's claims, arguing that the Release barred Marcel from raising claims regarding marks registered prior to its signing.[64] Rebutting the claim, Marcel argued that the Release did not bar "any claims as to infringing *uses* occurring after execution of the settlement agreement."[65] In other words, consistent with its understanding of the scope of the settlement agreement, Marcel believed that the Release did not bar any claims of infringement that occurred after 2003.[66] After the district court rejected Lucky Brand's argument, Lucky Brand raised the Release again in its answer.[67] However, the case proceeded towards trial, and Lucky Brand never raised this defense again in the course of the 2005 Action.[68]

Ultimately, the 2005 Action resolved in Marcel's favor. As a sanction for discovery misconduct, the district court entered a permanent injunction against Lucky Brand, enjoining it from using the "Get Lucky" mark.[69] On Marcel's second counterclaim, a jury found that Lucky Brand had infringed on Marcel's "Get Lucky" mark by using "Get Lucky" in conjunction with Lucky

60. *Id.* (emphasis added).

61. *Id.*

62. Lucky Brand Dungarees, Inc. v. Marcel Fashions Grp., 140 S. Ct. 1589, 1592 (2020).

63. *Id.*; *see also* Defendants' Answer, Affirmative Defenses, and Counterclaims to Plaintiffs' Complaint at 39, Lucky Brand Dungarees, Inc. v. Ally Apparel Resources LLC, No. 1:05-cv-06757 (S.D.N.Y. July 27. 2005).

64. *Lucky Brand*, 140 S. Ct. at 1592–93.

65. *Marcel II*, 898 F.3d at 235.

66. *See id.* at 234–35.

67. *Id.*

68. *Id.*

69. Marcel Fashions Grp. v. Lucky Brand Dungarees, Inc. (*Marcel I*), 779 F.3d 102, 106 (2d Cir. 2015).

Brand's own marks.[70] Following the verdict, the district court requested that the parties draft a final order.[71] Marcel's counsel proposed language that would have enjoined Lucky Brand from using not only the "Get Lucky" mark, but also Lucky Brand's own marks containing the word "Lucky."[72] Lucky Brand refused to accept Marcel's proposal, forcing Marcel's counsel to remove the language prior to submitting the proposed order to the court.[73] Ultimately, the district court incorporated the parties' submission into its final order, prohibiting Lucky Brand from using Marcel's "Get Lucky" mark.[74] The order also declared that Lucky Brand had infringed on Marcel's mark "by using GET LUCKY, the LUCKY BRAND trademarks, and any other trademarks including the word 'Lucky' after May 2003."[75] Notably, however, the order did not enjoin Lucky Brand from using the latter two marks.[76]

C. THE 2011 ACTION

1. The Parties' Return to District Court

In 2011, Marcel sued Lucky Brand again for continued infringement of Marcel's "Get Lucky" mark.[77] In its complaint, Marcel alleged that Lucky Brand's "post-2010 use of Lucky Brand's own marks—some of which used the word 'Lucky'—infringed Marcel's 'Get Lucky' mark in a manner that . . . was previously found infringing."[78] In response, Lucky Brand filed an answer and motion for summary judgment, without raising the Release defense in either.[79] The district court granted Lucky Brand's motion for summary judgment, finding that Marcel's new claims were precluded by the 2005 Action.[80] However, the Second Circuit disagreed, noting that the alleged infringements occurred after the 2005 Action had already concluded, meaning that preclusion did not apply.[81]

---

70. *Id.*
71. *Id.*
72. *Id.*
73. *Id.*
74. *Id.*
75. *Id.*
76. *See* Lucky Brand Dungarees, Inc. v. Marcel Fashions Grp., 140 S. Ct. 1589, 1593 n.2 (2020) ("The injunction did not enjoin, or even mention, Lucky Brand's use of any other marks or phrases containing the word 'Lucky.'").
77. Marcel Fashions Grp. v. Lucky Brand Dungarees, Inc. (*Marcel II*), 898 F.3d 232, 235 (2d Cir. 2018), *rev'd*, 140 S. Ct. at 1589.
78. *Lucky Brand*, 140 S. Ct. at 1593.
79. *Marcel II*, 898 F.3d at 235.
80. *Id.*
81. *Id.*

On remand, Lucky Brand, represented by new counsel, moved to dismiss the case based on a new argument. For the first time since its motion to dismiss in the 2005 Action, Lucky Brand claimed that the Release foreclosed Marcel's claims.[82] Marcel rebutted this assertion by arguing that claim preclusion applied to Lucky Brand's defense, since it could have been raised in the 2005 Action.[83] The district court sided with Lucky Brand and concluded that the settlement agreement barred Marcel's claims regarding Lucky's use of its own trademarks registered prior to the signing of the Release.[84] The court also rejected Marcel's counterargument, finding that Lucky Brand's defenses were not issue precluded because the Release had not been actually litigated and adjudicated in the 2005 Action.[85] Further, the court reasoned that claim preclusion did not apply, since Lucky Brand was "not asserting a claim against Marcel."[86] In response to the district court's dismissal, Marcel appealed.

2. *The Second Circuit's Novel Doctrine of Defense Preclusion*

On appeal, the Second Circuit vacated the lower court's decision and remanded the case.[87] Agreeing with Marcel's theory of "defense preclusion," the Second Circuit held that claim preclusion barred Lucky Brand's Release defense.[88] Emphasizing that Lucky Brand could have avoided the 2011 Action entirely if it had litigated the Release in the 2005 Action, the Second Circuit created a new "defense preclusion" doctrine.[89] The court held that claim preclusion will prevent a party from raising a defense where:

i. a previous action involved an adjudication on the merits;
ii. the previous action involved the same parties or those in privity with them;
iii. the defense was either asserted or could have been asserted, in the prior action; and
iv. the district court, in its discretion, concludes that preclusion of the defense is appropriate because efficiency concerns outweigh any unfairness to the party whose defense would be precluded.[90]

82. *Id.* at 235–36.
83. *Id.* at 236.
84. *Id.*
85. *Id.*
86. *Id.*
87. *Id.*
88. *Id.*
89. *Id.* at 239–41.
90. *Id.* at 241.

In justifying its new defense preclusion doctrine, the Second Circuit discussed at length the policy rationales justifying preclusion. Noting that the "principles animating the claim preclusion doctrine [do not] disappear when that which is sought to be precluded is a defense," the court emphasized that its new rule furthered the goals of finality and efficiency.[91] The court argued that defense preclusion would not only incentivize defendants to "litigate all their relevant defenses in an initial action," but also prevent them from filing "wasteful follow-on actions that would not have been filed had the defense been asserted (and maintained) at the first opportunity."[92] Further, the court argued that ruling against defense preclusion would make judicial victories vulnerable to "hidden defense[s]" that may later emerge.[93] Had Lucky Brand litigated the Release defense in the 2005 Action, the court reasoned, the parties and judicial system could have avoided "seven-plus years of litigation, involving 179 district court docket entries and two appeals."[94]

The Second Circuit's novel approach differed from other circuits' application of preclusion to defenses. For example, the Eleventh Circuit had held that a plaintiff could only bar a defendant from raising a defense on claim preclusion grounds if "the same cause of action [was] involved in both cases,"[95] an element that is notably absent from the Second Circuit's test. By creating this circuit split, the Second Circuit's new defense preclusion rule was ripe for Supreme Court review.

3. The Supreme Court's Rejection of "Defense Preclusion"

Writing for a unanimous Supreme Court, Justice Sotomayor struck down the Second Circuit's application of defense preclusion. In a narrow holding, the Court ruled that Marcel could not preclude Lucky Brand from raising new defenses in the 2011 Action.[96] The opinion proceeded by first noting that the Court had "never explicitly recognized 'defense preclusion' as a standalone category of res judicata."[97] Based on its precedent (or lack thereof), the Court reasoned, "any . . . preclusion of defenses must, at a minimum, satisfy the strictures of issue preclusion or claim preclusion."[98] The Court immediately ruled out issue preclusion because the 2005 Action did not adjudicate the issue

91. *Id.* at 239.
92. *Id.*
93. *Id.*
94. *Id.*
95. McKinnon v. Blue Cross & Blue Shield, 935 F.2d 1187, 1192 (11th Cir. 1991).
96. Lucky Brand Dungarees, Inc. v. Marcel Fashions Grp., 140 S. Ct. 1589, 1598 (2020).
97. *Id.* at 1595.
98. *Id.* at 1595; *see also Marcel II*, 898 F.3d at 236 n.3.

of the Release.[99] Turning its attention to claim preclusion, the Court held that defense preclusion could only apply if the 2011 Action and 2005 Action shared either the same cause of action or a "common nucleus of operative facts."[100]

Applying this standard, the Court held that the 2011 Action and the 2005 Action did not implicate the same cause of action.[101] Rather, the two cases were "grounded on different conduct, involving different marks, occurring at different times."[102] The Court also emphasized that the conduct triggering the 2011 Action occurred "after the conclusion of the 2005 Action,"[103] relying on the general principle that conduct occurring after the filing of a lawsuit gives rise to a new claim.[104] The Court stated that it would go no further in "determin[ing] when (if ever) applying claim preclusion to defenses may be appropriate, because a necessary predicate—identity of claims—[was] lacking."[105]

In the narrow context of this case, the Court was correct in finding that defense preclusion did not apply to the dispute between Lucky Brand and Marcel. As Justice Sotomayor made clear, preclusion always requires identity of claims and identity of parties between two suits.[106] As noted in Part I, these elements impose an outer limit on preclusion in order to afford all parties a day in court. Thus, the Court was correct to reject defense preclusion in this case, since Marcel's 2011 claims only dealt with Lucky Brand's use of its own marks in isolation, rather than their use in conjunction with Marcel's "Get Lucky" mark.

Despite the Court's sound reasoning in resolving the parties' dispute, the Court more broadly suggested in dicta that "there may be good reasons to question any application of claim preclusion to defenses," even where two cases shared the same cause of action and the same parties.[107] Thus, the Court's dicta left open the possibility of a *per se* rule against defense preclusion. As the remainder of this Note will argue, the Court's reasoning in favor of such a rule is problematic because it overemphasized the fairness concerns that could result from applying preclusion to defenses, while ignoring how such a rule would interact with the available remedies in trademark law.

---

99. *Lucky Brand*, 140 S. Ct. at 1595.; *see also supra* Section I.A.
100. *Lucky Brand*, 140 S. Ct. at 1595.
101. *Id.*
102. *Id.*
103. *Id.* at 1596.
104. *Id.* (citing Whole Woman's Health v. Hellerstedt, 579 U.S. 582 (2016)).
105. *Id.* at 1595 n.2.
106. *See supra* Section I.A (discussing the elements of claim preclusion).
107. *Lucky Brand*, 140 S. Ct. at 1595 n.2.

IV. THE COURT'S UNDUE EMPHASIS ON FAIRNESS

As indicated by their opinions, both the Second Circuit and the Supreme Court heavily emphasized fairness and due process concerns in questioning any application of preclusion to defenses. For instance, the fourth prong of the Second Circuit's defense preclusion test required lower courts to preclude a defense if the "efficiency concerns outweigh any unfairness to the party whose defense would be precluded."[108] This emphasis on fairness carried over to the parties' Supreme Court briefs[109] and Justice Sotomayor's opinion.[110]

However, the Court's fixation on fairness was inconsistent with its prior rulings, since it has traditionally favored efficiency over fairness when applying claim preclusion in cases like *Lucky Brand*. As noted, requiring identity of claims and mutuality of parties already protects the litigants' due process rights and addresses fairness concerns.[111] Once those elements of claim preclusion are met, courts are tasked with evenhandedly throwing out claims or issues, no matter how unfair it may seem, for the sake of efficiency.[112] While the Supreme Court's additional consideration of fairness may be desirable or welcome, the Court's suggestion against *any* application of defense preclusion, even where the parties and claims are identical to those of a prior suit, is questionable in light of preclusion's underlying goal of promoting judicial economy.

A. DISTINGUISHING BETWEEN MUTUAL AND NON-MUTUAL USES OF CLAIM PRECLUSION

The problems arising from the Court's analysis stemmed from its failure to distinguish between mutual and non-mutual uses of claim preclusion. Untangling these different uses is critical to uncovering the different interests at stake when deciding whether preclusion applies, including whether courts should assess any issues of fairness separately.

Before distinguishing between mutual and non-mutual uses of preclusion, it is important to remember that *Lucky Brand* dealt with only claim preclusion. As discussed in Section I.A, issue preclusion prevents parties from re-litigating issues that were actually litigated and decided, even in the context of a

108. Marcel Fashions Grp. v. Lucky Brand Dungarees, Inc. (*Marcel II*), 898 F.3d 232, 241 (2d Cir. 2018), *rev'd*, 140 S. Ct. at 1589.

109. *See* Brief for Respondent at 42–46, *Lucky Brand*, 140 S. Ct. at 1589 (No. 18-1086); Reply Brief for Petitioners at 21–22, *Lucky Brand*, 140 S. Ct. at 1589 (No. 18-1086).

110. Though the Supreme Court avoided explicitly using the words "fairness" or "due process," the Court noted that courts should not use preclusion to punish defendants for failing to raise certain defenses in a prior suit because of practical considerations. *Lucky Brand*, 140 S. Ct. at 1596.

111. *See supra* Section I.A.

112. *See supra* note 14–17 and accompanying text.

completely different claim.[113] By contrast, claim preclusion prevents the litigation of a claim or defense that either was or *could have* been litigated in a prior action.[114] As the Court and parties agreed, "issue preclusion [did] not apply" to this case since prior litigation did not address the issue regarding the Release.[115] Thus, the Court's opinion correctly limited its analysis to the application of *claim* preclusion to defenses.

Despite correctly finding that only claim preclusion was at issue, the Court failed to differentiate between mutual and non-mutual uses of claim preclusion. As the name implies, a mutual use of preclusion occurs where the parties in the subsequent action are identical to or "in privity"[116] with those involved in the previous case.[117] In contrast, non-mutual preclusion occurs when a plaintiff or defendant attempts to use preclusion against a stranger.[118]

Unlike issue preclusion, claim preclusion generally requires mutuality, with a few limited exceptions. The Supreme Court outlined these six limited exceptions in *Taylor v. Sturgell*, which held that claim preclusion could apply against a non-party to the previous suit where: (1) the non-party agreed to be bound by the prior judgment; (2) where a special relationship ("privity") exists between the prior and subsequent litigant; (3) where the party seeking to re-litigate a claim was fully and fairly represented in the previous case; (4) where the non-party somehow controlled the previous litigation; (5) where the second litigant is an agent of the first litigant; and (6) where statutory systems foreclose subsequent litigation by absent parties.[119] In finding exceptions that closely resembled mutuality, the *Taylor* Court emphasized that the common law of preclusion is "subject to due process limitations."[120] By requiring mutuality or one of the six *Taylor* factors, the Court mitigated any fairness and due process concerns that are implicated by allowing parties to assert claim preclusion against a stranger.

The mutuality requirement, along with the six *Taylor* factors, ensures that procedural due process limits are already built into preclusion doctrine. For instance, the *Taylor* Court reasoned:

> A person *who was not a party to a suit* generally has not had a "full and fair opportunity to litigate" the claims and issues settled in that

113. *See supra* Section I.A.
114. *See id.*
115. *Lucky Brand*, 140 S. Ct. at 1595.
116. Privity is defined as having a "substantive legal relationship[]." Taylor v. Sturgell, 533 U.S. 880, 894 n.8 (2008).
117. *See* Trammell, *supra* note 13, at 574.
118. *See id.*
119. *Taylor*, 553 U.S. at 893–95.
120. *Id.* at 891.

> [prior] suit. The application of claim and issue preclusion to nonparties thus runs up against the "deep-rooted historic tradition that everyone should have his own day in court."[121]

Because mutuality ensures that fairness concerns are given due consideration, the Court has rejected any additional fairness issues raised where mutuality was not an issue. For instance, in *Federated Department Stores v. Moitie*, the Ninth Circuit had held that, as a matter of "simple justice," it would be unfair to use res judicata to preclude non-appealing parties whose claims were "closely interwoven" with those of successfully appealing parties.[122] The Supreme Court reversed, holding that the Ninth Circuit's consideration of fairness was "an unprecedented departure from the accepted principles of res judicata."[123] The Court also emphasized that "[t]he doctrine of res judicata serves vital public interests beyond an individual judge's ad hoc determination of the equities in a particular case."[124] Further, any additional fairness considerations would make any application of claim preclusion subject to debate and lead to further litigation, which runs contrary to the purposes of preclusion.[125]

Recognizing the differences between mutual and non-mutual uses of preclusion is critical because it helps courts determine whether to preclude a claim. However, courts and parties often conflate mutual and non-mutual uses of claim preclusion,[126] thereby allowing consideration of policy concerns, like fairness issues, that are outside the scope of the legal question before them and distorting the ultimate outcome as the Court did in *Lucky Brand*.

B. LUCKY BRAND'S FIXATION ON FAIRNESS

The Court's failure to distinguish between mutual and non-mutual uses of preclusion led to it belaboring the fairness concerns implicated by applying claim preclusion to defenses. In its opening brief to the Supreme Court, Lucky Brand argued that the Second Circuit's application of defense preclusion was "fundamentally unfair."[127] It also dedicated an entire section of its reply brief to due process concerns, arguing that precluding Lucky Brand from raising its

---

121. *Id.* at 892–93 (emphasis added).

122. Federated Dep't Stores v. Moitie, 452 U.S. 395, 399–401 (1981).

123. *Id.* at 399.

124. *Id.* at 401; *see also* Heiser v. Woodruff, 327 U.S. 726, 733 (1946) (finding that there is "no principle of law or equity which sanctions the rejection by a federal court of the salutary principle of *res judicata*").

125. *See supra* Section I.A.

126. *See, e.g.*, Bursak, *supra* note 1, at 1656 n.24 (citing cases, like *Int'l Air Response v. United States*, 324 F.3d 1376, 1378 (Fed. Cir. 2003), where parties and courts have shown confusion as to whether claim preclusion or issue preclusion applied).

127. Brief for Petitioners at 46, Lucky Brand Dungarees, Inc. v. Marcel Fashions Grp., 140 S. Ct. 1589 (2020) (No. 18-1086).

Release defense would deprive it of not only the bargain struck in the Release, but also the ability to defend itself against new claims.[128]

The Court seemed receptive to Lucky Brand's fundamental fairness argument. In its opinion, the Court noted that "it [was] no surprise . . . that the Second Circuit held that Marcel's 2011 Action claims were not barred by the 2005 Action. By the same token, the 2005 Action could not bar Lucky Brand's 2011 defenses."[129] Thus, the Court seemed to suggest that courts must apply preclusion doctrine "symmetrically" between the two parties. However, mutuality was not an issue in *Lucky Brand*. The 2001 Action, the 2005 Action, and the 2011 Action all involved the same parties: Lucky Brand and Marcel. When mutuality exists, as it did here, the Court has already found that preclusion should favor efficiency and finality over all other considerations.[130] Thus, considerations of equity between the parties seem out of place in this inquiry.

Even if judicial balancing of equity between the parties had been appropriate, barring Lucky Brand's defense in this case would not have violated any fairness principles. Arguably, the Supreme Court may have emphasized these fairness concerns because Lucky Brand had not yet had its "day in court" on Marcel's *new* claims. Yet the Court seemed to suggest, in dicta, that even if the subsequent suit had involved identical claims, Lucky Brand should still be able to raise its defenses "even though they were equally available and relevant in both actions."[131] In justifying this, the Court noted that "various considerations, other than actual merits, may govern whether to bring a defense, such as the smallness of the amount or the value of the property in controversy, the difficulty of obtaining the necessary evidence, the expense of the litigation, and a party's own situation."[132] Notably, however, Lucky Brand *did* raise the Release defense in the 2005 Action but chose not to pursue it beyond its answer to Marcel's counterclaims. At oral arguments, counsel for Lucky Brand surmised that it had decided to forego the Release defense because it would not have been dispositive to the case.[133] But the *Moitie* Court already rejected a similar argument, finding that it would not recognize a "general equitable doctrine" when conducting its claim preclusion analysis,

128. Reply Brief at 21–22, *Lucky Brand*, 140 S. Ct. at 1589.

129. *Lucky Brand*, 140 S. Ct. at 1596.

130. *See supra* note 14–17 and accompanying text.

131. *Lucky Brand*, 140 S. Ct. at 1595 n.2.

132. *Id.* (citing Cromwell v. County of Sacramento, 94 U.S. 351, 356 (1877)) (internal quotation marks and alterations omitted).

133. Transcript of Oral Argument at 5–6, *Lucky Brand*, 140 S. Ct. at 1589.

especially where the non-appealing parties had "made a calculated choice to forego their appeals."[134]

By attaching significance to these fairness rationales, the Court overemphasized the due process concerns that defense preclusion potentially raises. As noted above, claim preclusion's mutuality requirement already protects the parties' due process rights. The Court carried over these protections to the defense preclusion context when it ruled that "any . . . preclusion of defenses must, at a minimum, satisfy the strictures of issue preclusion or claim preclusion."[135] By separately considering a defendant's due process rights, the Court double-counted the fairness issues and erroneously put a thumb on the scale for Lucky Brand, even though mutuality was met.

V. THE TROUBLING EFFECTS OF LUCKY BRAND COMBINED WITH *EBAY*

Like efficiency, the goal of delivering finality to the parties also drives preclusion doctrine. However, the Court did not adequately consider how a *per se* rule against defense preclusion would threaten finality of decisions in the trademark context. This is especially critical in light of the *eBay* injunction standard, which requires plaintiffs to affirmatively show four factors in order to obtain an injunction. This standard, originally designed for patents, came to replace the automatic injunction rule for trademarks as well, in effect heightening the standard for trademark injunctions. As a result, trademark owners are often now left without a forward-looking remedy to enforce in the future, even when they prevail on their infringement claims today.[136] This Part proceeds by explaining the *eBay* injunction test and arguing that the potential finality issues raised by the Court's dicta in *Lucky Brand*, along with the unique quirks of trademark law, demand revisiting the question of *eBay*'s applicability to trademark infringement cases.

A. EBAY'S INJUNCTION STANDARD

Prior to *eBay*, both patent and trademark courts almost always granted an injunction once infringement was shown.[137] Under this rule, courts used to

134. Federated Dep't Stores v. Moitie, 452 U.S. 394, 400 (1981).

135. *Lucky Brand*, 140 S. Ct. at 1595.

136. *See infra* Section IV.B.

137. *See* eBay Inc. v. MercExchange, L.L.C., 547 U.S. 388, 391 (2006) (rejecting the Federal Circuit's "general rule that courts will issue permanent injunctions against patent infringement absent exceptional circumstances"). Though *eBay* was a patent case, the *eBay* four-factor test has been extended to copyright and trademark infringement cases as well. Lemley, *supra* note 8, at 1798–1800.

issue injunctions automatically if a plaintiff prevailed on their infringement claims, unless there were "exceptional circumstances."[138]

However, courts and skeptics began to question the automatic injunction rule,[139] at least in patent cases. They argued that injunctions were an extremely harsh remedy that prevented not only the infringer's harmful activities, but also non-infringing activities that could be beneficial to the public.[140] This problem became more pronounced with the rise of non-practicing entities, or "patent trolls,"[141] in the early 2000s.[142]

In an attempt to deter suits brought by non-practicing entities, the Supreme Court rejected the automatic injunction rule in *eBay v. MercExchange*.[143] The Court found that the automatic injunction rule was too mechanical and rigid in light of the extraordinary property rights that came with an injunction. In so holding, the Court held that a patentee seeking a *permanent* injunction must first satisfy a four-factor test by demonstrating:

(1) that it has suffered an irreparable injury (the "irreparable harm" prong);

---

138. *eBay*, 547 U.S. at 392.

139. The central inquiry of this debate is whether the "property rule" or "liability rule" is best suited to remedy infringement cases. The "property rule" refers to the use of a property law remedy that involves deciding "who is to be given an initial entitlement." Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1092 (1972). In the trademark context, the "property rule" allows plaintiffs to keep the entitlement through an injunction unless and until they agree to a selling price. By contrast, the "liability rule" refers to courts "requiring transgressors to pay victims for harms suffered." Louis Kaplow & Steven Shavell, *Property Rules Versus Liability Rules: An Economic Analysis*, 109 HARV. L. REV. 713, 715 (1996). In the trademark context, courts can employ the "liability rule" by ordering the defendant to pay damages to the wronged plaintiff.

140. *See* Mark A. Lemley & Philip J. Weiser, *Should Property or Liability Rules Govern Information?*, 85 TEX. L. REV. 783, 784 (2007) (finding that injunctions that are not "well tailored to the scope of the property right at issue but necessarily restrain the use of property not owned by the plaintiff . . . can overwhelm the benefits of property rules in enforcing legal rights"). *See generally* Calabresi & Melamed, *supra* note 139 (suggesting that a property rule is desirable where there are few parties and low transaction costs).

141. *See* Michael S. Mireles, *Trademark Trolls: A Problem in the United States?*, 18 CHAP. L. REV. 815, 819 (2015) (defining a patent troll as an owner of a patent that does not actually practice the patented invention but taxes others who do).

142. James Bessen, *The Evidence Is In: Patent Trolls Do Hurt Innovation*, HARV. BUS. REV. (Nov. 2014), https://hbr.org/2014/07/the-evidence-is-in-patent-trolls-do-hurt-innovation (finding that patent trolls, without making a useful product of their own, create significant harm by "cost[ing] innovators money" and forcing them to "respond by investing less in R&D").

143. *See eBay*, 547 U.S. at 396–97 (Kennedy, J., concurring) (suggesting that injunctions may be appropriate when a patent owner uses the patent "not as a basis for producing and selling goods but, instead, primarily for obtaining licensing fees").

(2) that the remedies available at law, such as monetary damages, are inadequate to compensate for the injury (the "inadequate remedies at law" prong);
(3) that, considering the balance of hardships between the plaintiff and defendant, a remedy in equity is warranted; and
(4) that the public interest would not be disserved by a permanent injunction.[144]

Because non-practicing entities could not prove irreparable injury (since they do not actually practice their patents), the Court's four-factor test was successful in preventing "patent trolls" from winning broad rights to exclude and exercising that leverage in licensing negotiations.[145]

Although the *eBay* test was originally limited to injunctions in patent infringement suits, courts began applying it to other intellectual property disputes. In *Herb Reed*, the Ninth Circuit extended the *eBay* test to trademark infringement suits and held that trademark owners must establish "the likelihood of irreparable harm" before obtaining a *preliminary* injunction.[146] In the same breath, the Ninth Circuit also reiterated that the *eBay* test applied to permanent injunctions in the trademark context.[147] Thus, over the span of two cases in less than ten years, the trademark landscape changed from one in which injunctions were nearly guaranteed upon a showing of infringement[148] to one where they have become more rare due to the difficulty of showing the first two *eBay* factors.[149]

B. LUCKY BRAND'S DISMISSAL OF FINALITY CONCERNS

In light of *eBay* and *Herb Reed*, a *per se* rule against preclusion applied to defenses could potentially lead to absurd results for plaintiffs who prevail in a

144. *Id.* at 391 (majority opinion).

145. *See* Lemley, *supra* note 8, at 1796 ("[T]he results [of *eBay*] in patent cases have been salutary as a policy matter, allowing courts to deny injunctions to non-practicing entities that sought injunctions only to hold up defendants while still granting injunctions to those who need it.").

146. Herb Reed Enters., LLC v. Fla. Ent. Mgmt., 736 F.3d 1239, 1249 (9th Cir. 2013) (citations omitted); *see also* Ferring Pharm., Inc. v. Watson Pharm., Inc., 765 F.3d 205, 206 (3d Cir. 2014) (applying *eBay* to Lanham Act cases in the Third Circuit); Voice of the Arab World, Inc. v. MDTV Med. News Now, Inc., 645 F.3d 26, 33 (1st Cir. 2011) (applying *eBay* to Lanham Act cases in the First Circuit).

147. *Herb Reed*, 736 F.3d. at 1249.

148. Lemley, *supra* note 8, at 1797 n.19 (citing to circuit court cases that enforced the automatic injunction rule upon a showing of trademark infringement or, at the very least, automatically established "a likelihood of success on the merits and irreparable harm").

149. *See* Lemley, *supra* note 8, at 1806 (arguing that trademark owners have difficulty providing evidence of irreparable injury and that "the concepts of irreparable injury and the lack of an adequate remedy at law are inextricably intertwined").

trademark infringement case. Though allowing new defenses would not necessarily be dispositive to the outcome of an infringement case, it could still pose a risk to the litigation or even to the validity of the trademark itself. As Marcel properly noted in its brief, finality and efficiency are the cornerstones of preclusion doctrine.[150] Marcel further reasoned that applying preclusion to defenses would help "protect[] the finality of judgments," emphasizing that a "losing litigant deserves no rematch after a defeat fairly suffered."[151] In addition, Marcel argued that allowing Lucky Brand to raise the Release defense in response to the 2011 Action would undermine the judgment of the 2005 Action, allowing Lucky Brand to "flout prior judicial determinations of the parties' rights, in hopes of obtaining a more favorable result in a subsequent lawsuit."[152]

In an attempt to quickly dismiss Marcel's finality concerns, the Court noted that the Release defense in the 2011 Action did not "imperil the judgment of the 2005 Action because the lawsuits involved both different conduct and different trademarks."[153] Even outside the context of the parties' case, the Court also argued that defense preclusion was superfluous, since plaintiffs could avoid facing new defenses that threaten a prior judgment by obtaining "supplemental relief in the original action."[154] Thus, the Court reasoned that a per se rule against defense preclusion would have no impact on the finality of prior judgments.

However, the Court's recommendation of obtaining "supplemental relief" is unrealistic since prevailing trademark plaintiffs often do not have an injunction to enforce. The new *eBay* standard for injunctions has left successful trademark plaintiffs without injunctions, and thus without any relief, to enforce at a future date. Between January 1, 2009 (the first date for which data is available on Lex Machina), and October 19, 2020, about 57,000 federal trademark cases were filed.[155] Of the 1,293 cases decided on the merits in favor of the plaintiff,[156] only 722 permanent injunctions were granted.[157] These

150. Brief for Respondent at 22–24, Lucky Brand Dungarees, Inc. v. Marcel Fashions Grp., 140 S. Ct. 1589 (2020) (No. 18-1086).

151. *Id.* at 5.

152. *Id.* at 37.

153. *Lucky Brand*, 140 S. Ct. at 1595.

154. *Id.* at 1597 n.3.

155. This data was obtained using Lex Machina after selecting "Trademark" for "Case Types."

156. This includes cases decided on the pleadings, at summary judgment, after trial, or as a matter of law.

157. This data was obtained using Lex Machina after selecting "Trademark" for "Case Types." Notably, damages awards in trademark infringement suits generally tracked trends for

figures suggest that over 40% of prevailing plaintiffs are left without an injunction to enforce against defendants who continue infringing. This statistic is particularly striking given that less than 3% of prevailing plaintiffs win any monetary damages.[158] Indeed, there have been cases in which trademark owners won their case but were awarded neither damages nor an injunction.[159]

The specific facts of *Lucky Brand* also highlight the futility of the Court's "solution." As noted in Section II.B, Marcel could not, despite its best efforts, obtain an injunction in the 2005 Action regarding Lucky Brand's uses of its own trademarks in conjunction with the "Get Lucky" mark. Thus, Marcel could never have availed itself of the Supreme Court's suggested course of obtaining supplemental relief, even if the claims in the 2005 Action and 2011 Action had been identical.

Such a result, however, is not unique to Marcel. As a result of *Lucky Brand*'s dicta, the failure to obtain an injunction, whether due to the heightened injunction standard under *eBay* or a lack of bargaining power (which appeared to be the case with Marcel) could make future plaintiffs vulnerable to new defenses.[160] Thus, rethinking *eBay*'s applicability to trademark law may help get ahead of the problem raised by the Court's dicta in *Lucky Brand*.

C. TRADEMARK'S QUIRKS AND THE RESULTING INCOMPATIBILITY WITH EBAY

A prevailing plaintiff in a trademark infringement suit typically has a menu of options when it comes to seeking a remedy. These options can be simplified into two broad categories: (1) injunctions, including both preliminary and permanent injunctions, and (2) monetary damages, which includes the plaintiff's lost profits, disgorgement of the defendant's profits as a result of infringement, and attorney's fees.[161]

Lucky Brand's dicta and its potential impact on the finality of trademark decisions call for a more searching inquiry into the current injunction standard. In addition to exacerbating potential finality issues if future courts refuse to

---

all cases with damages awards. This suggests that courts were not awarding higher damages in trademark cases to make up for the lack of injunctions.

158. Lemley, *supra* note 8, at 1808 n.79.

159. *See id.* at 1808–09 n.80 (citing two federal court cases in which the court failed to award any type of remedy to the prevailing plaintiffs).

160. *See, e.g.*, Response Brief of Appellees Phh Mortg. Corp. at 13–14, Davis v. Ocwen Loan Servicing, LLC, No. 20-10547 (5th Cir. Sept. 2, 2020) (arguing that a statute of limitations defense against foreclosure could have been raised in a prior action); Reply Brief of Petitioners at 4–5, Estate of Steinbeck v. Kaffaga, 141 S. Ct. 236 (2020) (No. 19-1181) (relying on *Lucky Brand* to raise a validity defense regarding a 1983 agreement splitting royalties on John Steinbeck's work).

161. Lanham Act §§ 34–35, 15 U.S.C. §§ 1116–1117.

apply preclusion to defenses, the *eBay* test is fundamentally incompatible with the unique purposes and characteristics of trademark law. This incompatibility, at its core, is due to trademarks being different from other types of intellectual property, such as patents and copyrights, in key areas.

The differences between trademarks and other intellectual property can be summarized into four points. First, injunctions issued upon a showing of infringement always help further the ultimate purpose of trademark law, which is to improve the quality of information in the marketplace. This is not always true in patent or copyright law, since some amount of infringement and access to information may be beneficial for more innovation and creation. Second, the heightened injunction standard under *eBay*/*Herb Reed* often leaves prevailing trademark owners without any remedy due to the inherent difficulties of accurately assessing trademark damages. Third, establishing the first prong of the *eBay* test, the "likelihood of irreparable harm," is more challenging in trademark infringement suits. Lastly, *eBay* is fundamentally incompatible with trademark law because it was meant to solve a problem that does not exist as applied to trademarks.

1. Injunctions Further the Underlying Purpose of Trademark Law

At its core, trademark law serves to relay information about the nature and quality of goods to buyers, thereby reducing consumer transaction costs.[162] Thus, unlike patent or copyright law, trademark law was never intended to reward "the fruits of intellectual labor" or to incentivize creation.[163] Lending support to this theory, federal trademark law, unlike the patent and copyright regimes, is constitutionally grounded in the Commerce Clause, not the Intellectual Property Clause.[164]

Due to these differences, injunctions better serve the unique policy goals underlying trademark law than they do for patent or copyright law. Consumers, as the "intended beneficiaries" of trademark law, should enjoy the benefits of a remedy once infringement is found.[165] An injunction fulfills this role by preventing the defendant from using the infringing mark, thereby directly

162. *See* Mark A. Lemley, *The Modern Lanham Act and the Death of Common Sense*, 108 YALE L.J. 1687, 1688 (1999) ("Trademarks are a compact and efficient means of communicating information to consumers."). *But see* Mark P. McKenna, *The Normative Foundations of Trademark Law*, 82 NOTRE DAME L. REV. 1839, 1843 (2007) (arguing that, traditionally, "the mark was only an instrumentality of the producer's business, which was the ultimate object of protection").

163. Trade-Mark Cases, 100 U.S. 82, 93–94 (1879).

164. U.S. Pat. & Trademark Off., *Introduction*, TRADEMARK MANUAL OF EXAMINING PROCEDURE (2013), https://tmep.uspto.gov/RDMS/TMEP/Oct2013#/Oct2013/introductiond1e1.html.

165. Lemley, *supra* note 8, at 1809.

reducing the potential for consumer confusion in the future. If a court were to find infringement without granting an injunction, "[c]onsumers [would] continue to make purchases under false pretenses and either blame the trademark owner or become less likely to trust anything in the market thereafter."[166]

In contrast, monetary damages are ill-suited for compensating the harms caused by trademark infringement.[167] For instance, no amount of damages can "unconfuse" consumers once they are confused by an infringer, since the harm has already been done. Additionally, courts are often unable to accurately calculate the exact damage done to a company's goodwill due to the infringement.[168] As a result, damages will often be highly speculative, either under- or over-compensating prevailing plaintiffs. Even if a court were to accurately award monetary damages (without enjoining the infringing activity), the public would stand to lose as the defendant continued to confuse consumers.[169]

Moreover, the use of money damages actually runs *contrary* to the underlying policies served by trademark law. Because damages awards go directly to the trademark owner, these awards suggest that an infringer is compensating the trademark owner, either for lost profits or as a reasonable royalty for using the mark. Granting money damages thus supports the view that trademark law protects the owner's interest in protecting the value of their mark, not the interests of consumers. Although subject to debate, some scholars argue that this is inconsistent with trademark's purpose because it monopolizes distinct product features, thereby conferring some amount of market power upon the owner, while only maintaining a "tenuous relationship to consumer deception."[170]

---

166. *Id.* at 1810.

167. Although damages could deter a defendant from re-infringing, there are two counterarguments to this point. First, damages serving a deterrence function only address the potential harm of future confusion, not the confusion that has already occurred and was the focus of an infringement action. Second, courts do not appear to have significantly increased damages awards in trademark infringement cases to make up for the difficulty trademark owners face in obtaining injunctions. *See supra* note 157.

168. Lemley, *supra* note 8, at 1803.

169. *Id.* at 1810.

170. Glynn S. Lunney Jr., *Trademark Monopolies*, 48 EMORY L.J. 367, 372 (1999). *But see* McKenna, *supra* note 162, at 1844–45 (arguing that trademark law serves an incentive function for firms to invest in their products and services, similar to patent and copyright law); William M. Landes & Richard A. Posner, *Trademark Law: An Economic Perspective*, 30 J.L. & ECON. 265, 271 (1987) ("An entirely different benefit of trademark protection derives from the incentives that such protection creates to invest resources not in maintaining quality but in investing new words.").

The benefits of injunctions in the trademark context become more pronounced when compared to their use in patent and copyright cases. While trademarks are only useful when they are "used to designate a single brand,"[171] patents and copyrights prevent the use of non-rivalrous, informational goods. As a result, patent and copyright protections come at a cost, since they artificially limit the public's access to new inventions and creative works. To avoid creating too many access barriers, patent and copyright law try to balance the countervailing interests of incentivizing creation and providing access to protected works.[172] For example, when providing access is optimal, courts may prefer awarding damages, since an injunction would restrict access and curb further creation. Thus, damages in this context can arguably help further the goals of patent law by allowing the infringing (yet potentially useful) product to exist, thereby encouraging more creation while compensating the patent owner.[173]

By contrast, there is no parallel balancing act in trademark law. As discussed above, trademarks are only concerned with preventing consumer confusion. Confusion-based infringement immediately harms consumers in a manner that can only be remedied by enjoining the defendant's conduct. Despite injunctions best serving the policy rationales motivating trademark law, *eBay* has created barriers preventing trademark owners from enjoining infringing use. Thus, the only other remedy trademark owners have is damages, which often do not further the public's interest in preventing consumer confusion and reducing transaction costs.

2. *A Heightened Injunction Standard Leaves Trademark Plaintiffs Without a Remedy*

Injunctions are not only uniquely suited to serve the underlying purposes of trademark law but are also administratively preferable. Under the Lanham Act, a prevailing plaintiff can recover "(1) defendant's profits, (2) any damages sustained by the plaintiff, and (3) the costs of the action."[174] Despite the different ways of calculating monetary damages, courts should prefer injunctions in theory, since monetary damages are difficult to calculate.

Calculating accurate damages based on either the defendant's profits or the plaintiff's lost profits is difficult since the calculation relies on a counterfactual

171. Landes & Posner, *supra* note 170, at 274.

172. PETER S. MENELL, MARK A LEMLEY, ROBERT P MERGES & SHYAMKRISHNA BALGANESH, INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE: 2020 at 514 (2020). Though this source only makes this assertion in the context of copyright law, it can be applied to patent law as well.

173. Lemley, *supra* note 8, at 1809.

174. Lanham Act § 35, 15 U.S.C. § 1117.

world in which infringement never occurred.[175] When calculating defendant's profits, for example, the court must determine defendant's sales less the sales that would have occurred without the infringement. Showing plaintiff's lost profits is equally difficult because it assumes not only a counterfactual of non-infringement but also "proof of actual, not merely likely, confusion."[176] Additionally, unlike in patent cases where courts can calculate damages based on a reasonable licensing fee, trademark cases require courts to calculate intangible harms, such as consumer confusion and "damage to goodwill."[177]

Due to the speculative nature of damages in the trademark context, courts and plaintiffs may resort to injunctions as a remedy.[178] However, as discussed in Section IV.B, the heightened standard for injunctions has left prevailing trademark owners without any other recourse.[179] Thus, in mechanically applying *eBay* to trademark cases, courts have failed to create an incentive structure that encourages trademark owners to file infringement cases and police unlawful uses of their marks.

3. *The Difficulty of Showing "The Likelihood of Irreparable Harm" in Trademark Infringement Suits*

Notwithstanding the practical and normative considerations of trademark law, establishing "the likelihood of irreparable harm," the first prong of the *eBay* test, is particularly challenging in trademark cases.[180] This difficulty arises because of the similarities between the "irreparable harm" and "inadequate remedies at law" prongs of the *eBay* test. Black's Law Dictionary defines "irreparable injury" as "[a]n injury that cannot be adequately measured or compensated by money."[181] Under this definition, the first prong appears identical to the second prong of *eBay*'s four-factor test, which requires plaintiffs to demonstrate "that remedies available at law, such as monetary damages, are inadequate to compensate for that injury."[182] Because the first two prongs are identical, they "tend to stand or fall together."[183]

---

175. *See* Lemley, *supra* note 8, at 1802 ("[I]t is very hard to reconstruct what would have happened in the but-for world in which infringement did not occur.").

176. *Id.* at 1807.

177. *Id.* at 1802 & n.49, 1803.

178. *Id.* at 1806 ("Trademark law is . . . much more reluctant to award damages than any other area of IP law.").

179. *See supra* Section IV.B (explaining how *eBay* has made it more difficult for trademark owners to obtain injunctions).

180. This is not to say that making such a showing is a simple matter in patent or copyright infringement cases.

181. *Injury*, BLACK'S LAW DICTIONARY (11th ed. 2019).

182. eBay, Inc. v. MercExchange, L.L.C., 547 U.S. 388, 391 (2006).

183. Lemley, *supra* note 8, at 1802.

Proving that an injury is irreparable or incapable of being adequately compensated by money damages is generally difficult in intellectual property disputes, which are "business cases."[184] Because infringement suits usually occur between business entities, the resulting harm is usually economic in nature.[185] Thus, compared to plaintiffs in tort cases, who can easily assert that money damages are ill-suited to compensate for death or bodily harm, trademark plaintiffs (and patentees for that matter) often find it difficult to assert that some amount of monetary damages would fail to make them whole.[186]

Despite this common challenge across all IP infringement cases, trademark cases face a unique set of challenges due to the difficulties in quantifying the harms caused by trademark infringement. As noted above, courts have difficulty in measuring the harm caused by infringement because customers are often not even aware that they are confused.[187] However, at the same time, courts cannot automatically assume that all of an infringer's sales are a result of the confusion.[188] Thus, measuring any harm incurred by trademark holders in terms of monetary damages is a challenging endeavor, since it is impossible to accurately evaluate the damage to the plaintiff's reputation and goodwill.[189]

4. eBay *is a Remedy for a Non-Existent Problem in Trademark Law*

Lastly, *eBay* was designed to solve a problem that does not exist in the trademark context: trolls. The Supreme Court ultimately decided *eBay* to address the increasing number of patent infringement suits filed by patent trolls. Joined by Justices Stevens, Souter, and Breyer, Justice Kennedy's concurrence noted that the patent industry had become rampant with firms that were not concerned with "producing and selling goods but, instead, primarily [with] obtaining licensing fees."[190] While practitioners and scholars alike have lauded *eBay* for affording courts more flexibility to deny injunctive relief requested by non-practicing entities, *eBay* has been less successful in the trademark context because a parallel "trademark troll" problem does not exist.[191]

---

184. *Id.*

185. *See id.* ("I doubt there are very many circumstances in which an IP owner whose rights had been infringed wouldn't be satisfied with *some* amount of money.").

186. *Id.*

187. *Id.* at 1803.

188. *Id.*

189. *Id.*

190. eBay, Inc. v. MercExchange, L.L.C., 547 U.S. 388, 396 (2006) (Kennedy, J., concurring).

191. *See generally* Mireles, *supra* note 141 (explaining why the Lanham Act has prevented, and will likely continue preventing, the development of "trademark trolls").

Unlike patent law, trademark law requires and incentivizes ongoing usage to maintain exclusive rights in a mark. The Lanham Act's "use in commerce" requirement for eligible marks primarily guards against trademark trolls. As the name suggests, the "use in commerce" requirement instructs applicants to certify that they are either using their marks in commerce or intend to do so at a future date.[192] This step in the application process requires trademark owners to actually use their mark, unlike patent trolls that do not actually practice their patent while taxing others who do.[193] Additionally, the Lanham Act provides additional safeguards by requiring trademark owners to submit affidavits every ten years to confirm the mark's use in commerce.[194] Once an owner ceases to use a mark in commerce, trademark law deems the mark "abandoned."[195] Non-use thus creates a defense against infringement claims, thereby preventing a hold-up problem in which non-practicing entities either force others to license their mark or prohibit use. Given these protections already built into the Lanham Act, *eBay* is less applicable in the trademark context.

VI. RECONCILING *LUCKY BRAND* THROUGH THE AUTOMATIC INJUNCTION RULE

Though fundamentally a civil procedure case, *Lucky Brand* casts further doubt on the mechanical application of *eBay* to trademark cases and underscores the need to reform the injunction standard for trademark infringement cases. While *eBay* was undoubtedly valuable in preventing frivolous patent infringement litigation, its extension to trademark law was short-sighted[196] Due to the widespread judicial application of *eBay*, the Supreme Court's denial of certiorari in *Herb Reed*, and the Court's apparent willingness to abandon defense preclusion, the best solution to address the problems raised by this Note may be a legislative one. In order to strengthen the trademark regime and avoid the potential issues created by the Court's dicta in *Lucky Brand*, Congress should overturn *Herb Reed* and carve out a trademark exception to *eBay*'s four-factor test. In its stead, Congress should revive the automatic injunction rule for plaintiffs who prevail on the merits in consumer confusion cases.

192. Lanham Act § 1, 15 U.S.C. § 1051.

193. *See* Mireles, *supra* note 141, at 819, 828 (explaining troll characteristics and the use requirement).

194. Lanham Act § 8, 15 U.S.C. § 1058.

195. Lanham Act § 45, 15 U.S.C. § 1127.

196. *See, e.g.*, DAN L. BURK & MARK A. LEMLEY, THE PATENT CRISIS AND HOW THE COURTS CAN SOLVE IT 139 (2009) (describing the results of *eBay* as "encouraging" at the district court level).

Scholars and legislators have offered a number of alternative suggestions to reinforce the carrots and sticks of trademark law in light of *eBay*, but they either fall short or ignore the inherent shortcomings of the *eBay* rule. For instance, some experts have proposed overturning *eBay* in its entirety for all intellectual property cases.[197] However, overruling *eBay* would undermine its success in stopping frivolous patent suits filed by non-practicing entities. Others have suggested using *eBay*'s four factors as guiding questions for consideration, rather than as mandatory elements for plaintiffs to affirmatively prove to enjoin an infringer's use.[198] Though using *eBay*'s four-pronged test as guiding factors rather than as elements would undoubtedly ease the burden on plaintiffs, this solution overlooks the fundamental flaws of the *eBay* factors. As outlined in Section III.B, the first two prongs of *eBay* are extremely difficult to prove in the context of trademark law. The fourth prong, requiring that the public interest not be disserved by an injunction, is also redundant, since an injunction will almost always serve the public interest by protecting consumers from confusion once there has been a finding of trademark infringement. In a further attempt to reform this area of the law, Congress passed the Trademark Modernization Act as part of the COVID-19 relief bill.[199] One of the provisions, effective as of December 27, 2020, amended the Lanham Act and restored a "rebuttable presumption of irreparable harm upon a finding of a violation."[200] Again, the same issues that exist under *eBay* would persist even if plaintiffs won a rebuttable presumption of irreparable harm, since they would still have to affirmatively show the other three prongs of the *eBay* test. Despite the shortcomings of the Trademark Modernization Act, it nevertheless reflects Congress's appetite to reform this area of the law.

In contrast to the proposals discussed above, creating an automatic injunction rule for trademark infringement cases would make for better policy given the purposes of trademark law and practical judicial considerations, without compromising fairness. First, such a rule would better align with trademark law's purpose of consumer protection. A plaintiff's victory in a trademark infringement suit suggests that the plaintiff has met all the elements

197. *See, e.g.*, Gene Quinn, *Why* eBay v. MercExchange *Should, But Won't, Be Overruled*, IP WATCHDOG (Feb. 16, 2020), https://www.ipwatchdog.com/2020/02/16/ebay-v-mercexchange-wont-overruled/id=118929/ (arguing that "Congress should explicitly overrule *eBay*).

198. *See, e.g.*, Lemley, *supra* note 8, at 1812–14 (advocating against a "rote application of the *eBay* factors from interfering with the legitimate purposes of trademark law").

199. Neil Friedman, Gina Kim, Ryan McGonigle & R. Kent Roberts, *Trademark Modernization Act of 2020 to Strengthen Trademark Owners' Rights*, JD SUPRA (Jan. 25, 2021), https://www.jdsupra.com/legalnews/trademark-modernization-act-of-2020-to-7775539/.

200. Trademark Modernization Act of 2020, H.R. 6196, 116th Cong. § 6(a) (2020) (amending Lanham Act § 34(a), 15 U.S.C. § 1116(a)).

of showing infringement, including the likelihood of confusion. Thus, once the court makes such a finding, an injunction should automatically follow to immediately cease any further use that could cause consumer confusion.

Second, a reliable remedy under an automatic injunction rule will incentivize trademark owners to initiate litigation in the first place. Companies, especially those less well funded, will be more willing to bear the costs of litigation and file a trademark infringement suit if they know that a remedy will almost certainly follow a favorable outcome. This is especially important considering the limited remedies that trademark owners currently have at their disposal, as demonstrated by the small number of cases in which courts grant monetary damages.[201] Thus, an automatic injunction rule will help re-align trademark law's incentive structure with its purposes by encouraging trademark owners to not only prosecute unfair uses of their marks but also police activity that could potentially harm consumers and the market.

Finally, and most central to the focus of this Note, an automatic injunction rule best protects trademark owners facing continued infringement in a world without defense preclusion, as the Supreme Court seemed to favor in *Lucky Brand*. Plaintiffs who find themselves in such circumstances will be able to seek a judgment-enforcement action per the Court's suggestion, rather than filing a new infringement suit and making their marks vulnerable to new defenses. Had Marcel received an injunction following the 2005 Action to enjoin Lucky Brand's infringing activity, the small Floridian retailer would have had a better chance of stopping future infringement. Perhaps Lucky Brand would have also thought twice before continuing its infringing activity. Instead, the parties returned to court in 2011 and litigated their case all the way to the Supreme Court, only to find themselves back in district court with the case stayed indefinitely after Lucky Brand commenced bankruptcy proceedings.[202] The prolonging of the 2011 Action likely did not do either party any favors, and both Lucky Brand and Marcel would have benefited from an order enjoining all infringing activity from the 2005 Action.

While an automatic injunction rule may seem overly punitive, especially to innocent infringers, the current regime under *eBay* has proven to be unworkable for the purposes of trademark law. The Court's dicta in *Lucky Brand* further highlights the impracticability of the four-factor injunction test. If the Court believes that litigants are better off without preclusion of defenses,

201. *See* Lemley, *supra* note 8, at 1808 n.79 (finding that less than 3% of prevailing plaintiffs win any monetary damages).

202. Memo Endorsed on Notice of Suggestion of Pendency of Bankruptcy and Automatic Stay of Proceedings and Request for Stay of the Full Proceedings, Marcel Fashions Grp. v. Lucky Brand Dungarees, Inc., No. 1:11-cv-05523 (S.D.N.Y. July 7, 2020).

then the law must at least allow prevailing trademark owners to easily obtain injunctions and secure their marks against future infringement without opening themselves up to new defenses.

VII. CONCLUSION

Despite ruling in favor of Lucky Brand, the Supreme Court did not dispose of the dispute between the two parties. Since neither the Second Circuit nor the Supreme Court ruled on the merits of the Release defense, the parties still need to litigate the scope of the settlement agreement to determine whether it bars Marcel's claims. Thus, the twenty-year saga between the two parties will continue—a result that seems to run contrary to the normative underpinnings of preclusion doctrine.

Though the impact of the Supreme Court's decision on the outcome of the case is yet to be determined, *Lucky Brand* still provides a number of takeaways from a litigation perspective. Most obviously, it serves as a lesson for trademark owners to seek broad injunctions after a favorable judgment to better protect their marks in cases of future infringement. Had Marcel stood its ground and enjoined Lucky Brand's use of its own trademarks, it may have avoided the 2011 Action and the Release defense altogether.

Less obviously, *Lucky Brand* sheds light on the fundamental issues regarding *eBay*'s application to trademark law. *Lucky Brand* may provide a basis for courts to reject any application of defense preclusion, forcing prevailing trademark plaintiffs facing continued infringement to either silently accept a competitor's use of their marks or file a new case and risk opening themselves up to new defenses. As this Note argues, courts and legislators will need to seriously consider the impact such a rule may have in the trademark context, where continued infringement between repeat players is common and injunctions under the *eBay* standard are few and far between.

A NEW SCA FRAMEWORK TO UNCLOTHE THE BUSINESS MODEL ATTACK

Meet Mehta†

I. INTRODUCTION

The Stored Communications Act (SCA) is one of the main statutory structures protecting user communications in the online realm. The SCA prevents specific types of online service providers from voluntarily disclosing their subscribers' communications or being compelled to do so, except in very limited circumstances.[1] As online services have steadily grown, becoming essential and ever-prevalent aspects of our daily lives, so has the importance of the SCA. Today, with many of our communications and interactions happening over social media and other online services, it is essential that the SCA is there to protect our information.

Although the importance of the SCA to protecting user privacy is evident, courts have often found it difficult to apply to modern services and technologies.[2] The SCA was enacted in 1986, before the arrival of the modern internet, and is thus severely outdated. It relies on simplistic service classifications (i.e., communication services vs. storage services) that no longer apply in order to create a hierarchy of communication protections. Courts have therefore "struggled to analyze problems involving modern technology within the confines of [the SCA] framework" and have called on Congress to update the SCA.[3] The SCA, however, has not yet been updated, leaving judges to wade through a "confusing and uncertain area of the law."[4]

Perhaps due to the fact that courts have found it frustrating to apply the SCA to modern technologies, courts and, to a large extent, the legal community as a whole, have often operated under the premise that social media companies are regulated by the SCA.[5] This assumption, however, was most recently challenged in *Facebook v. Superior Court*, in which a criminal defendant seeking

---

DOI: https://doi.org/10.15779/Z38FT8DK74

† J.D. 2022, University of California, Berkeley, School of Law.

1. *See* 18 U.S.C. §§ 2702–03.
2. *See* Facebook, Inc. v. Superior Ct. of San Diego Cnty., 471 P.3d 383, 406–07 (Cal. 2020) (Cantil-Sakauye, C.J., concurring).
3. *Id.* at 407 (quoting Konop v. Hawaiian Airlines, 302 F.3d 868, 874 (9th Cir. 2002)).
4. *See id.* (quoting *Konop*, 302 F.3d at 874).
5. *See id.* at 412 (Cuéllar, J., concurring).

to subpoena Facebook claimed that Facebook did not qualify for the SCA's disclosure protections.[6]

In *Facebook*, the defendant advanced the business model theory, which argues that a company's business model of mining, analyzing, and sharing user content with third parties to facilitate targeted advertisements places the company outside of the disclosure provisions of the SCA.[7] According to the defendant, Facebook used such a business model and therefore contravened the SCA's explicit prohibition against the voluntary disclosure of user communications.[8] Thus, the defendant argued, Facebook could not use the SCA as a shield to protect itself from complying with a subpoena requesting user communications.[9] Though the court noted the potential importance of the business model theory and recognized that it deserved greater exploration, it did not adjudicate the issue; instead, it disposed of the case on other grounds.[10]

A deeper look in this Note, however, reveals that the business model theory is an impotent attack on social media services. Due to the communicative nature of social media and the industry trend of using advertisement-based business models that neither sell user information nor analyze user communications,[11] the business model attack falls flat: that is, it does not effectively change the application of the SCA's disclosure protections. Nonetheless, due to the sheer number of different social media services and business models available, courts should still implement business model scrutiny and engage in a fact-specific, case-by-case analysis when determining the exact scope of a particular service's SCA protections. Specifically, courts should analyze a service's business model as one step in a three-step analytical framework to help shape their analysis and solidify their reasoning.

Part II of this Note describes and provides background on the SCA. With that understanding in place, Part III discusses *Facebook v. Superior Court*, with a

6. *See id.* at 402 (majority opinion).

7. *See id.*

8. *See id.* at 402–03.

9. *See id.*

10. *See id.* ("We conclude that the trial court below abused its discretion when ruling on the motion to quash by failing to apply the [applicable] test. . . . We will not assess the underlying merits of the business model thesis.").

11. *See* Alfred Ng, *What Does it Actually Mean When a Company Says, "We Do Not Sell Your Data"?*, THE MARKUP (Sept. 2, 2021), https://themarkup.org/the-breakdown/2021/09/02/what-does-it-actually-mean-when-a-company-says-we-do-not-sell-your-data [https://perma.cc/LU6J-3K7X]; *infra* note 169 and accompanying text.

special emphasis on Chief Justice Cantil-Sakauye's concurrence, where she quickly explores the business model theory.

Part IV begins with a short analysis and critique of Chief Justice Cantil-Sakauye's concurrence. Then, Part IV argues for a new three-step analytical framework for scrutinizing online and social media services under the SCA. The three steps consist of: (1) classifying the service as a communication or storage service; (2) determining whether the service authorizes access to user data, and if so, for what reasons; and (3) determining whether the business model or the service's interaction with user data limits the service's SCA protections. This analytical framework is meant to serve as a structuring tool, guiding courts away from improvised and ad-hoc analysis to a systematic method of analyzing modern services under the SCA.

After outlining the three-step analytical approach, Part IV then gives examples of putting that framework into practice. Part IV uses the proposed framework on three social media/web services: Gmail, Facebook Messenger, and Facebook Wallposts.

Finally, Part IV extrapolates from these examples and the business models of popular social media services to reach a final conclusion: the business model theory will be largely unsuccessful in attacking SCA disclosure protections provided to social media services. As the SCA stringently protects communication services, and as social media services are inherently communicative and are increasingly neither sharing nor analyzing their users' communications, the business model theory will not have a noticeable effect on SCA protections for social media sites.

II. A BACKGROUND OF THE SCA

In order to understand the impact of the business model theory on social media companies' disclosure protections under the SCA, it is important to first understand what the SCA is, why it was passed, and how it presents difficulties to courts in the modern world. By limiting specified entities from divulging their users' online and electronic communications, the SCA extends Fourth Amendment-like privacy protections in the online and digital realms. Although protecting online privacy is an ever-growing concern, because the SCA was enacted prior to the advent of the modern internet, courts sometimes find it difficult to apply the SCA to modern technologies and services.

A. THE STORED COMMUNICATIONS ACT

The SCA is a federal statute that governs how and when a service provider can disclose, either voluntarily or compulsorily, stored wired and electronic

communications.[12] Enacted in 1986 as part of the broader Electronic Communications Privacy Act (ECPA), the SCA serves to "represent[] a fair balance between the privacy expectations of citizens and the legitimate needs of law enforcement."[13] As the main statutory and legal framework protecting communications and subscriber information stored by online service providers (OSPs), the SCA plays a vital role in safeguarding Americans' online privacy interests. However, as the SCA was "written prior to the advent of the [modern] Internet," and has not been meaningfully updated, "the . . . statutory framework is ill-suited to address modern forms of communication."[14] Therefore, courts are often frustrated as they "struggle[] to analyze problems involving modern technology within the confines" of the statute.[15]

1. *The Purpose of the Stored Communications Act*

Congress enacted the SCA to extend privacy protections to electronic and internet communications that were largely left unprotected by the Fourth Amendment and the Wiretap Act of 1968.[16] The Fourth Amendment protects people from the government's arbitrary intrusions into their "houses, papers, and effects";[17] a person using the Internet, however, "does not have a physical 'home,' nor really any private space at all."[18] The Wiretap Act, on the other hand, addressed the unauthorized aural interception of "conversations using 'hard' telephone lines, but did not apply to interception of computer and other digital and electronic communications."[19]

The Fourth Amendment's third-party doctrine raised further doubts as to whether online communications would be protected. When accessing internet services, a user divulges their communications to multiple third parties. First, a user must connect to an internet service provider (ISP), such as Comcast or Verizon. The ISP then connects the user to the requested online service provider and acts as an intermediary for all communications to and from that service. Further, when a person uses an OSP, such as an email or online messaging service, the OSP will often keep a record of the user's communications for system integrity and user convenience. As such, online

12. *See* 18 U.S.C. § 2702.
13. H.R. REP. NO. 99-647, at 19 (1986).
14. Konop v. Hawaiian Airlines, Inc., 302 F.3d 868, 874 (9th Cir. 2002).
15. *Id.*
16. *See* S. REP. NO. 99-541, at 2–3 (1986).
17. U.S. CONST. amend. IV.
18. Orin S. Kerr, *A User's Guide to the Stored Communications Act, and a Legislator's Guide to Amending It*, 72 GEO. WASH. L. REV. 1208, 1209 (2004).
19. *Electronic Communications Privacy Act of 1986 (ECPA), 18 U.S.C. §§ 2510–23*, BUREAU OF JUST. ASSISTANCE, https://it.ojp.gov/PrivacyLiberty/authorities/statutes/1285 [https://perma.cc/G3CV-QYQM].

communications are often disclosed to (at least) two third parties, the ISP and the OSP, "that hold and process a user's information on the user's behalf."[20] As of 1986, however, the Supreme Court had regularly upheld the third-party doctrine, which states that there is no reasonable expectation of privacy for information divulged to third parties; consequently, no Fourth Amendment protection applies to that information.[21] In 1986, there was uncertainty over what extent, or if at all, the Fourth Amendment would protect computer and online communications.[22]

Congress emphasized these statutory and constitutional gaps when describing the purpose and necessity of the SCA and ECPA:[23] "This gap results in legal uncertainty. . . . Congress must act to protect the privacy of our citizens. If we do not, we will promote the gradual erosion of this precious right."[24]

The actual text of the SCA reveals another important goal of the statute: to provide privacy protections for electronic information *proportionate* to the privacy interest the information actually implicates. The SCA is filled with dichotomies and distinctions that form the basis of legal categories that are treated differently under the statute.[25] The statute draws distinctions between content and non-content information, voluntary and compelled disclosure, public and nonpublic service providers, and communication and computing

---

20. Kerr, *supra* note 18, at 1210.

21. *See, e.g.*, Smith v. Maryland, 442 U.S. 735, 743–44 (1979) (holding that there is no expectation of privacy for dialed phone numbers conveyed to telephone companies); United States v. Miller, 425 U.S. 435, 443 (1976) (holding that there is no expectation of privacy for deposit slips and checks given to a bank). The Court in *Miller* emphasized:

> [T]he Fourth Amendment does not prohibit the obtaining of information revealed to a third party and conveyed by [them] to government authorities, even if the information is revealed on the assumption that it will be used only for a limited purpose and the confidence placed in the third party will not be betrayed.

Miller, 425 U.S. at 443. However, this doctrine is evolving in the modern age. *See, e.g.*, Carpenter v. United States, 138 S. Ct. 2206, 2223 (2018) (refusing to extend the third-party doctrine to cell-site location information that is automatically sent to wireless carriers); United States v. Warshak, 631 F.3d 266, 288 (6th Cir. 2010) (holding that there is a reasonable expectation of privacy in the *contents* of an email held on third-party servers).

22. *See* S. REP. NO. 99-541, at 3 (1986). More recent jurisprudence has drawn a distinction between non-content online information, which is not protected by the Fourth Amendment, and content information, which may be. *See* Kerr, *supra* note 18, at 1210–11 nn.12–14.

23. *See* S. REP. NO. 99-541, at 3, 5; H.R. REP. NO. 99-647, at 19 (1986).

24. *See* S. REP. NO. 99-541, at 5.

25. *See* Kerr, *supra* note 18, at 1223–34.

services.[26] Based on these distinctions, the statute provides differing levels of privacy protection. For example, greater privacy protection is afforded to content information, providers of services to the public, and communication services than to their respective counterparts.[27] The tiered protection mechanism of the SCA reflects the drafters' judgments about what kinds of information concern greater privacy interests and thus deserve greater protection.

Combining the legislative history of the SCA and its structured protection mechanism reveals a concrete aim. The purpose of the SCA is to cover the gaps in Fourth Amendment privacy protections for online communications, but only to provide just enough protection as is necessary, depending on the seriousness of the privacy interests involved.

2. *The Structure of the Stored Communications Act*

At its core, the SCA serves its purpose by limiting certain types of service providers from either voluntarily disclosing their subscribers' communications or being compelled to do so, while also balancing the interests of law enforcement.[28] To illustrate, service providers included under the SCA can knowingly divulge their users' communications *only* to the addressee or intended recipient of such communication, with the consent of the originator or the addressee of the communication, to a governmental entity "if the provider . . . believes that an emergency involving danger of death or serious physical injury . . . requires disclosure,"[29] or under a few other circumstances listed in 18 U.S.C. § 2702(b). Similarly, covered service providers cannot be forced by the government to divulge their users' communications except via a warrant, subpoena, or court order (depending on the type of service provided).[30] As such, communications held by covered service providers are afforded significant privacy protections.

The SCA, however, does not apply to all online and electronic services; it only provides privacy protections for communications held by two types of service providers: providers of electronic communication services (ECS) and providers of remote computing services (RCS).[31] This is because in 1986 these were essentially the only two types of online services offered: communication services, for sending and receiving messages such as email, and remote

26. *See* 18 U.S.C. §§ 2702–03.

27. *See id.*; *see also* Kerr, *supra* note 18, at 1222–33 (discussing in much greater detail the SCA's structured protection mechanism and dichotomies).

28. *See* 18 U.S.C. §§ 2702–03.

29. 18 U.S.C. § 2702(b)(8).

30. 18 U.S.C. § 2703.

31. *See* 18 U.S.C. §§ 2702–03.

computing services, for storage and outsourcing computing tasks.[32] It is clear that communication services implicate privacy concerns. Users have an interest in keeping their private messages private, even though such services typically make copies and store those messages while in flight. Oftentimes, communication services store private messages in temporary electronic storage for several months for user benefit and system integrity.[33] Likewise, remote computing services implicate similar privacy concerns.[34] Users send their private information to storage or outsourcing services, which then keep copies of that information for extended periods of time.[35]

By only regulating ECS and RCS providers, the SCA thus "freez[es] into the law the understandings of computer network use as of 1986."[36] Because the SCA provides different levels of regulations for ECS and RCS providers,[37] it is implicit that a service cannot be, at one time, both an ECS and an RCS. Therefore, current SCA jurisprudence requires delineating services as either ECS, RCS, or neither. Courts, however, often find it difficult to adjudicate modern cases under the statute, as services do not neatly fall into one category or the other.[38] Nonetheless, figuring out the exact classification of a service is particularly important, as that determines not only if, but also under what standards, the provider is regulated and its communications protected by the SCA.

a) Electronic Communications Services

An electronic communications service is "any service which provides . . . the ability to send or receive wire or electronic communications."[39] Under the SCA, a public ECS provider is prohibited from voluntarily sharing the contents of its stored user communications to any third parties, save for specified exceptions listed in § 2702(b).[40] Importantly, "contents" has an expansive meaning within the SCA. It encompasses not just

32. *See* S. REP. NO. 99-541, at 3 (1986).
33. *See id.*
34. *See id.*
35. *See id.*
36. Kerr, *supra* note 18, at 1214.
37. *See* 18 U.S.C. §§ 2702–03.
38. *See, e.g.*, Flagg v. City of Detroit, 252 F.R.D. 346, 360–63 (E.D. Mich. 2008) (discussing whether SkyTel provided ECS or RCS in relation to stored text messages and admitting that its conclusion might be "mistaken").
39. 18 U.S.C. § 2510(15).
40. 18 U.S.C § 2702(a)(1).

the actual substance of a communication, but also "any information concerning the . . . purport, or meaning of that communication."[41]

The SCA further regulates how a governmental entity may *require* the disclosure of communications or records stored by ECS providers. The statute requires a warrant to compel disclosure of communication contents stored for 180 days or less.[42] There are less stringent requirements, such as requirements for only a subpoena or a court order, for communications held longer than 180 days[43] and for non-content information, such as user records.[44]

b) Remote Computing Services

On the other hand, a remote computing service is one that offers computer storage or processing services.[45] Unlike the near-absolute prohibition on voluntary disclosure of ECS-held communications, the analogous prohibition on RCS-held communications is much weaker. The SCA prohibits RCS providers from voluntarily sharing their stored user content only if three qualifying conditions are all met: (1) the content is carried strictly on behalf of a customer; (2) the content is carried solely for providing storage or processing services; and (3) the provider is not authorized to access the content for any reason other than providing storage or processing services.[46]

The SCA also regulates compelled disclosures of information held by RCS providers. The disclosure protections here are also much weaker than those available to ECS providers. To even be eligible for compelled disclosure protections, an RCS provider has to meet the same three qualifying conditions outlined in the preceding paragraph.[47] Even if the conditions are met, a governmental entity can require disclosure from an RCS provider much more easily than from an ECS provider. To compel disclosure of user content, the government can either (1) obtain a warrant or (2) provide notice to the subscriber and obtain a subpoena or court order.[48] For non-content information, the government needs only a subpoena or a court order; it does not need to notify the subscriber.[49]

41. 18 U.S.C § 2510(8).
42. 18 U.S.C § 2703(a).
43. 18 U.S.C § 2703(b).
44. 18 U.S.C § 2703(c).
45. 18 U.S.C § 2711(2).
46. 18 U.S.C § 2702(a)(2).
47. 18 U.S.C § 2703(b)(2).
48. 18 U.S.C § 2703(b).
49. 18 U.S.C § 2703(c).

c) Significance of the Classification; Difficulty in Classifying

As the discussion, *supra*, illustrates, a service's classification as an ECS or an RCS is significant because it has a tremendous impact on the scope of privacy protections afforded to user information. Not only does the SCA provide generally weaker protections to information held by RCS providers, but it also conditions those protections on the RCS providers meeting the three qualifying conditions. One of the practical effects of these conditions, and one that is central to this paper, is that if an RCS provider is permitted to access its user communications for any reason unrelated to storage or processing, its stored information receives no disclosure protections. On the other hand, there is no such restriction for ECS providers.

While it is extremely important to delineate services as an ECS, RCS, or neither, it is often difficult and tricky to do so, as many modern service providers are multifunctional and provide multiple different services.[50] A specific entity, providing just a single service, "can act as [a] provider[] of ECS in some contexts, [a] provider[] of RCS in other contexts, and as neither in some contexts as well."[51] Further, many modern entities provide a plethora of different services (e.g., Google provides email, chat, maps, search engine, storage, document editing, and many other services). As such, courts have observed that defining whether an entity provides ECS or RCS is a context-specific inquiry. The distinction between the two "serves to define the service that is being provided at a particular time (or as to a particular piece of

50. Courts also find the classification difficult due to an "inherent structural flaw[] in the statute." *See* Eric R. Hinz, Note, *A Distinctionless Distinction: Why the RCS/ECS Distinction in the Stored Communications Act Does Not Work*, 88 NOTRE DAME L. REV. 489, 501 (2012). There is uncertainty over whether storage "for purposes of backup protection" should be read expansively or narrowly. *Id.* If a communication is stored for backup purposes, then the service is an ECS; otherwise, the service is an RCS (or neither).

51. Kerr, *supra* note 18, at 1215. Professor Kerr further explains that an entity can act as an ECS provider, RCS provider, or neither with respect to not only different communications, but also with respect to different copies of the same communication.

> In the case of a public provider, for example, files held in intermediate "electronic storage" are protected under the ECS rules; meanwhile, files held for long-term storage by that same provider are protected by the RCS rules. The same treatment exists for different copies of the same communication: a provider can act as an ECS with respect to one copy of a communication, as an RCS with respect to another copy, and as neither an ECS nor an RCS with respect to a third copy.

Id. at 1216 (footnotes omitted).

electronic communication at a particular time), rather than to define the service provider itself."[52]

d) Calls to Modernize the SCA

As the SCA was adopted in 1986 and was "written prior to the advent of the Internet and the World Wide Web,"[53] courts have often called upon Congress to modernize the SCA. In *Konop v. Hawaiian Airlines*, the Ninth Circuit stated that "the [SCA] is ill-suited to address modern forms of communication," and that "[c]ourts have struggled to analyze problems involving modern technology within the confines of this statutory framework."[54] The *Konop* court noted that "until Congress brings the laws in line with modern technology, protection of the Internet and websites . . . will remain a confusing and uncertain area of the law."[55] Similarly, in 2013, a frustrated federal district court wrote that "[m]ost courts, including this one, would prefer that Congress update the [SCA] to take into account the invention of the Internet."[56] While discussing the shortcomings of the SCA, courts have also specifically pointed out the difficulties in determining the scope of the statute's application to social media platforms and communications.[57]

III. *FACEBOOK V. SUPERIOR COURT OF SAN DIEGO COUNTY*

With the background of the SCA in mind, Part III of this Note turns to *Facebook v. Superior Court*, where the business model theory most recently materialized. Section III.A summarizes the case and the majority opinion as background, since the case was decided on separate grounds not involving the applicability of the SCA to social media companies. Section III.B of this Part then analyzes Chief Justice Cantil-Sakauye's concurrence, in which she explored the business model theory in greater depth.

52. *In re* United States, 665 F. Supp. 2d 1210, 1214 (D. Or. 2009).

53. *See* Konop v. Hawaiian Airlines, Inc., 302 F.3d 868, 874 (9th Cir. 2002).

54. *Id.*

55. *Id.*

56. Ehling v. Monmouth-Ocean Hosp. Serv. Corp., 961 F. Supp. 2d 659, 666 n.2 (D.N.J. 2013).

57. State v. Johnson, 538 S.W.3d 32, 68 (Tenn. Crim. App. 2017); *see also In re* Application of State for Commc'ns Data Warrants to Obtain the Contents of Stored Commc'ns from Twitter, Inc., 154 A.3d 169, 177 (N.J. Super. Ct. App. Div. 2017) ("Courts have expressed frustration with the failure to update the federal statute to keep pace with the advent of the Internet and social media platforms . . . .").

A. SUMMARY OF THE CASE AND MAJORITY OPINION

In *Facebook*, a defendant on trial for attempted murder by shooting sought to subpoena his victim's Facebook communications.[58] The defendant, Lance Touchstone, argued that he needed his victim's Facebook communications, both before and after the shooting, to investigate an affirmative self-defense claim and to impeach the victim's character.[59]

Facebook moved to quash the subpoena, but the superior court denied the motion, finding that there was good cause.[60] After a timely appeal by Facebook, however, the Supreme Court of California remanded the case to the superior court with directions to vacate its order. The California Supreme Court held that the trial court "erred by conducting an incomplete assessment of the relevant factors and interests."[61] The court used this case to point out the relevant seven factors ("Alhambra factors") that a court "should explicitly consider and balance in ruling on a motion to quash a subpoena . . . directed to a third party."[62]

By deciding and remanding the case based on the foundational issue of subpoena validity, the court did not have a chance to address the defendant's business model theory. The defendant argued that due to Facebook's business model, Facebook could not use the SCA as a shield to protect itself from non-compliance with the defendant's subpoena.[63] Since Facebook mines, analyzes, and shares information about its users' communications to facilitate targeted advertisements, the defendant asserted that the SCA's disclosure protections, codified in 18 U.S.C. §§ 2702(a) and 2703(b), did not apply and that Facebook must therefore comply with the subpoena.[64] And while Facebook urged the court to address this substantive issue, the court, noting the issue's importance, declined to do so, as it had disposed of the case on other grounds.[65]

---

58. Facebook, Inc. v. Superior Ct. of San Diego Cnty., 471 P.3d 383, 387 (Cal. 2020).

59. *Id.*

60. *Id.* at 399.

61. *Id.* at 387.

62. *Id.* The seven factors were originally articulated in *City of Alhambra v. Superior Court*, 252 Cal. Rptr. 789 (Ct. App. 1988), *superseded in part by constitutional amendment*, CAL. CONST. art. I, § 30(c), *as recognized in Garcia v. Superior Court*, 163 P.3d 939, 946–48 (Cal. 2007), and consist of: (1) a "plausible justification," (2) the adequacy of description of the materials sought, (3) the availability of the sought materials from other sources, (4) whether any rights of the third party or "any protected governmental interest" would be violated, (5) the timeliness of the request, (6) the potential of delay of trial, and (7) whether there would be an undue burden on a document-producing third party. *Id.* at 799–800.

63. *Facebook*, 471 P.3d at 402.

64. *Id.* at 402–03.

65. *Id.* at 402.

B. CHIEF JUSTICE CANTIL-SAKAUYE'S CONCURRENCE

Realizing the importance and potential implications of the defendant's business model theory, Chief Justice Cantil-Sakauye wrote a separate concurrence to explore the theory in more detail.[66] The Chief Justice first gave an overview of the business model argument, then provided background by reviewing the SCA's definition of ECS and RCS, and then lastly summarized the parties' contentions regarding Facebook's ECS/RCS status.

At a high level, the Chief Justice characterized the business model theory as an attack that sought to place Facebook and other similarly situated social media services "outside the ambit of . . . the Stored Communications Act."[67] The theory asserts that a service's "authorization to undertake, and its practice of" accessing its users' communications outside of certain limited circumstances "renders [the service] subject to a viable state subpoena."[68] For example, a service that mines, analyzes, and shares its subscribers' communications, or is authorized to do so, is precluded "from qualifying under the SCA *as a provider that is prohibited by the Act from disclosing user content*."[69] Therefore, such a service "cannot hold up the Act as a shield that protects it from complying with a viable state subpoena."[70]

In particular, the Chief Justice interpreted the business model theory as arguing that a service that accesses its users' communications for reasons outside the listed circumstances in the SCA "does not qualify as an entity that provides either ECS or RCS with respect to the sought communications."[71] The SCA only "covers, and prohibits disclosure of" communications held by ECS and RCS providers.[72] If an entity does not provide either in regard to a specific communication, "the entity cannot rely upon the SCA" to protect itself against a valid subpoena seeking that particular communication.[73]

When defining ECS and RCS, the Chief Justice emphasized the limitations put upon qualifying service providers in the disclosure provisions of §§ 2702(a) and 2703(b). As described above, an ECS provider is barred from knowingly divulging the contents of any stored user communications.[74] For RCS, the Chief Justice noted that several legal commentators, relying on the statutory

66. *Id.* at 403–04 (Cantil-Sakauye, C.J., concurring).
67. *Id.* at 403.
68. *Id.* at 404.
69. *Id.* at 404–05.
70. *Id.* at 405.
71. *Id.*
72. *Id.*
73. *Id.*
74. *See* 18 U.S.C. § 2702(a)(1).

language of § 2702, have reasoned that if an entity is "'authorized to access the contents of any [RCS communications] for purposes . . . other than storage or computer processing' . . . the Act's bar on disclosure is inapplicable. In other words, . . . such an entity would not be acting as an RCS."[75]

With the business model theory described and the ECS/RCS definitions reviewed, the Chief Justice considered the parties' arguments as to whether Facebook provided ECS, RCS, or neither. The criminal defendant argued that Facebook does not act as an ECS provider, as "(1) Facebook is authorized to [and does] mine, analyze, and share with third party advertisers licensed information about its users' content . . . , and (2) Facebook stores users' communications indefinitely, lets users share the stored data with others, and facilitates manipulation of the data by the user."[76] Facebook, on the other hand, asserted that it qualified as an ECS provider, as it either kept its users' communications in "temporary or intermediate storage" or stored them "for purposes of backup protection."[77] It further insisted that having access to its users' content was irrelevant in determining whether it was an ECS provider, and stated that a number of courts had already found it to be an ECS provider.[78]

In analyzing these arguments, the Chief Justice first noted that determining the ECS/RCS status of a service is a "context-dependent inquiry" that defines the service in relation to a particular communication at a particular time.[79] The Chief Justice pointed to federal decisions that held that when services such as email hold a message after it has been opened or accessed, the service transforms from an ECS to an RCS.[80] The Chief Justice analogized those cases to the current one. She thus stated,

> [W]hether Facebook should be found to qualify as a provider of ECS . . . appears open to question. Moreover, assuming that Facebook might qualify initially . . . as an entity that provides ECS, it . . . may also be obligated to establish its qualification as an entity that provides RCS with respect to stored communications sought in a viable state subpoena.[81]

75. *Facebook*, 471 P.3d at 406 (Cantil-Sakauye, C.J., concurring) (quoting 18 U.S.C. § 2702(a)(2)(B)) (footnote and emphasis omitted).
76. *Id.* at 408.
77. *Id.* (quoting 18 U.S.C § 2510(17)).
78. *Id.*
79. *Id.*
80. *Id.* at 409.
81. *Id.*

The Chief Justice then analyzed the parties' arguments about whether Facebook qualified as an RCS provider. The Chief Justice first reiterated that due to the statutory language of § 2702, several legal commentators and at least one court had found that if an entity is authorized to access its users' communications for purposes other than providing storage or computer processing, "the entity may not, or does not, qualify under the SCA as one that provides RCS."[82] Using this argument, the criminal defendant asserted that since Facebook, as necessitated by its business model, does mine, analyze, and share information about its users' content with their authorization, it cannot qualify as an RCS provider.[83] Following this argument, those activities went beyond merely providing storage or processing services and "demonstrate[] that Facebook is authorized to act *in precisely the manner the statute says it must not* if it wishes to qualify as a provider of RCS."[84] Facebook, on the other hand, contended that everything it was authorized to do fell under the umbrella of computer processing services.[85] The Chief Justice noted that Facebook cited a federal decision, legislative history, and an academic article by Orin Kerr to support its claim, all allegedly showing that computer processing services should be broadly construed.[86] Unconvinced, the Chief Justice questioned whether those sources really supported Facebook's assertion; in fact, she stated that the sources "may suggest the opposite—that ['computer processing services'] was intended to have a narrow, rather than broad, interpretation."[87]

Facebook also argued that every court that had previously considered whether Facebook was qualified as an ECS or RCS provider found that it met at least one of the tests.[88] The Chief Justice, however, remained unconvinced and simply stated that no court had ever considered the criminal defendant's specific claim that due to Facebook's business model, it did not qualify as an ECS or RCS provider, thus falling outside the purview of the SCA.[89] That issue remains unresolved.[90]

Going beyond the statutory language and precedent, Facebook also asserted that its policy considerations necessitated that it qualified under the SCA. "[C]oncluding otherwise would (1) unduly disrupt and impair technological innovation, (2) disappoint users' settled privacy expectations,

82. *Id.*
83. *Id.* at 409–10.
84. *Id.* at 410.
85. *Id.*
86. *Id.*
87. *Id.*
88. *Id.*
89. *Id.*
90. *Id.*

and (3) frustrate its ability to protect against malware."[91] Citing "practical marketplace reasons," the Chief Justice questioned whether disqualifying Facebook as a protected entity under the SCA would actually lead to "disruptions or voluntary disclosures . . . absent legal compulsion."[92] Further, the Chief Justice stated that a narrower construction of "computer processing" could still include measures to counteract malware, while excluding mining and analyzing user data for targeted advertisements.[93] Finally, the Chief Justice stated that "as a matter of policy, . . . finding Facebook to lie outside the SCA might have the beneficial effect of spurring long-needed congressional adjustment of the outdated [Stored Communications] Act."[94]

IV. PROPOSED FRAMEWORK AND APPLICATIONS

With an understanding of the SCA in place and with the basic structure and mechanisms of the business model theory explained by Chief Justice Cantil-Sakauye, this Note can now turn to its analysis. In this Part, this Note first analyzes how the business model theory affects modern social media services and their disclosure protections under the SCA. Section IV.A gives a brief critique of the Chief Justice's concurrence in *Facebook* to highlight key interpretative differences in the business model theory. This Note then discusses its three-step analytical framework in Section IV.B, with the hopes of giving courts a systematic way of analyzing modern services under the SCA. Following the introduction of this framework, Section IV.C demonstrates its application to three web services. Finally, Section IV.C uses the worked-out examples to draw out general lessons and extrapolations to reach an overall conclusion: the business model theory will not have a noticeable effect on social media services and their disclosure protections.

A. CRITIQUE OF THE CHIEF JUSTICE'S CONCURRENCE

The Chief Justice's brief analysis of the business model theory, though helpful in outlining the theory's main ideas and arguments, had a few notable shortcomings. First, and most importantly, the Chief Justice overgeneralized the business model theory. She interpreted it to mean that a successful business model attack rendered a service completely outside the purview of the SCA, instead of just the SCA's disclosure provisions. Second, the Chief Justice failed to acknowledge the underlying strangeness of the business model theory:

91. *Id.*
92. *Id.* at 410–11.
93. *Id.* at 411 n.14.
94. *Id.*

mainly that the theory further punishes consumers for a service's insecure or privacy-harming business model (e.g., a service that shares user data with third parties for ad targeting). Finally, while the Chief Justice made a valid point that marketplace concerns will prevent services from carelessly sharing user data even without the SCA protections in place, such important privacy rights deserve to be statutorily protected anyway.

Throughout the concurrence, the Chief Justice overgeneralized the business model theory. Relying on the criminal defendant's characterization and a number of legal scholars, the Chief Justice interpreted the theory as stating that an electronic communication or computing service, by accessing or having the authority to access user communications for reasons other than providing communication, storage, or processing functions, could not be classified as an ECS or RCS, thus falling completely outside of the SCA's scope.[95] The business model theory, when more properly stated, is more focused; it is better interpreted as asserting two specific claims. First, if an ECS's business model requires it to share user communications to third parties, it violates § 2702(a)(1), which bars an ECS provider from "knowingly divulg[ing]"[96] user communications. Second, and much more significantly, the business model theory states that if an RCS has the authorization to access user communications outside of providing storage or processing services, that RCS falls outside of the disclosure provisions set forth in §§ 2702(a)(2) and 2703(b), making its communications unprotected from voluntary and compelled disclosures. Importantly, the business model theory does not remove the service from being an ECS or RCS; the service still retains its classifications and is still bound by the rest of the SCA.[97] It is simply the disclosure provisions that are violated or no longer apply.

The Chief Justice's overgeneralization of the business model theory was largely dependent on conflating the conditions of the disclosure provisions with the actual definitional requirements of ECS and RCS. The definition of ECS, set out in § 2510(15), states that an ECS is "any service which provides . . . the ability to send or receive wire or electronic communications."[98] If an ECS shares its stored communications to third parties, as it is prohibited to do under disclosure provision § 2702(a)(1), it

95. *See id.* at 405–06, 408–10.

96. 18 U.S.C. § 2702(a)(1).

97. For example, an RCS that accesses its user communications for reasons unnecessary to provide storage or processing services loses SCA §§ 2702(a) and 2703(b) disclosure protections for its user communications. However, it is still bound by § 2703(c), which governs compelled disclosures of subscriber records and information.

98. 18 U.S.C. § 2510(15).

merely violates that subsection;[99] it still remains an ECS. Similarly, the definition of RCS is set out in § 2711(2) and means any service that provides the public "computer storage or processing services."[100] If an RCS is authorized to access its users' communications for reasons unrelated to computer storage or computer processing, it simply means that the RCS fails the condition to receive disclosure protections under §§ 2702(a)(2) and 2703(b). The service still remains an RCS, but its stored communications are no longer protected.

Further, the Chief Justice failed to address the inherent oddity in the business model attack, which further punishes consumers for a service's already problematic business model. Boiled down, the RCS prong of the SCA ends up stripping people of their online privacy because an RCS provider is using a business model that itself decreases user privacy (i.e., requiring authorization to access user communications for non-storage or non-processing reasons). In effect, a successful business model attack harms user privacy because user privacy is already being harmed.

Proponents of the business model theory may argue that when users authorize a service provider to access their communications for reasons unrelated to the central (i.e., storage or processing) service, the users no longer have an expectation of privacy. And while this argument may have made sense in the 1980s, where online services only had one function and were not so intrinsically linked with our day-to-day lives, it no longer holds weight. Most modern entities are multifaceted and provide several different services, both ECS and RCS. When a user authorizes an entity to access their information, they do so by agreeing to a privacy or a data policy that concerns the entity as a whole. Most policies do not differentiate between services, and as such do not differentiate the authorizations given to ECS components versus RCS components. Therefore, courts should not be looking at authorization in the abstract, as that deals with the entire entity. Instead, courts should be looking at if and how the RCS components actually access user information.

From a more policy-oriented perspective, online services are now so linked with our daily lives that it is wrong to say that because a user authorizes a company to access their data to accomplish non-storage/non-processing

99. 18 U.S.C § 2707(a) provides that "[any] subscriber, or other person aggrieved by any violation of this chapter . . . may, in a civil action, recover from the person or entity . . . which engaged in that violation such relief as may be appropriate." 18 U.S.C. § 2707(a). Section 2707(b) lists three appropriate remedies available to plaintiffs who bring a civil action under the SCA: "preliminary and other equitable or declaratory relief," damages, and reasonable attorney's fees and costs. 18 U.S.C. §§ 2707(b)(1)–(3).

100. 18 U.S.C. § 2711(2).

tasks, they no longer deserve any privacy protection. And while marketplace concerns[101] make it unlikely that online entities would voluntarily disclose user communications, such privacy rights are so important that they deserve to be statutorily protected regardless. Relying on an entity's goodwill and reputational concerns is not good enough.

B. PROPOSED ANALYTICAL FRAMEWORK

While Chief Justice Cantil-Sakauye's concurrence in *Facebook* provides insight into the business model theory, how it might apply to Facebook, and the potential policy considerations it implicates, the theory still "deserves additional and focused attention."[102] Given social media's pervasive nature and ever-growing influence in our society, it is vital to understand how the business model theory affects the SCA disclosure protections provided to social media communications.[103]

A closer look at the business model theory reveals that it has a minimal impact on disclosure protections provided to most social media services. Due to the inherent communicative purpose of social media and the nature of modern ad-targeting business models, where user data is neither shared nor sold, the business model attack largely falls flat. Nevertheless, due to the numerous services and business models available, business model scrutiny, which is too-often ignored and disregarded by current courts, is an important step in SCA analysis. Courts should make a conscious effort to add this step into their analysis to better and more fully determine the scope of a service's disclosure protections. In particular, courts should implement business model scrutiny as one step within a three-step analytical framework when analyzing modern services under the SCA.

The three-step framework is meant to serve as a structuring tool, leading courts to a more concrete and systematic analysis of services under the SCA. In the first step, courts should classify the service in question as an ECS, RCS, or neither. This is likely the most arduous step in the framework. In the second step, assuming that the service fits as an ECS or RCS, courts should identify the service's business model and determine exactly what authorizations the service has to access consumer data. In the third and final step, courts should determine how the service's business model and authorizations to access and use user data impact the service's compliance with the SCA's disclosure provisions. Specifically, if the service is an ECS and divulges user

101. Facebook, Inc. v. Superior Ct. of San Diego Cnty., 471 P.3d 383, 410–11 (Cal. 2020) (Cantil-Sakauye, C.J., concurring).

102. *Id.* at 411.

103. *Id.* at 412 (Cuéllar, J., concurring).

communications to third parties, it violates the SCA and can be liable for civil damages.[104] If the service is an RCS and has authorization to or does access consumer data for reasons other than providing storage or processing services, its user communications are not protected from disclosures under the SCA.

This analytical framework is necessary given that there is little statutory guidance in applying the SCA to modern technologies and services, leaving courts confused and frustrated.[105] Left to their own devices, courts currently engage in a case-by-case, ad-hoc analysis, which often leaves important steps ignored, such as business model scrutiny. This also results in hard-to-understand opinions and weak analysis, which further muddies the water.

1. Step One: Classification of Service as an ECS, RCS, or Neither

Here, courts should use traditional analytical methods, shaped by statutory interpretation, precedent, peer courts, and academic and legal research, to determine the classification of a service as an ECS, RCS, or neither. Courts should keep in mind that this is a context and fact-specific inquiry that defines not the entire entity or corporation, but rather the specific service as it relates to a particular piece of information at a particular time.[106]

a) Determining Whether a Service is an ECS

As defined *supra*, an ECS is "any service which provides to users thereof the ability to send or receive wire or electronic communications."[107]

In order to add certainty and avoid contravention of the statute's intent, courts should classify a service as an ECS if the relevant piece of communication can be viewed and edited or responded to by at least one non-sending, non-service party. This conception would modernize the SCA by adding certainty as to whether certain modern services can be classified as an ECS or not. For example, it would definitively allow providers of emails, chats, wall and forum posts, tweets, shared pictures, and editable documents to be considered ECS providers. Recipients of these communications can message back, leave a comment, "like" the photo, or edit the shared documents.

104. *See supra* note 99.

105. *See, e.g.*, Konop v. Hawaiian Airlines, Inc., 302 F.3d 868, 874 (9th Cir. 2002); Ehling v. Monmouth-Ocean Hosp. Serv. Corp., 961 F. Supp. 2d 659, 666 n.2 (D.N.J. 2013); State v. Johnson, 538 S.W.3d 32, 68 (Tenn. Crim. App. 2017); Anzaldua v. Ne. Ambulance & Fire Prot. Dist., 793 F.3d 822, 839 n.5 (8th Cir. 2015).

106. *See In re* United States, 665 F. Supp. 2d 1210, 1214 (D. Or. 2009); Flagg v. City of Detroit, 252 F.R.D. 346, 362 (E.D. Mich. 2008); *Facebook*, 471 P.3d at 408 (Cantil-Sakauye, C.J., concurring); Kerr, *supra* note 18, at 1215 & n.48.

107. 18 U.S.C. § 2510(15).

Therefore, a service in relation to such a communication can be quickly and easily classified as an ECS, at least for certain periods of time.

The requirement that at least one other party can view or respond to the communication also ensures that the service is actually being used to communicate and share information, rather than being used for rudimentary storage or backup purposes. Thus, if a person sends a message or email to only themself or changes their privacy settings so that their tweets or wallposts are only visible to themself, the service, in that particular situation, is not acting as an ECS. This would prevent anyone from "gaming" the SCA by making it seem like they are sending communications, when in reality they are just storing information (as ECS communications are provided higher privacy protections than RCS communications). A change in privacy settings of a particular communication, therefore, may lead to a change in classification of that service. For example, if a user posts a picture on Facebook for their friends, but then changes the privacy settings so that only they can still see the picture, Facebook would go from an ECS provider to an RCS provider or neither in that context.

As services can change between ECS, RCS, and neither even in relation to a specific piece of communication, timing plays an important role in the service's classification. For instance, multiple courts have held that webmail services are ECS while an email is in transit and sits unopened by the recipient, but transform into RCS once that email has been accessed.[108] As Professor Orin Kerr notes,

> The thinking is that when an e-mail customer leaves a copy of an already accessed e-mail stored on a server, that copy is no longer "incident to transmission" nor a backup copy of a file that is incident to transmission: rather, it is just in remote storage like any other file held by an RCS.[109]

108. *See, e.g.*, United States v. Weaver, 636 F. Supp. 2d 769, 772–73 (C.D. Ill. 2009); Fraser v. Nationwide Mut. Ins. Co., 135 F. Supp. 2d 623, 635–38 (E.D. Pa. 2001), *vacated in part on other grounds*, 352 F.3d 107 (3d Cir. 2003). *But see* Theofel v. Farey-Jones, 359 F.3d 1066, 1076 (9th Cir. 2004) (holding that an email on a server, regardless of whether it has been accessed, is protected under the ECS rules until "the underlying message has expired in the normal course"); Quon v. Arch Wireless Operating Co., 529 F.3d 892, 902–03 (9th Cir. 2008) (re-affirming *Theofel's* holding that an ECS that stores an already-accessed communication remains an ECS indefinitely), *rev'd on other grounds sub nom.* City of Ontario v. Quon, 560 U.S. 746 (2010). The view espoused by *Theofel* has been attacked by courts and legal scholars. *See Weaver*, 636 F. Supp. 2d at 772–73; Kerr, *supra* note 18, at 1217 & n.61.

109. Kerr, *supra* note 18, at 1216. This thinking also comports with legislative intent. *See* H.R. REP. NO. 99-647, at 64–65 (1986) (noting that an accessed email still kept on a provider's server is covered under sections relating to remote computing services).

This reasoning can be easily extended to text and online chat messages; if the message has been accessed by the intended recipient but is still stored by the service, the service is acting as an RCS in regard to the recipient's copy of the communication.[110]

If courts cannot easily tell when the communication has been opened or accessed, they should default to the 180-day lifespan for ECS communications asserted in the SCA; thereafter, they should treat the service as an RCS. For communications such as tweets or wallposts, where there is just one copy of the communication that is viewable by a large group of "friends" or "followers," determining when every "friend" has accessed the communication would be tough, if not impossible. Courts can thus grant those communications perpetual protection under the ECS provisions, but this is likely to be unsatisfactory; not only would this treat tweets and wallposts differently from email and chat, but it would also seem illogical to say that an entity is still providing an electronic communication service in relation to a message posted, say, five years ago. Instead, courts should use the SCA-drawn lifespan for ECS communications, 180 days, as the default upper limit. Section 2703(a) provides strong protection against required disclosures for ECS contents held in electronic storage for 180 days or less.[111] It provides weaker protection for contents held in storage for longer than 180 days; the protection given to contents in long-term storage is the same as that afforded to contents held by RCS providers.[112] Courts should use this distinction and hold that once a provider stores an ECS communication for longer than 180 days, the service transforms to an RCS in relation to that communication.[113]

This Part's conception of ECS keeps pace with modern developments while still pursuing the statutory purposes of the SCA and remaining faithful to legislative intent and court precedent. The 1986 House Report on the SCA classified "electronic bulletin boards," which offer "interested persons [the ability to] communicate openly with the public to exchange . . . information," as a feature, specifically, of some *electronic communication services*.[114] As such, courts have repeatedly held that bulletin board services are protected under

110. *See Flagg*, 252 F.R.D. at 362–63 (holding that Skytel, a text message service provider, was initially an ECS provider, but transformed into an RCS provider after the communication in question was accessed and stored).

111. *See* 18 U.S.C. § 2703(a).

112. *See id.*

113. *See* Kerr, *supra* note 18, at 1217 n.61.

114. H.R. REP. NO. 99-647, at 62 (1986).

the SCA,[115] and at least two courts have explicitly found that such services fall under the umbrella of ECS.[116] The electronic bulletin boards of 1986 can be easily analogized to present-day forums, such as Reddit, and to services that offer wallposts, tweets, and multimedia sharing, such as Facebook, Twitter, and Instagram.[117] Like bulletin boards, these modern day services allow users to share thoughts and information: a user can post a particular topic, which may come in the form of a thread, a wallpost, or even a picture, and other users can respond with their thoughts, opinions, or additional information. Protecting these services under ECS, which is afforded stronger protections than RCS, would further the purpose of the SCA; it would provide strong Fourth Amendment-like protections to online social communications that implicate important privacy interests.

b) Determining Whether a Service is RCS

If a service is not classified as an ECS, courts then have to determine if the service is an RCS. As defined in § 2711(2), a remote computing service is "the provision to the public of computer storage or processing services."[118]

As a threshold matter, courts should only question whether a service is an RCS if they have already concluded that it is not an ECS. This is because communications held by ECS providers are afforded greater privacy protections than those held by RCS providers.[119] Thus, if a particular service can be classified as both an ECS and an RCS in regard to a particular communication, it should by default be considered an ECS and the communication afforded the appropriate protections. To do otherwise would be to provide lesser, RCS-style protections to a communication technically held by an ECS; this would go against the terms of the SCA. Relatedly, by only reaching the RCS determination inquiry after holding that the service failed to

115. *See, e.g.*, United States v. Steiger, 318 F.3d 1039, 1049 (11th Cir. 2003); Steve Jackson Games, Inc. v. U.S. Secret Serv., 36 F.3d 457, 462 (5th Cir. 1994); Konop v. Hawaiian Airlines, Inc., 302 F.3d 868, 875 (9th Cir. 2002).

116. *See Konop*, 302 F.3d at 875, 879; Kaufman v. Nest Seekers, LLC, No. 05 CV 6782 (GBD), 2006 WL 2807177, at *5 (S.D.N.Y. Sept. 26, 2006). *But see* Steve Jackson Games, Inc. v. U.S. Secret Serv., 816 F. Supp. 432, 443 (W.D. Tex. 1993) (finding that a bulletin board service was an RCS), *aff'd*, 36 F.3d at 457.

117. *See* Crispin v. Christian Audigier, Inc., 717 F. Supp. 2d 965, 980–81 (C.D. Cal. 2010) (finding Facebook and Myspace wallposts to be analogous to bulletin board services).

118. 18 U.S.C. § 2711(2).

119. *See* 18 U.S.C. § 2702(a) (providing a blanket prohibition against voluntary disclosure of ECS-held communications but requiring multiple conditions for the prohibition to apply to RCS-held communications); 18 U.S.C. § 2703 (generally requiring a warrant to compel disclosure of ECS-held communications, but only a subpoena or court order for an RCS-held communication).

qualify as an ECS, courts can preempt the question asking whether all modern, online ECS providers, because they invariably require storage and processing functionalities, necessarily provide RCS as well. While courts can dispose of this question by (1) noting that the SCA specifically protects ECS communications "in electronic storage"[120] and then differentiating electronic storage from the RCS conception of storage and by (2) highlighting the higher level of protections afforded to ECS communications, this question merely serves to distract and confuse courts. This question is simply unnecessary, as once an ECS determination has been made, the classification inquiry should stop.

In regard to RCS, the concept of computer storage is relatively straightforward. It simply refers to storing information on another entity's servers for either space, backup, or user convenience considerations. There are countless services in the market that specialize in offering such services: some of the more popular ones today are Dropbox, Google Drive, and Microsoft OneDrive. Additionally, computer storage service also incorporates services that were ECS but transformed into RCS. As described above, if a communication service retains the communication after the recipient has already accessed it, or if the service stores the communication for longer than 180 days, the service transforms from an ECS into an RCS.[121] It makes sense to conceptualize those "drop-down" services as providing computer storage.

Defining processing services, however, is more difficult. In an abstract sense, every modern online technology can be described as a processing service. When you interact with an online service, the service processes your inputs and provides you with the appropriate output(s). For example, when you use a search engine, the search engine processes each keystroke, displaying them on your monitor. Once the query is typed up and you hit the "Search" button, the search engine searches its index for the appropriate content, ranks it in terms of relevancy (usually), and then displays the results.[122] Similarly, when you are online shopping and click on a link for "Men's Shoes," the service pulls up its list of men's shoes for sale, ranks them using some algorithm (e.g., popularity over the last quarter), and then displays the results. If you then sort by "Price Low–High," the service changes the order of the results to show you the cheapest shoes first.

---

120. *See* 18 U.S.C. § 2702(a)(1).

121. *See supra* Section IV.B.1.a.

122. Britney Muller, *How Search Engines Work: Crawling, Indexing, and Ranking*, MOZ, https://moz.com/beginners-guide-to-seo/how-search-engines-operate [https://perma.cc/U9KK-3AC3] (last visited Aug. 17, 2022).

Legislative history, however, shows that the drafters of the SCA meant processing services to be more narrowly defined. The Senate Report talks about "businesses of all sizes transmit[ting] their records to remote computers to obtain sophisticated data processing services."[123] This shows that the legislators intended "processing services" to mean those used when outsourcing tasks and functions. "In the era before spreadsheets, a company might send raw data to a remote computing service and ask the service to crunch numbers to calculate its payroll."[124] As such, processing services, in relation to the SCA, should only include services that are used when outsourcing "sophisticated" data processing or other computationally-intensive tasks. Currently, some popular processing services include Amazon AWS, Microsoft Azure, and Google Cloud. These services may be used to run analytics on Big Data, train Artificial Intelligence, provide network security and monitoring, render animations, and much more.

2. *Step Two: Analyzing the Service's Business Model and Interaction with Consumer Data*

If courts have classified a service as an ECS or RCS, as opposed to neither, they should then determine the data authorizations the service has and how and why it interacts with consumer data. Here courts should specifically look for whether the service divulges its users' communications to third parties or if it has authorization to and does access its users' communications for reasons unnecessary to the service it provides. Such disclosures or data accesses would implicate violations of the SCA or disqualify the service from the SCA's disclosure provisions (as examined in Step Three).[125]

Scrutinizing a service's business model may entail examining the company's annual reports, reading the company's data, privacy, and security policies, analyzing the dominant models used by the company's peers, and simply asking or requiring the company to disclose its pertinent model(s). It is important to note, however, that the business model of interest is that of the service, and not of the entire entity. A few high-level examples of business/revenue models include selling consumer data, using targeted advertisements, and selling additional or premium features for a cost.

Courts should then seek to understand if, how, and why the service accesses user data in relation to its business model. This will likely require reading and understanding the service's data and privacy policies. It should be noted, however, that it might be difficult to disentangle how and why different

123. *See* S. REP. NO. 99-541, at 3 (1986).
124. Kerr, *supra* note 18, at 1230.
125. *See* 18 U.S.C. §§ 2702(a), 2703(b).

services of a particular company access consumer data; most policies are written to apply one set of rules to an entity's entire suite of services, rather than different sets to different services. More precisely, courts should look at the service's treatment of the particular content or communication in question, as that will ultimately decide if the service is violating the SCA or is disqualified from the SCA's disclosure provisions in regard to that distinct communication.

If the service is an RCS, courts should ask if users have specifically authorized the service, and not the entire entity, to access their data for any reasons unrelated to providing storage or processing services. This is an extremely difficult question because, as noted above, privacy and data policies are not usually separated service-by-service or by ECS vs. RCS components. A user authorizing an entity to access its data is not the same as the user authorizing the specific RCS component. Therefore, courts should also look to whether the RCS actually accesses user data for non-storage or processing reasons, as that is likely the best proof of authorization. If appropriate, courts can also ask if users can change the service's authorization settings. If so, courts should then look at the actual authorization settings in play, as they relate to particular communications at the time of the inquiry.

If the service is an ECS, courts should ask if the service knowingly shares or sells its users' communications to third parties. These ECS and RCS specific questions are important, as they will guide the courts in Step Three when determining whether the service violates or is disqualified from the disclosure provisions of the SCA.

3. *Step Three: Determining Whether the Service's Business Model Violates the SCA or Disqualifies It from the SCA's Disclosure Provisions*

Once courts understand the service's business model and how and why it interacts with consumer data, they should look at §§ 2702 and 2703 and determine if the service violates or is disqualified from the SCA's disclosure provisions. Section 2702(a)(1) states that an ECS provider cannot "knowingly divulge" or share its users' communications to third parties.[126] On the other hand, RCS-held user communications are only protected from voluntary or compelled disclosure if the provider is not authorized to access its users' communications for any reason other than providing computer storage or processing services.[127]

As such, if the service is classified as an ECS, and the service does knowingly divulge the contents of the user communications in question, then

126. *See* 18 U.S.C. § 2702(a)(1).
127. *See* 18 U.S.C. §§ 2702(a)(2), 2703(b).

the service violates the voluntary disclosure provision set out in § 2702(a)(1) and would be liable for civil penalties.

Similarly, if the service is classified as an RCS, and is authorized to access the communication in question for any non-storage or non-processing reasons, then the service is disqualified from the voluntary and compelled disclosure provisions set forth in §§ 2702(a)(2) and 2703(b). In other words, the RCS provider can share user information with who and whatever it wants, and it must comply with simple government subpoenas. However, as explained in Step Two, it is very difficult to see if a particular RCS service or functionality actually has that authorization. Therefore, courts, in most circumstances, should look to actual access and not just authorization (which tends to be entity-level authorization) when determining whether the RCS provider should be disqualified from the SCA disclosure provisions. That is, courts should only disqualify an RCS provider if it actually accesses user communications for reasons unrelated to computer storage or processing.

Further, courts should interpret the disqualifying condition narrowly, allowing RCS providers to access consumer data not just for the core storage or processing service, but also for all supplementary services necessary to maintain system integrity, preserve user security, and provide a viable and robust service. Such a conception would afford service providers flexibility in ensuring that their service remains secure, complies with present or future regulations regarding content monitoring, and provides useful functionalities such as email search. This way, if a storage service accesses the consumer data to scan for malware[128] or provide search results, the consumer data can still qualify for disclosure protections.

C. EXAMPLES OF USING THE ANALYTICAL FRAMEWORK

With the proposed analytical framework explained, it would be beneficial to see how it works on modern online services. First, Part C first analyzes Gmail's webmail service through the lens of the three-step framework. This example shows how the framework works in a scenario already analyzed by and litigated in courts (i.e., scrutinizing email/webmail under the SCA), thus serving as a tool to make baseline comparisons and expectations. Further, this example also illustrates the intricacies of the SCA and demonstrate how services are moving towards (or already using) an ad-based business model that does not analyze user communications. Second, Part C analyzes Facebook's

128. *See* Davey Winder, *Google Confirms New AI Tool Scans 300 Billion Gmail Attachments Every Week*, FORBES (Feb. 28, 2020, 7:35 AM), https://www.forbes.com/sites/daveywinder/2020/02/28/google-confirms-new-ai-tool-scans-300-billion-gmail-attachments-every-week/?sh=812e4b3edd1f [https://perma.cc/S5CV-SQUX].

Messenger service. This example shows how most, if not all, messaging services can be directly analogized to email for analysis under the SCA. Third, Part C applies the three-step framework to Facebook Wallposts. This example covers non-messaging social media services where the communications are often sent to mass groups of people ("friends" or "followers") while having no "read receipt" feature, showing how the lack of the feature creates unique challenges under the current SCA jurisprudence that the proposed framework and conceptualizations help mitigate.

1. Gmail

Imagine Person A, who uses Gmail, sending an email to Person B who, for simplicity's sake, also uses Gmail. As Person A writes the email, a "draft" is created and periodically saved. Once A finishes writing and sends the email to Person B, the draft is completed and Gmail removes the "Draft" label and affixes the "Sent" label to the email. This way, if A wants to read the email again, they can do so by clicking on the "Sent" label and accessing all the sent emails.[129] At the same time, a copy is created and sent to Person B's inbox, where it is affixed with the "Inbox" label. Therefore, there are (at least) two copies: Person A's copy with the "Sent" label and Person B's copy with the "Inbox" label.

a) Step One: Determining if Gmail is an ECS, RCS, or Neither

Determining whether Gmail is an ECS, RCS, or neither depends on which copy of the email is in question. For example, if the government is trying to compel disclosure of Person *A*'s copy (either while in the "draft" stage or after it was sent), then Gmail is acting as an RCS, not an ECS. In either the draft or the sent stage, Person *A*'s copy cannot be viewed, edited, or responded to by any other non-sending, non-service party. It also cannot be said that either copy is in "intermediate storage . . . incidental to the electronic transmission" of the communication or that either copy is being stored for "backup protection" in relation to a communication service.[130] It is more compelling to think of Gmail indefinitely storing Person *A*'s copy as providing a storage service for user convenience, i.e., later access by Person *A*.

In regard to Person *B*'s copy of the email, Gmail's status as an ECS or RCS depends on whether Person *B* has accessed the email or not. As described above,[131] if Person *B* has not yet accessed the email, then the email remains in

129. *See* Facebook, Inc. v. Superior Ct. of San Diego Cnty., 471 P.3d 383, 372 n.14 (Cal. 2020) (Cantil-Sakauye, C.J., concurring).

130. *See* 18 U.S.C. §§ 2510(17)(A)–(B).

131. *See supra* Section IV.B.1.a.

"intermediate storage . . . incidental to the electronic transmission," and Gmail still serves as an ECS.[132] On the other hand, if Person *B* has already accessed the email, then Gmail, in relation to that particular email, is an RCS: the email "is no longer 'incident to transmission[,]' . . . it is just in remote storage like any other file held by an RCS."[133]

In sum, Step One finds that Gmail is an RCS with respect to Person *A*'s copy, and that Gmail is an ECS *or* RCS with respect to Person *B*'s copy, depending on timing.

b) Step Two: Gmail's Business Model and Interactions with Data

Gmail has two primary business models. There is a paid version of Gmail sold to businesses as part of Google Workspace, a collection of productivity and "collaboration tools like Gmail, Calendar, Meet, Chat, Drive, Docs, . . . and more."[134] There is also a free version of Gmail that serves "useful and relevant," i.e., targeted, ads to its users.[135] This Section will focus on the free version of Gmail and how it interacts with its subscribers' communications.

While Gmail's free version uses targeted ads, it does not share user content with third parties. Gmail's targeted ads are "shown to [a person] based on [their] online activity while [they're] signed into Google."[136] Google does not "scan or read . . . Gmail messages to show . . . ads."[137] Further, Google "does not sell . . . personal information," nor does it "share . . . personal information with advertisers, unless [a subscriber] ha[s] asked [them] to."[138] Gmail also allows users "[t]o opt-out of the use of personal information for personalized

132. *See* 18 U.S.C. § 2510(17)(A).

133. Kerr, *supra* note 18, at 1216.

134. GOOGLE WORKSPACE, https://workspace.google.com/ [https://perma.cc/3TKF-JTH9] (last visited Aug. 17, 2022).

135. *How Gmail Ads Work*, GMAIL HELP, https://support.google.com/mail/answer/6603?hl=en&ref_topic=3394218 [https://perma.cc/85QP-XNHF] (last visited Aug. 17, 2022).

136. *Id.*

137. *Id.* It is worthwhile to note that Google did scan user emails for ad personalization in the past. It stopped this practice in mid-2017. Daisuke Wakabayashi, *Google Will No Longer Scan Gmail for Ad Targeting*, N.Y. TIMES (June 23, 2017), https://www.nytimes.com/2017/06/23/technology/gmail-ads.html [https://perma.cc/NS28-U5YZ]. Even so, the analysis would still remain the same for pre-2017 emails—courts should look at the actual authorization settings in play at the time of the inquiry when determining the scope of the SCA protections. Doing otherwise would, at least to a certain extent, render user authorization changes moot. This can upset user privacy expectations and can lead to fractured protections, depending on how often the authorization settings are changed.

138. *How Gmail Ads Work*, *supra* note 135.

Gmail ads" by "turn[ing] off Ads Personalization."[139] From this information it is evident that, at the very least, Gmail does not access or share user content for targeted advertisements.

While Gmail does not access emails for ad targeting, it does access them to provide user convenience functionalities. For example, Gmail has a search functionality that allows a user to find particular emails by typing in key words or filling in who the sender or recipient was. To present relevant results, Gmail must scan the user's emails to determine which ones appropriately match the inquiry. Further, Gmail also has a function that automatically tags some emails as important. "Gmail analyzes [a user's] new incoming messages to predict what's important, considering things like how [they've] treated similar messages in the past, how directly the message is addressed to [the user], and many other factors."[140] Users can opt out of this functionality. There are many other Gmail functionalities, such as auto-categorization of social media emails, spam and malware detection, and reply and follow-up reminders, that involve Gmail accessing user communications.

In sum, Step Two finds that while Gmail accesses user communications, it does so for supplementary services needed to provide user and system integrity and user convenience.

c) Step Three: Determining Whether Gmail Violates or is Disqualified from the Disclosure Provisions of the SCA

First, turn to Person A. Under the SCA, in order for RCS communications to qualify for disclosure protection, the RCS must not have authorization to access user communications for reasons other than providing storage or processing services.[141] As found in Step Two, while Gmail may not access its user emails to provide targeted advertisements, it still does so to provide several supplemental functionalities, such as search, spam detection, and auto-categorization. For some of these services, such as search and follow-up reminders, Gmail accesses user emails while serving as an RCS. Such functionalities, however, should be looked at as necessary for providing a modern, robust email service. Realistically, no email service can survive without a built-in search function. These functionalities exist to enhance user convenience by providing a service where a user can easily access and interact

139. *Id.*

140. Anne P. Mitchell, *Gmail 'Skip the Inbox' Filter Not Working? This May Be Why – and How to Fix It*, INTERNET PATROL (Apr. 8, 2019), https://www.theinternetpatrol.com/gmail-skip-the-inbox-filter-not-working-this-may-be-why-and-how-to-fix-it [https://perma.cc/NZ22-WBME].

141. *See* 18 U.S.C. §§ 2702(a)(2), 2703(b).

with their past emails. As such, these supplementary services should be conceptualized under the umbrella of storage services. And as § 2702(a)(2) extends protection as long as the RCS is only authorized to access user content for storage or computing services, Person A's copy of the email should be protected from voluntary and compelled disclosure.

Now turn to Person B. As seen in Step One, Gmail may either be acting as an ECS or an RCS, depending on whether B has accessed the email or not. If Person B has not accessed the email, then Gmail is acting as an ECS. And as established in Step Two, Gmail does not share the contents of email communications to third parties for advertising purposes. As such, it is unlikely that Gmail shares user communications for any other reason as well. Under this assumption, Gmail, acting as an ECS, does not violate the SCA's voluntary disclosure provision set forth in § 2702(a)(1). If Person B has accessed the email, then Gmail is acting as an RCS. Here, the same reasoning would apply as it did for Person A's email—Gmail, as an RCS, would not be disqualified from the SCA's disclosure provisions, and Person B's email would be protected.

2. *Facebook Messenger*

Now instead of an email exchange, imagine Person A and Person B talking over Facebook Messenger. As in the Gmail example, begin by examining Person A sending a message to Person B. When Person A sends the message, a copy is retained on Person A's chat window, and another copy is sent to Person B's. The analysis in Step One, determining whether Messenger is an ECS, RCS, or neither, is the same as in the Gmail example; messaging and text services can be easily analogized to email services. Messenger, in relation to Person A's copy of the message, acts as an RCS. In relation to Person B's copy, Messenger acts as an ECS if *B* has not yet accessed the message, and an RCS otherwise.

Similar to Gmail, Messenger does not "use the content of . . . messages . . . for ad targeting,"[142] and it does not sell information to advertisers.[143] Instead, targeted ads are shown "based on [the user's] activity across Meta technologies," the user's "activity with other businesses," the user's "activity on other websites and apps," and the user's location.[144]

142. *You Control Your Messenger Experience*, MESSENGER, https://www.messenger.com/privacy [https://perma.cc/6KGL-85ES] (last visited Sep. 8, 2022).

143. *About Facebook Ads*, FACEBOOK, https://www.facebook.com/ads/about/?entry_product=ad_preferences [https://perma.cc/P6A6-U4AX] (last visited Aug. 17, 2022).

144. *Id.*

Messenger, however, does scan and collect information from chats to "improve the product experience[] and keep people safe and secure."[145] For example, Messenger automatically scans chat messages to stop scammers, prevent phishing, safeguard minors,[146] protect users from malware and viruses, and prevent the spread of child exploitation imagery.[147] Further, users can report a chat as violating Facebook's "[C]ommunity [S]tandards," which may prompt human moderators to take a look at the chat and take it down.[148]

While some of these services, such as search or account investigations, involve Messenger accessing user communications as an RCS for reasons unrelated to storage, they still do not disqualify Messenger from the RCS disclosure provisions of the SCA. As in the Gmail example, these functionalities should be seen as supplementary to providing a robust and user-friendly storage service that protects its users and maintains both system and entity integrity. As such, Messenger, as an RCS, should not be disqualified from the SCA's disclosure protections.

Similarly, Messenger, acting as an ECS, does not violate the SCA disclosure provision set forth in § 2702(a)(1). It does not share or knowingly divulge its user communications with third parties, and thus complies with the provision.

3. Facebook Wallposts

Here, imagine Person A posting a status onto their "Facebook Wall." Unlike the Gmail and Messenger examples, there is only one copy of the message—Person A's. By default, only A's Facebook friends can view the wallpost, and they can do so either by going to Person A's profile or by exploring their own "Newsfeed." The analysis done here can be analogized to picture and video sharing services, forums, and other status-update/profile-creation services, such as Twitter and LinkedIn.

a) Step One: Determining if Facebook is an ECS, RCS, or Neither

Facebook, in relation to Person A's wallpost, should be regarded as an ECS for the first six months. The post is viewable by A's friends, who can comment on it, "like" it, or share it, giving credence to the notion that Facebook is acting as a communication service.

145. *You Control Your Messenger Experience*, *supra* note 142.

146. *Id.*

147. Alanna Petroff, *Yes, Facebook is Scanning your Messages for Abuse*, CNN BUS. (Apr. 5, 2018, 10:18 PM), https://money.cnn.com/2018/04/05/technology/facebook-messenger-messages-privacy/index.html [https://perma.cc/Y95K-4JM7].

148. *See id.*

The wallpost can be further analogized to a post made on a private electronic bulletin board service. As described above, courts have repeatedly held that electronic bulletin boards come under the purview of the SCA.[149] And while courts are more split on whether bulletin board services are ECS or RCS, legislative history seems to favor the ECS conception.[150] Using this bulletin board analogy, the court in *Crispin v. Christian Audigier, Inc.* held that Facebook, in relation to a wallpost, is an ECS.[151] In so concluding, the court reasoned that wallposts must be "stored for backup purposes," as they could not be "protectable as a form of temporary, intermediate storage."[152] For Facebook wallposts, "there is no temporary, intermediate step . . . . Unlike an email, there is no step whereby a Facebook wall posting must be opened, at which point it is deemed received."[153] The better conception, however, is that the wallpost is in temporary storage. While Facebook is the literal "final destination"[154] of the wallpost, in a more real sense the final destinations are the actual recipients. This is the same reason why even though a recipient's inbox is the final destination of an email, the email is still considered to be in temporary storage until it is accessed. Therefore, Person A's wallpost should be considered an electronic communication held in temporary storage.

As it is understandably difficult, or even impossible, to confirm when all of A's friends have accessed the wallpost, Facebook should remain an ECS for the default, SCA-drawn lifespan of 180 days.[155] After that time, Facebook should be thought of as an RCS in relation to *A*'s wallpost. At that point, the wallpost should be considered to be in remote storage.

b) Step Two: Facebook's Business Model and Interactions with Consumer Data

Facebook, like most social media companies, relies on a business model built on targeted ads. Although it relies on ad targeting, Facebook does not "sell any [user] information to anyone, and . . . never will."[156] It does, however, share limited information with partners and advertisers. The information

149. *See supra* note 115 and accompanying text.

150. *See supra* note 114 and accompanying text; Konop v. Hawaiian Airlines, Inc., 302 F.3d 868, 875 (9th Cir. 2002).

151. *See* 717 F. Supp. 2d 965, 988–89 (C.D. Cal. 2010).

152. *Id.* at 989.

153. *Id.*

154. *Id.* at 988 (quoting Snow v. DIRECTV, Inc., No. 2:04-CV-515FTM33SPC, 2005 WL 1226158, at *3 (M.D. Fla. May 9, 2005), *report and recommendation adopted*, 2005 WL 1266435 (M.D. Fla. May 27, 2005), *aff'd*, 450 F.3d 1314 (11th Cir. 2006)).

155. *See supra* note 113 and accompanying text.

156. *Data Policy*, FACEBOOK (Aug. 21, 2020), https://www.facebook.com/full_data_use_policy [https://perma.cc/8UJ7-G4XJ].

shared is "aggregated statistics" and "reports about the kinds of people" that interact with the advertiser's Facebook Pages, posts, and advertisements; importantly, Facebook does not share user communications with third parties.[157]

Nonetheless, in order to personalize ads and provide an integrated experience across its numerous products, Facebook has the authorization to, and does, collect "the content [and] communications" that users provide when using their products.[158] Facebook uses this information to provide, personalize and improve their products."[159] In essence, this means that they "connect information" across Facebook Products in order to provide a seamless experience and to better tailor ads.[160] At a broad level, in selecting what ads to show a particular individual, Facebook uses location data; profile information, such as age, gender, education, job title, and relationship status; interests, including listed hobbies, activities, and page interactions; and behaviors, such as mobile and device usage.[161]

Like Messenger, Facebook, aside from collecting user information for ad personalization, also scans and collects information to provide supplementary services necessary to "[p]romote safety, integrity and security."[162] Facebook uses this information to "combat harmful conduct, detect and prevent spam and other bad experiences, . . . [and] investigate suspicious activity or violations of [Facebook's] terms or policies."[163]

c) Step Three: Determining Whether Facebook Violates or is Disqualified from the Disclosure Provisions of the SCA

In the first six months, while Facebook is acting as an ECS in relation to A's wallpost, Facebook will not violate the disclosure provision set forth in § 2702(a)(1) of the SCA. As Facebook does not sell or share the contents of a user's wallposts (or any other user communications, for that matter), it does not "knowingly divulge . . . the contents of a communication" to a third party.[164] Therefore, it complies with the ECS disclosure provision of the SCA.

157. *See id.*
158. *Id.*
159. *Id.*
160. *See id.*
161. *See id.*; *About Reaching New Audiences*, FACEBOOK FOR BUS. (Nov. 11, 2020), https://www.facebook.com/business/help/717368264947302?id=176276233019487&helpref=page_content [https://perma.cc/4TYS-84WH].
162. *See Data Policy*, *supra* note 156.
163. *Id.*
164. *See* 18 U.S.C. § 2702(a)(1).

Perhaps counterintuitively due to its ad personalization, Facebook, acting as an RCS in relation to *A*'s wallpost after 180 days, will not be disqualified from the voluntary and compelled disclosure provisions of the SCA. Facebook is authorized to and does access user communications, including wallposts, for several reasons, including search, suspicious activity investigation, and spam and malware detection. Many of these interactions, however, happen while Facebook is acting as an ECS, i.e., shortly after the communication is posted.[165] For example, Facebook has a number of A.I. systems that can automatically detect, and sometimes take down, hate speech, misinformation, pornography, and other policy-violating materials.[166] After Facebook has transformed into an RCS, however, its communication interactions are likely limited to search and user report investigations. Like with Messenger, these interactions should be viewed as supplementary to providing a secure and user-friendly storage service.

Specifically in regard to targeted advertisements, Facebook does not mine or analyze wallposts to personalize ads. Rather than analyzing status updates and posted pictures, Facebook looks at profile information, life updates, listed hobbies, and user interactions with business and interest pages. Even if wallposts were analyzed for ad purposes, it is likely that the analysis would happen shortly after the communication was posted and then be quickly integrated into the user's ad profile. It is unlikely that Facebook would analyze a wallpost 180 days or later for ad personalization. As such, Facebook, acting as an RCS in respect to *A*'s wallpost, should not be disqualified from the voluntary and compelled disclosure provisions of the SCA.

D. LESSONS AND EXTRAPOLATIONS FROM USING THE FRAMEWORK

In each of the Gmail, Messenger, and Facebook examples above, the business model theory did not hold that the service was either violating or disqualified from the disclosure provisions of the SCA. For most social media services, this result is to be expected as a consequence of the information-sharing nature of these services and the ad-targeting business models they apply.

---

165. *See generally* Jeremy Kahn, *Facebook Makes Strides Using A.I. to Automatically Find Hate Speech and COVID-19 Misinformation*, FORTUNE (May 12, 2020, 11:21 AM), https://fortune.com/2020/05/12/facebook-a-i-hate-speech-covid-19-misinformation [https://perma.cc/T7U9-EX72] (explaining that Facebook's A.I. systems are getting better at automatically detecting and removing hate speech and misinformation).

166. *See id.*; Mark Sullivan, *Facebook's AI for Detecting Hate Speech is Facing its Biggest Challenge Yet*, FAST CO. (Aug. 14, 2020), https://www.fastcompany.com/90539275/facebooks-ai-for-detecting-hate-speech-is-facing-its-biggest-challenge-yet [https://perma.cc/65CP-6KQ3].

As social media has a general communicative purpose, with users sharing information with their friends and followers, it is compelling to think of these services as ECS. Subscribers use these services to connect with others, participate in conversations, and share various parts of their lives. As such, courts will likely see many of these services as providing, in relation to most pieces of information, "the ability to send or receive . . . electronic communications."[167] Of course, as described above in the Gmail and Facebook examples, some of these services will "drop down" to RCS when the communication in question has been accessed or when enough time, i.e., 180 days, has passed.

Further, even though the business model of most social media services is based on targeted advertisements, that fact is unlikely to prevent the services from complying with the SCA. Most popular social media services do not actually sell or share user communications with third parties to facilitate their ad targeting; indeed, data policies often explicitly state that the service does not sell or share user content.[168] As such, when these services are acting as ECS in relation to a particular communication, they will not be found to violate the disclosure provisions of the SCA.

It is also unlikely that social media services, when they "drop down" to RCS, will be disqualified from receiving voluntary and compelled disclosure protections under the SCA. Most major social media services do not scan and analyze user communications for ad personalization; instead, the services rely on, amongst other things, demographics, profile information, and specified interests.[169] And even if these services were to scan communications for ad

167. *See* 18 U.S.C. § 2510(15).

168. *See, e.g.*, *Data Policy*, *supra* note 156 ("[Facebook does not] sell any of your information to anyone, and [they] never will."); *Twitter Privacy Policy*, TWITTER (June 18, 2020), https://twitter.com/en/privacy [https://perma.cc/WL35-NWT6] ("[Twitter] use[s] information about whom you have communicated with and when (but not the content of those communications) . . . ."); *Reddit Privacy Policy*, REDDIT (Sept. 15, 2020), https://www.redditinc.com/policies/privacy-policy-october-15-2020 [https://perma.cc/PY9H-TKSZ] ("Reddit only shares nonpublic information about you in the following ways. We do not sell this information."); *Privacy Policy*, SNAP INC. (Sept. 14, 2020), https://www.snap.com/en-US/privacy/privacy-policy [https://perma.cc/59EW-QUA4].

169. *See* Petroff, *supra* note 147; *How Gmail Ads Work*, *supra* note 135. The fact that social media services do not analyze user communications is most apparent through privacy/data policies and instructional web pages sharing how the service actually personalizes ads. *See, e.g.*, *About Reaching New Audiences*, *supra* note 161; *Targeting Options for LinkedIn Advertisements*, LINKEDIN HELP, https://www.linkedin.com/help/lms/answer/722/targeting-options-for-linkedin-advertisements?lang=en [https://perma.cc/R4JM-W3HR] (last visited Aug. 17, 2022); *About Targeting for Video Campaigns*, YOUTUBE HELP, https://support.google.com/youtube/answer/2454017?hl=en [https://perma.cc/UVJ2-34TG] (last visited Aug. 17, 2022).

personalization, it is likely to happen early after a communication is posted, while the service is still acting as an ECS.[170] Further, while many social media services have authorization to access user communications for functionalities such as search and user report investigations, these functionalities should be considered supplementary to the core RCS, and thus allowed.

While the business model theory will not diminish the disclosure protections afforded to most social media services, courts should still use business model scrutiny and the three-step framework in their analysis. Due to the sheer and expanding number of social media services and the various business models available, it is still important to use context-specific, case-by-case analysis. Courts should consider each service separately and, using business model scrutiny and the three-step framework, should determine if the social media service violates or is disqualified from the disclosure protections of the SCA. Adhering to this framework will result in better reasoned and structured analysis, which will not only be responsive to modern technologies and services, but will also help clarify the muddy waters of SCA jurisprudence.

V. CONCLUSION

The proposed framework and conceptualizations laid out in this Note would help modernize courts' understanding of the SCA and guide them through this murky and confusing area of the law. The framework helps courts supplement their analysis, strengthen their structure, and provide consistent results. Moreover, the modernized conceptualization of the SCA underlying the framework helps resolve the main inherent oddity of the business model attack: it seeks to use a privacy statute to diminish the privacy rights of internet users, specifically in a context where they are already vulnerable. By viewing supplementary functionalities of services as a broader part of the service itself, courts can ensure that users and their communications are appropriately protected in the digital realm. When analyzing the business model attack through this modernized lens, a conclusion becomes clear: due to the nature of current ad-targeted business models, the attack will have little, if any, impact on social media services and their SCA disclosure protections.

170. *See* 18 U.S.C. §§ 2702(a)(1)–(2). An RCS provider, if it wants to receive SCA disclosure protections, cannot access user communications for any reasons unrelated to providing storage and processing services. There is no such restriction for ECS providers. Therefore, an ECS service can access user communications for ad targeting.

DIVIDING TO UNIFY THE INTERNET

Joseph A. Kroon†

I. INTRODUCTION

The internet began as a place for the world to connect. A place for information to flow freely. A place for people across the globe to communicate with each other within fractions of a second. The internet was to be "a world that is both everywhere and nowhere."[1] At least, that was what the early internet pioneers wanted it to be.

Today, the internet is separating into geopolitical spheres, mirroring the exact borders the internet pioneers wanted to dismantle.[2] The "Great Firewall of China" has confined Chinese residents into a China-specific internet sphere.[3] Russia requires all cloud and platform data to be stored within Russian borders.[4] India has enacted similar data localization laws.[5] This is the rise of "Internet Balkanization."

The European Union is the latest geopolitical sphere to erect an internet border. On July 16, 2020, the European Union Court of Justice (hereinafter "CJEU") decided *Data Protection Commissioner v. Facebook Ireland, Ltd* (hereinafter "*Schrems II*").[6] Maximillian Schrems, an Austrian privacy activist, brought the case.[7] Schrems requested to suspend data transfers from the European Union to the United States for two reasons. First, he argued that the EU–U.S. Privacy Shield, the EU–U.S. data transfer treaty, did not provide EU citizens with adequate protection mandated by the General Data Protection Regulation (hereinafter "GDPR") and the EU Charter.[8] Second, he argued that Standard

---

DOI: https://doi.org/10.15779/Z382J6854B

† J.D., University of California, Berkeley School of Law, 2022.

1. John Perry Barlow, *A Declaration of Independence of Cyberspace*, ELEC. FRONTIER FOUND. (Feb. 8, 1996), https://www.eff.org/cyberspace-independence.
2. *Id.*
3. *The Great Firewall of China*, BLOOMBERG (Nov. 5, 2018), https://www.bloomberg.com/quicktake/great-firewall-of-china.
4. Anupam Chander & Uyên P. Lê, *Data Nationalism*, 64 EMORY L.J. 677, 712–15 (2015).
5. *Id.* at 694–97.
6. Case C-311/18, Data Protection Commissioner v. Facebook Ireland, Ltd., 2020 E.C.R. (hereinafter *Schrems II*).
7. *Id.* ¶ 50.
8. *Id.* ¶ 55.

Contractual Clauses (SCCs), a contract mechanism which allows private entities to transfer data, were incompatible with the Charter.[9] The CJEU ultimately held that the EU–U.S. Privacy Shield was invalid. It also held that the SCCs were still acceptable but reaffirmed their limitations.[10] Therefore, the CJEU *Schrems II* decision limits how data is transferred between the European Union and other countries. Another digital legal border was erected.

This Note argues that *Schrems II* paradoxically promotes Internet Balkanization by maintaining both data users' privacy interests and the open internet ideals. The CJEU rectified high privacy expectations, erected a digital legal border, and fragmented the internet. However, *Schrems II* erects this border by establishing an idea this Note calls "digital autonomy":[11] users' ability to use the open internet free from adverse foreign entities' processing. By establishing digital autonomy, the EU maintained the open internet.

To illustrate, this Note explores the implications of *Schrems II* in a global, digital world. Part II elaborates on the general privacy interest in an increasing global and digital world. It begins with an overview of information privacy, specifically how privacy relates to individual autonomy and how privacy interests have changed in a digital environment. Then it explains what Internet Balkanization is and how it is rising today. Part III of this Note gives an overview of *Schrems II*. It begins with the Snowden revelations and the aftermath of countries adopting more privacy and data localization policies, and then it elaborates on the pertinent GDPR provisions and the *Schrems II* decision. Finally, Part IV argues that *Schrems II* maintains a core open internet ideal—autonomy of individual data users—by promoting Internet Balkanization.

II. PRIVACY INTEREST IN A GLOBAL, DIGITALIZED WORLD

Today, information is one the most valuable pieces of capital,[12] but according to Maximillian Schrems, information is "none of a company's business";[13] and these two forces conflicted with each other in a recent,

9. *Id.*

10. *Id.* ¶ 105.

11. *See infra* pp. 53–55.

12. *The World's Most Valuable Resource Is No Longer Oil, but Data*, THE ECONOMIST (May 6, 2017), https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

13. *See FAQs*, NOYB, https://noyb.eu/en/faqs (referencing how NOYB, Maximillian Schrem's privacy advocacy group, believes that data user's privacy is "none of a company's business").

dominating CJEU case. But what is "information privacy," the central concern Schrems fought for? And what value does information have?

The first Section of this Part reviews information privacy. It discusses what information privacy initially concerned, how it relates to autonomy, and how it is becoming more pertinent due to the greater accessibility of personal information. The second Section examines an open internet—what it is and its societal benefits—and then discusses the open internet's adversary: Internet Balkanization, its rise, and its pros and cons.

A. INFORMATION PRIVACY AND AUTONOMY IN THE DIGITAL AGE

Privacy is the "claim of individuals . . . to determine for themselves when, how, and to what extent information about them is communicated to others."[14] There are different categories of information privacy concerns. Legal scholars have categorized privacy as decisional privacy,[15] physical privacy, proprietary privacy, and information privacy. *Schrems II* and this Note primarily concern information privacy.

This Section discusses information privacy, autonomy, and the relationship between the two. First, this Section summarizes what information privacy is and how the digital era has made information privacy an ever-increasing issue. Then, this Section discusses what autonomy is, how autonomy is a product of information privacy, and how data users' digital experience impacts their autonomy.

1. *An Overview of Information Privacy and the Digital Era's Impact on Information Privacy*

Information privacy is the right "to control the flow of our personal information."[16] It includes having control over the information that makes up who we are and how society views us.[17] This encompasses information that define individuals objectively (e.g., name, address, age) and information that define individuals subjectively (e.g., political views, sexual preferences, spiritual perspectives).[18]

Society benefits when individuals have control over their information. For instance, privacy promotes different democratic values like freedom of economic choice and freedom of political participation. Privacy allows

14. ALAN WESTIN, PRIVACY AND FREEDOM 7 (1967).

15. Decisional privacy is the individual's right to make choices about their body and their family.

16. Karl Manheim & Lyric Kaplan, *Artificial Intelligence: Risks to Privacy and Democracy*, 21 YALE J. L. & TECH. 106, 118 (2019).

17. *See id.*

18. *See id.*

individuals to buy what they want or vote for who they want without fear that their choices will be broadcasted publicly, resulting in scrutiny. Also, privacy is fundamental to individual autonomy.[19]

However, information privacy is not a right to control all information at all times. For instance, users cannot assert a privacy right if the information has already been disclosed or if the public has a strong interest in the information.

First, information privacy is limited by whether the information has already been disclosed. Privacy law protects individual's direct control over their information, but it does not limit anyone else from sharing the individual's information unless there is a legal obligation prohibiting disclosure.[20] Individuals no longer have complete control over information after it has been disclosed to someone. Thus, privacy is a limited right, narrowed by who lawfully has access to information.

Second, information privacy is limited by public interests. Legislation should make limitations in order to "meet objectives of [the public's] general interests."[21] Recognized interests include counterintelligence, public security, and contractual performance.[22] These are exceptions legislation decided are more important an individual's privacy.[23] Therefore, people should not have monopoles over their data; their data submits to greater public interests.

In sum, information privacy is control over one's information, but that control only extends as far as one shares their data and what other interests the public is concerned with.

---

19. *See generally* Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373 (2000) ("Autonomy in a contingent world requires a zone of relative insulation from outside scrutiny and interference--a field of operation within which to engage in the conscious construction of self.").

20. *See* Nader v. Gen. Motors Corp., 25 N.Y.2d 560, 569 (1970) (recognizing that interviewing third parties could hardly be an invasion of plaintiff's privacy). Some statutes prohibit third parties from sharing information that individuals have disclosed to them. This includes the GDPR.

21. Case C-311/18, Data Protection Commissioner v. Facebook Ireland, Ltd., 2020 E.C.R. ¶ 174.

22. *See* 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), art. 49, 2016 O.J. (L 119) 1 [hereinafter GDPR] (enumerating some exceptions where data transfers can occur without an adequacy agreement nor Safeguard Agreement).

23. *See id.*

However, the right to privacy was theorized pre-cell phones, pre-social media, pre-globalization,[24] and time has exacerbated information privacy concerns in two ways. An unprecedented amount of data is being collected, and that data discloses more intimate details than ever before.

Today, more data than ever is being collected. At its inception, information privacy concerned the details one normally disclosed while in the confines of their home: information individuals usually shared with their "inner circle," close family and friends.[25] In the digital era, parties in the "inner circle" are now applications we use, websites we visit, or surveillance camera on the streets we walk.[26] For example, a data user's private social media posts are still legally accessible by going through an entity the data user has previously authorized access to. The internet has given companies so much information that they now have a "God's eye" of the market.[27] Facebook knows what people like; Apple knows what apps people download; Amazon knows what people buy. This knowledge gives companies insight into the entire society.

Additionally, data users now disclose more intimate details than before. By using cultivated data, artificial intelligence (AI) can find surprising patterns in data sets that were impossible to discover with previous technology.[28] For example, one study used telephone metadata (e.g., the numbers dialed and length of calls) to infer that a person had cardiac arrhythmia.[29] Indeed, the U.S. Supreme Court has recognized this concern in the context of GPS location.[30] For example, developers used Instagram feeds to create an AI that predicts

---

24. Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193 (1890).

25. *Id.* at 195.

26. *Facial Recognition Technology*, ACLU, https://www.aclu.org/issues/privacy-technology/surveillance-technologies/face-recognition-technology.

27. *See* THE ECONOMIST, *supra* note 12.

28. There is an argument to be had about whether AI conflicts with information privacy because AI can be created with anonymous data. There are two main reasons why anonymized data is not a simple fix for AI. First, some data independently identifies data users (e.g., name, address, phone number, etc.) Because this data is independently identifiable, they are prohibited from AI developers wanting to use the anonymized data exception. Second, a combination of different unidentifiable data together can identify a data user. Although information may originally be anonymized, it can eventually be de-anonymized. For more of this discussion, see Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010).

29. Bjorn Carey, *Scientists Show Telephone Metadata Can Reveal Surprisingly Sensitive Personal Information*, PHYS ORG (May 17, 2016), https://phys.org/news/2016-05-scientists-metadata-reveal-surprisingly-sensitive.html.

30. United States v. Jones, 565 U.S. 400, 416 (2012) (Sotomayor, J., concurring) ("I would ask whether people reasonably expect that their movements will be recorded and aggregated in a manner that enables the government to ascertain, more or less at will, their political and religious beliefs, sexual habits, and so on.").

data users' depression levels.[31] In another example, AI used facial recognition to predict someone's sexual orientation.[32] This is information data users may have disclosed in intimate settings but now unknowingly disclose to the internet.

Information privacy originally concerned the right to control our data. But data, today, is colossal.

2. *An Overview of Autonomy and How Data Users Have Autonomy in a Digital Environment*

Abstractly, autonomy is the freedom to make one's own decisions, detached from adverse influences. Although individual autonomy is not a modern phenomenon, the digital landscape has altered how autonomy can be impacted. Autonomy is an individual's capacity to make their own decisions, but in the digital landscape, more and more adverse influences impact an individual's autonomy. But maintaining information privacy protects autonomy.

Autonomy is a person's capacity to make their own decisions.[33] It is to be free from external influences that manipulate one's thought process: free form adverse political influences, free form social influences, and free to "engage in the conscious construction of self."[34] The idea that people are free to be oneself, to be autonomous, is a fundamental belief in democratic societies.[35] Autonomy promotes a "vital diversity" of ideas that establish the foundation for democratic debates.[36] However, realistically being completely free from adverse influences is unachievable; it is an aspirational goal.

In the digital landscape, independence from adverse digital influences (i.e., a digital third party) provides users an environment to become autonomous. These adverse digital influences could be other data users (e.g., other data users people communicate with), websites data users visit (e.g., manipulating the data user's experience), or possible surveillance (e.g., the mere knowledge of an adverse user's presence could manipulate a data user's experience). The purest

---

31. Andrew G. Reece & Christopher M. Danforth, *Instagram Photos Reveal Predictive Markers of Depression*, 6 EPJ DATA SCI. 15, 15–16 (2017).

32. Yilun Wang, Michal Kosinski, *Deep Neural Networks Are More Accurate than Humans at Detecting Sexual Orientation from Facial Images.*, 114 J. PERSONALITY & SOC. PSYCHOL. 246 (2018) (using users' dating app profiles to create an AI system that predicts peoples' sexual orientation).

33. Cohen, *supra* note 19, at 1424 (establishing that autonomy is "how we develop the capacity and facility for choice").

34. *Id.* at 1424.

35. *Id.* at 1426.

36. *Id.* at 1425.

illustration depicting full data user autonomy would be a person surfing through the internet with no surveillance, no tracking, and no incitements manipulating the user's next step. In theory, the data user would visit any website whenever they wanted or do anything they want, without disclosing a hint of their identity. That is not reality, and for good reasons too.[37] But as an illustration, that is closest experience to complete digital autonomy.

Once adverse parties get involved, autonomy erodes. Attaching one's identity in a chat room may incentive them to post less-controversial comments; a search engine operator feeding curated articles may direct the data user to visit different pages than they originally intended.[38] This is the more realistic digital experience. The more third parties manipulate a data user's digital experience, the more a data user risks losing autonomy.[39] A one-off third party may influence a data user to visit an online retail store before they proceed with their intended business. But as more parties become involved, more of the data user's actions become susceptible to the third party's whims.[40] Arguably, the data user chose to visit a different website (a click-bait article perhaps) due to their own interests therefore never losing their autonomy. However, that mistakes the user's independent reaction—independently clicking the click-bait article link—as autonomy. This is an incorrect conclusion because a data user's autonomy comes from their original intentions, what they originally wanted to do. Although the data user may have originally wanted to research the article's topic, they may not have originally gone to that specific click-bait article. The more adverse influences there are, the more a data user's original intentions are manipulated, inevitably losing a portion of their digital autonomy.

To achieve autonomy, digital or otherwise, information privacy is necessary. Arguably, to be truly autonomous, people need to be relatively insulated from external influences, influences that would manipulate a "conscious construction of self."[41] Privacy nurtures this insulated environment and prevents "pervasive monitoring" that would push individuals "toward the bland and the mainstream," stifling individualism.[42] In a democratic society,

37. There are some instances we would want digital monitoring. For example, sometimes we want websites to guide our experience, resulting in a more efficient experience. But also, we want some monitoring to track egregious actions like in sex trafficking or child pornography.

38. *See* Jessica van der Schalk, *The Quest for Autonomy in the Digital Age*, FREEDOM LAB (Aug. 5, 2019), https://freedomlab.org/the-quest-for-autonomy-in-the-digital-age/.

39. *See* Cohen, *supra* note 19, at 1424.

40. *See id.*

41. *Id.*

42. *Id.* at 1426.

individualism is valued because it propagates fundamental democratic values like diverse ideas, diverse voices. Therefore, information privacy is "a constitutive element of a civil society."[43]

B. AN OVERVIEW OF THE OPEN INTERNET, AND ITS ADVERSARY: INTERNET BALKANIZATION

Privacy is not an absolute right.[44] In some instances, access to information counteracts data users' privacy interests. For one, the "[f]lows of personal data . . . [is] necessary for the expansion of international trade and international cooperation," the benefits of an open internet.[45]

This Section begins by summarizing what an "open internet" is and an open internet's economic and social benefits. Then, this Section discusses the rise of an open internet's adversary: Internet Balkanization, what it is, what causes it, and what its debatable pros and cons are.

1. An Overview of the Open Internet

As a preliminary question, what is an open internet? At the beginning of the internet, early pioneers believed the internet to be a place "naturally independent of the tyrannies [governments] seek to impose."[46] That is, governments have no place to confine internet users. This, ideally, would create a "world where anyone, anywhere may express [their] beliefs, no matter how singular, without fear of being coerced into silence or conformity."[47] An open internet is an open frontier of information, free form sovereign confines.

By enabling people to be connected, an open internet generates economic and social benefits.[48] From an economic viewpoint, an open internet minimizes production costs for companies.[49] It also facilitates more international trade, with both business partners and customers.[50] Also, an open

43. *Id.* at 1427.

44. *See* Case C-311/18, Data Protection Commissioner v. Facebook Ireland, Ltd., 2020 E.C.R.. ¶ 172.

45. GDPR, *supra* note 22, recital 101.

46. Barlow, *supra* note 1.

47. *Id.*

48. GDPR, *supra* note 22, recital 6 ("Technology has transformed both the economy and social life, and should further facilitate the free flow of personal data within the Union and the transfer to third countries and international organisations, while ensuring a high level of the protection of personal data.").

49. *See* Dillon Reisman, *Where Is Your Data, Really?: The Technical Case Against Data Localization*, LAWFARE (May 22, 2017), https://www.lawfareblog.com/where-your-data-really-technical-case-against-data-localization (discussing how data localization is expensive work).

50. OECD DIGITAL ECONOMY POLICY PAPER, ECONOMIC AND SOCIAL BENEFITS OF INTERNET OPENNESS, at 18 (2016) [hereinafter ECONOMIC AND SOCIAL BENEFITS OF INTERNET OPENNESS].

internet stimulates innovation.[51] Much of the digital revolution was enabled by resources being easily communicated via the internet.[52] On the other hand, arguments are being made that an open internet may not be as good for economic growth as some may believe.[53] This is an important debate to be had, but not the focus of this Note.

Also, an open internet has social benefits. With an open internet, people have a forum to communicate with other people regardless of their geographic location. Open communication enables people to connect with close ones, access education, or connect with social services like medical professionals. An open internet "facilitates the exchange of knowledge, ideas, interests, and viewpoints" from news media sources across the globe.[54] Internet users can also express their beliefs to a wider audience. Open internet's accessibility enabled journalists to investigate and widely report about the Uighur "re-education" camps[55] or about the Rohingya genocide.[56] An open internet enables one's voice to be heard, regardless of where they are in the world.

However, data users cannot rely on an open internet. states regularly adopt measures that fragment it. By fragmenting the open internet, data users lose the early cyber pioneers' dream for the internet. Scholars have called this fragmenting phenomenon "Internet Balkanization."

2. *The Rise of Internet Balkanization*

Opposing the open internet is Internet Balkanization. Internet Balkanization refers to the global internet fragmenting into different geopolitical borders, preventing internet users from accessing each other due to technical, legal, or physical barriers.[57] The internet originated as a universal internet, one spanning across all nations without borders. However, over the

---

51. *Id.*

52. Flaviu Mircea, *Impacts of the Digital Revolution*, MEDIUM (Nov. 26, 2018), https://medium.com/@flaviu.mirc/impacts-of-the-digital-revolution-e561e1cbc8bd ("Another important aspect of the Digital Revolution is the constant development of the communication technology, the main components being the widespread use of Internet and mobile phones, as well as other mobile communication devices.").

53. Jennifer Daskal, Paul Ohm & Pierre de Vries, *Debate: We Need to Protect Strong National Borders on the Internet*, 17 COLO. TECH. L.J. 13, 26 (2018); *see also* ECONOMIC AND SOCIAL BENEFITS OF INTERNET OPENNESS, *supra* note 50, at 22.

54. ECONOMIC AND SOCIAL BENEFITS OF INTERNET OPENNESS, *supra* note 50, at 23.

55. *Data Leak Reveals How China 'Brainwashes' Uighurs in Prison Camps*, BBC (Nov. 24, 2019), https://www.bbc.com/news/world-asia-china-50511063.

56. *Myanmar's Genocide Against Rohingya Not over, Says Rights Group*, THE GUARDIAN (Nov. 23, 2020), https://www.theguardian.com/world/2020/nov/23/myanmar-is-still-committing-genocide-against-rohingya-says-rights-group.

57. Kevin Werbach, *The Centripetal Network: How the Internet Holds Itself Together, and the Forces Tearing It Apart*, 42 U.C. DAVIS L. REV. 343, 373–74 (2008).

past decade states have started adopting more legal barriers, creating "internet borders."[58] Thus, the open internet is fragmenting into geopolitical internet spheres.[59]

These internet borders, unlike geographic borders, are not defined by one line. They change based on the type of data and what entities can do with the data. For example, some states prohibit a certain type of data from crossing its internet border, but more pervious for other types of data.[60] Or on the other hand, some states' internet borders are due to the lack of internet cables, a mechanical border.[61] These are all ways nations created internet borders, fragmenting the global internet.

Originally, more authoritarian states had internet barriers, but today more democratic states are also adopting internet borders. China has the "Great Chinese Firewall,"[62] which uses filters to block certain IP addresses, keywords, internet addresses, and so on.[63] As another example, Iran wants to make its own internet: an "isolated domestic intranet that can be used to promote Islamic content."[64] In recent years, however, imposing internet borders has spread beyond authoritarian states. Post-Snowden, more democratic states have also adopted internet barriers.[65] Germany proposed to build an EU internet network, intending to keep data user information within the EU.[66] Australia adopted the Personally Controlled Electronic Health Records Act, which prohibits health data from leaving Australia's internet borders.[67] These states' willingness to adopt internet borders, and the active steps they've taken to do so, makes Internet Balkanization an inevitability.

Although Internet Balkanization is the open internet's adversary, Internet Balkanization does not necessarily counteract the open internet's benefits. It is

58. Sally Adee, *The Global Internet Is Disintegrating. What Comes Next?*, BBC (May 14, 2019), https://www.bbc.com/future/article/20190514-the-global-internet-is-disintegrating-what-comes-next (discussing how North Korea is secluded because it lacks internet cables connecting it to the rest of the world).

59. This Note uses the term "internet spheres" to describe the different geopolitical fragments the global internet is separating into.

60. *See* Chadner & Lê, *supra* note 4, at 683. Australia has the Personally Controlled Electronic Health Records (PCEHR) Act which prohibits the transfer of healthcare data outside of Australia, with some exceptions.

61. Adee, *supra* note 58.

62. *Id.*

63. *Id.*

64. Matt Vasilogambros, *Iran's Own Internet*, THE ATLANTIC (Aug. 29, 2016), https://www.theatlantic.com/news/archive/2016/08/irans-own-internet/497894/; *Iran Rolls out Domestic Internet*, BBC (Aug. 29, 2016), https://www.bbc.com/news/technology-37212456.

65. For a more extensive discussion about the rise of data localization, see *infra* p. 15.

66. Chander & Lê, *supra* note 4, at 694.

67. *Id.* at 683.

important to understand Internet Balkanization's pros and cons before resolving its potential adverse effects.

3. *The Pros and Cons of Internet Balkanization*

Recognizing that Internet Balkanization is happening, the pros and cons should be assessed. This Section begins by discussing Internet Balkanization's three main benefits: it rebalances internet control, adds beneficial friction to innovation, and possibly supports economic development by providing space for smaller companies. This Section also discusses Internet Balkanization's two main detriments: the downsides of centralizing data and possibly deterring economic development.

First, one possible benefit is that Internet Balkanization rebalances governmental control over local communication, speech, and commerce. Internet Balkanization creates internet "borders," like geographic borders. Parallel to how a state controls what goes through geographic borders, Internet Balkanization shifts more power to states.[68] This shift gives states more control over what goes in and out of states' internet, which can be beneficial. For example, in the Cambridge Analytical scandal,[69] the American government could have had more control over how Cambridge Analytica used American data for campaign ads. However, countries can use internet borders problematically, like geographic borders. These issues will be discussed later in this Section.[70] Overall, creating these internet borders gives countries the opportunity to properly protect data users' internet experience.

Second, another possible benefit is that Internet Balkanization adds friction to innovation so that engineers can incorporate human values into the systems.[71] Without any friction, the internet is uncontrolled, a digital Wild West. This makes the internet extremely efficient, but it also makes the internet susceptible to malicious manipulation or adverse consequences. For example, adding friction—coined as "desirable inefficiency"—creates a space for corrective human intervention. There is time to assess the technology before

68. *See* Daskal et al., *supra* note 53, at 21; *see also* David R. Johnson & David Post, *Law and Borders - the Rise of Law in Cyberspace*, 48 STAN. L. REV. 1367, 1370 (1996) ("The Net thus radically subverts the system of rule-making based on borders between physical spaces, at least with respect to the claim that Cyberspace should naturally be governed by territorially defined rules.").

69. Nicholas Confessore, *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far*, N.Y. TIMES (Apr. 4, 2018), https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html.

70. *Infra* p. 20.

71. Note that this can also be a con because it would disrupt innovation, which harms everyone. This point will be discussed below.

engineers implement it. This is a "design strategy for building systems that are fairer and more trustworthy."[72]

The third possible benefit is that Internet Balkanization supports economic development by providing a space for fair competition. Internet Balkanization uncontestably increases the barriers to entry for companies.[73] Countries set requirements about how the tech companies can use the internet. For instance, South Korea has a law that makes it difficult to build location services on foreign APIs (application programming interfaces).[74] Australia prohibits transferring any personally identifiable health data outside of Australia.[75] Thus companies need to build data centers within Australia or outsource to local services.[76] All of this results in companies allocating resources to comply with the requirements.

Restrictions also make it more difficult for companies to utilize the internet by creating legislative hurdles, thus making it harder for mid-size and smaller companies to enter markers. Arguably, only the Tech Titans (e.g., Amazon, Google, Facebook) have the resources to comply with different requirements across borders.[77] However, even the largest companies do not have enough resources to comply with every country's restriction.[78] Companies must choose which markets to enter and which ones to leave alone. In the left-alone markets, smaller, competitive companies can grow, ideally to a size that would compete with the large tech companies. Therefore, Internet Balkanization potentially promotes economic development by leaving a space for smaller companies to grow to a competitive size.

There are two main cons to Internet Balkanization: (1) the cautions of centralizing data and (2) the economic concerns.

First, centralizing data within geopolitical spheres risks violating human rights. Internet Balkanization makes it hard for outside parties to access information by adding more barriers to the data and by centralizing the data.[79] Information is opaque, and this opaqueness risks violating human rights. Countries can more easily censor information, stifling dissenting speech.[80] In

72. Paul Ohm & Jonathan Frankle, *Desirable Inefficiencies*, 70 FLA. L.R. 777, 785 (2018).
73. *See* Daskal et al., *supra* note 53, at 22, 26.
74. Chander & Lê, *supra* note 4, at 704.
75. *Id.* at 683.
76. *Id.*
77. *See* Daskal et al., *supra* note 53, at 26.
78. *Id.* at 22.
79. ECONOMIC AND SOCIAL BENEFITS OF INTERNET OPENNESS, *supra* note 50, at 23.
80. *See* Jeffrey Rosen, *Google's Gatekeepers*, N.Y. TIMES (Nov. 28, 2008), https://www.nytimes.com/2008/11/30/magazine/30google-t.html; Clifford J. Levy, *Russia Uses*

2022, Russia suppressed online dissenters of the Ukrainian invasion.[81] Saudi Arabia has imprisoned several journalist who "stray away from the pro-government narrative."[82] The Electronic Frontier Foundation, ACLU, and public interest organizations have expressed that Internet Balkanization "undermines [the] rights [to Freedom of Expression and association] and threatens the potential of the Internet as a powerful tool for advancing human rights and democracy."[83]

Second, Internet Balkanization arguably deters economic development by increasing barriers to entry. Although it may promote economic development because it leaves a space for small to mid-size companies to grow.[84] The flip side argues that Internet Balkanization stifles tech developments by ossifying large tech companies. Arguably, only the Tech Titans (e.g., Amazon, Google, Facebook) have the resources to comply with different border requirements.[85] Therefore, they are the only ones with the resources to span across multiple internet spheres. Only large tech companies can penetrate the large markets because of the barriers to entry.[86] The result is that Internet Balkanization ossifies large tech companies' power in major markets, chilling the market.

Because of the inevitable development of Internet Balkanization, states should guide Internet Balkanization in a way that protects the original open internet benefits while maintaining data users' information privacy rights, particularly if countries want to continue benefiting from the global, open internet. And one way to maintain the ideals of the open internet is to provide individuals more autonomy over their data. *Schrems II* accomplishes this.

Microsoft to Suppress Dissent, N.Y. TIMES (Sept. 11, 2010), http://www.nytimes.com/2010/09/12/world/europe/12raids.html.

81. Robert McMahon, *Russia Is Censoring News on the Ware in Ukraine. Foreign Media Are Trying to Get Around That*, COUNCIL ON FOREIGN RELS. (Mar. 18, 2020), https://www.cfr.org/in-brief/russia-censoring-news-war-ukraine-foreign-media-are-trying-get-around.

82. *Ten Most Censored Countries*, COMM. TO PROTECT JOURNALISTS (Sept. 10, 2019), https://cpj.org/reports/2019/09/10-most-censored-eritrea-north-korea-turkmenistan-journalist/.

83. Letter from ACLU et al., to Senator and Chairman Patrick J. Leahy and Senator and Ranking Member Jeff Sessions (Oct. 26, 2010).

84. *See* Daskal et al., *supra* note 53, at 22.

85. *See id.* at 26.

86. *See id.*

III. *SCHREMS II* OVERVIEW

On July 16, 2020, the CJEU decided *Schrems II*. However, like all cases, that decision did not occur in a vacuum. The environment was one of increasing globalization and increasing privacy concerns.[87]

This Section begins by illustrating *Schrems II*'s landscape. This includes a discussion about the Snowden revelations and the GDPR. Then, this Section discusses the two Safeguard Agreements at issue in *Schrems II*: the EU–U.S. Privacy Shield, and the Standard Contractual Clauses (SCCs). Finally, this Section explains the *Schrems II* decision.

A. SNOWDEN, THE GDPR, AND THE DEVELOPMENT BEFORE *SCHREMS II*

Schrems II's journey to the CJEU began at least seven years prior, starting with an American: Edward Snowden. This Section begins by summarizing the Snowden revelations and Snowden's impact on privacy laws. Then, this Section further explores those privacy laws, specifically data localization laws and the GDPR.

In June 2013, Edward Snowden, a former U.S. National Security Agency (NSA) agent leaked classified materials that revealed NSA wiretapping and data collection activities.[88] The materials revealed how the United States covertly wiretapped prominent international leaders, including Brazil's former president Dilma Rouseff and Germany's Chancellor Angela Merkel.[89] Snowden's revelation incited uproar, with some calling him a "hero" and others calling him a "grandiose narcissist."[90] The revelations, more importantly, impacted international privacy laws.

Following Snowden's revelations, governments began reevaluating their privacy laws. One type of law governments began implementing was "data localization" laws.[91] Data localization is a process where governments "limit the storage, movement, and/or processing of digital data to specific

87. Chander & Lê, *supra* note 4, at 679.

88. DANIEL J. SOLOVE & PAUL M. SCHWARTZ, INFORMATION PRIVACY 455 (Rachel E. Barkow et al. eds., 6th ed. 2018).

89. *Id.*; Brian Winter, *Brazil's Rousseff Wants U.S. Apology for NSA Spying*, REUTERS (Sept. 4, 2013), https://www.reuters.com/article/us-usa-security-snowden-brazil/exclusive-brazils-rousseff-wants-u-s-apology-for-nsa-spying-idUSBRE98314N20130904.

90. SOLOVE & SCHWARTZ, *supra* note 88, at 477.

91. *See generally* Jonah Force Hill, *The Growth of Data Localization Post-Snowden: Analysis and Recommendations for U.S. Policy Makers and Business Leaders*, LAWFARE RES. PAP. SER. (2014) (analyzing the rise of data localization laws post-Snowden); *see also* Chander & Lê, *supra* note 4, at 694 ("In February 2014, Chancellor Angela Merkel proposed that Europe build out its own internet infrastructure designed to keep data within Europe.").

geographies, jurisdictions, and companies."[92] Although data localization has been around since the 1990s, primarily in more authoritarian states, since then more democratic countries have also introduced data localization laws.[93] Most notoriously, Russia began implementing a plan, requiring all of Russia's information to be stored on servers physically located in Russia.[94] This secludes any Russian information within its borders. However, following Snowden, more democratic countries also began adopting data localization laws, including Australia, Canada, and others.[95] Australia requires all health data to confined within Australia.[96] Canadian provinces require public institutions to store data user's personal information within Canada.[97] In sum, data localization has become a common practice.

In addition to data localization, governments reevaluated their general privacy laws, particularly the European Union. The European Union reevaluated their Proposed Data Protection Regulation and their Safe Harbor Agreement with the United States.[98] In 2018, the European Union passed the GDPR, the state's primary privacy legislation. The GDPR regulates how entities collect, transfer, process, and disclose data subjects'[99] personal data,[100] and it has become the standard privacy law for EU citizens.

In particular, the GDPR's Articles 44, 45, and 46 were pertinent to *Schrems II*. Article 44 sets the foundation for international data transfers.[101] If a company plans to transfer EU data internationally, they need to ensure EU data users have the same level of protection provided by the GDPR.[102]

Article 45 elaborates when a party can transfer data under an EU Commission Adequacy Decision, such as data transfer treaties. The European

92. Hill, *supra* note 91, at 1.

93. STEPHEN J. EZELL, ROBERT D. ATKINSON & MICHELLE A. WEIN, LOCALIZATION AS BARRIERS TO TRADE: THREAT TO THE GLOBAL INNOVATION ECONOMY 20 (2013).

94. Adee, *supra* note 58.

95. Hill, *supra* note 91, at 4 n.3.

96. Chander & Lê, *supra* note 4, at 683.

97. *See id.* at 685 ("British Columbia and Nova Scotia have enacted laws requiring that personal information held by public institutions--schools, universities, hospitals, government-owned utilities, and public agencies--be stored and accessed only in Canada unless one of a few limited exceptions applies.").

98. SOLOVE & SCHWARTZ, *supra* note 88.

99. "Data subjects" is defined as the person(s) whose data is collected and processed. GDPR, *supra* note 22, art. 4.

100. *Id.* arts. 2, 4.

101. *Id.* art. 44.

102. GDPR art. 44 specifies that "[a]ny transfer of personal data . . . to a third country . . . shall take place only if . . . conditions . . . are compiled with by controller . . . including for onward transfers." Further, Article 44 mandates that all provisions shall be applied the same level of protection to EU citizens so that "this Regulation is not undermined."

Union Commission, the EU's executive branch, can evaluate a third-party country's data protection standards. If the third-party's country "ensures an adequate level of protection," then the third party does not need a special authorization before they transfer data to an entity in that third-party country.[103]

Article 45(2) elaborates on elements the Commission should consider when determine a country's data protection standards. These elements include "respect for human rights and fundamental freedoms . . . national security and criminal law and access of the public authorities to personal data," and additional elements.[104] After assessing the country's level of protection, the Commission can implement an act ensuring "an adequate level of protection."[105] Implementation allows parties to transfer data to that country without needing specific authorization.

Article 46 provides other viable agreements that would enable transnational data transfers absent an Article 45 Adequacy Agreement.[106] A controller can transfer data internationally "only if the controller or processor has provided appropriate safeguards."[107] Some do not require authorization. These include "a legally binding and enforceable instrument between public authorities," "binding corporate rules," and others.[108] Some safeguards do require authorization before data is transferred. These include "contractual clauses."[109] If a company, in a country not recognized under Article 44, wants to transfer EU data internationally, an Article 46 agreement (hereinafter "Safeguard Agreement") is the next viable option.

The GDPR also contains recitals, providing additional context to the GDPR Articles. They recognize the balance between technology's need for data and the importance of protecting personal data.[110] That is, for instance, it

---

103. *Id.* art. 45(1).

104. *Id.* art. 45(2).

105. *Id.* art. 45(3).

106. "In the absence of a decision pursuant to Article 45(3), a controller or processor may transfer personal data to a third country or an international organisation only if the controller or processor has provided appropriate safeguards, and on condition that enforceable data subject rights and effective legal remedies for data subjects are available." *Id.* art. 46(1).

107. A "controller" is defined as any entity that "determines the purposes and means of the processing of personal data." A "processor" is defined as "a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller." *Id.* art. 4.

108. *Id.* art. 46(2).

109. *Id.* art. 46(3).

110. GDPR, *supra* note 22, recital 6 ("[S]haring of personal data has increased significantly . . . and should further facilitate the free flow of personal data within the Union

acknowledges the importance of international collaboration in data transfers but also guards against "undermin[ing]" the "level of protection of natural persons in the [European] Union."[111] Additionally, the recitals recommend what factors the Commission should take into account when assessing transnational data transfers.[112] If foreign government's privacy laws do not protect EU consumer data adequately, then foreign data controllers "should take measures to compensate for the lack of data protection in a third country by way of appropriate safeguards for the data subject."[113] Controllers should fill the gap in the absence of an EU Adequacy Decision. In addition to the Articles, the recitals illustrate the GDPR's expectations.

Following the Snowden revelations, governments recalibrated their privacy policies, whether that meant enacting data localization laws or adopting standardized privacy policies. Although the Snowden revelations occurred almost a decade ago, governments are still grappling with its implications, leading to *Schrems II*.

B. THE EU–U.S. PRIVACY SHIELD & SCCS, THE TWO SAFEGUARD PROVISIONS THE CJEU REVIEWED IN *SCHREMS II*

Under the GDPR, non-EU entities (also called third-country entities) can transfer EU data user information under an Article 45 Adequacy Decision or an Article 46 Safeguard Agreement. In *Schrems II*, the EU–U.S. Privacy Shield Adequacy Decision was the Article 45 Adequacy Decision at issue, and Standard Contractual Clauses were the Article 46 Safeguard Agreements at issue.[114] Before reviewing *Schrems II*, this Section provides background regarding both the EU–U.S. Privacy Shield and SCCs.

After collaborating with U.S. authority, the EU Commission adopted the EU–U.S. Privacy Shield to transfer data between the two unions. The Privacy Shield contained new protocols in addition to existing ones, but the EU Commission was particularly concerned with three of them.[115] First was the Ombudsperson Clause. Second was the U.S. Executive Branch's foreign surveillance authority. Finally, the EU Commission reviewed the Foreign

---

and the transfer to third countries and international organizations, while ensuring a high level of the protection of personal data.").

111. GDPR, *supra* note 22, recital 101.

112. GDPR, *supra* note 22, recital 104 specifies that when judging data transfers to foreign countries, the GDPR recommended that "the Commission should . . . take into account how a particular third country respects the rule of law . . . including legislation concerning public security."

113. GDPR, *supra* note 22, recital 108.

114. *Schrems II*, ¶ 1.

115. Case C-311/18, Data Protection Commissioner v. Facebook Ireland, Ltd., 2020 E.C.R., ¶ 42.

Intelligence Surveillance Act (FISA). These three protocols are addressed in turn.

To supply sufficient remedies to EU data subjects,[116] as mandated by the GDPR,[117] the United States created a new Ombudsperson Mechanism.[118] The Ombudsperson Mechanism created a Senior Coordinator (or "Ombudsperson") in the State Department. This Ombudsperson was foreign governments' point of contact to raise concerns about U.S. intelligence activities.[119] In response to these concerns, the Ombudsperson can rely on existing U.S. review mechanisms to remedy intelligence agency's non-compliance.[120] This system created a reactive response to non-compliance.

The second component the EU Commission assessed was the U.S. Executive's foreign surveillance authority. Under the Constitution, the Executive branch has authority over national security.[121] Congress can limit that authority, but the President can still govern intelligence agencies within those limitations. The EU Commission reviewed two key Presidential tools, Executive Order 12333 (E.O. 12333) and Presidential Policy Directive 28 (PPD-28). E.O. 12333, when enacted in 1981, established new surveillance authorities for U.S. intelligence agencies.[122] E.O. 12333 allows intelligence agencies to collect and retain EU data users' information that is travelling to the United States. Furthermore, E.O. 12333 is not governed by statute and does not give EU data users redress.[123]

The EU Commission also considered PPD-28, an executive order. PPD-28 imposed limitations on "signals intelligence operations," binding U.S. intelligence authorities.[124] The purpose of PPD-28 was to clarify what the

116. Under FISA, the U.S. provided some remedies to EU Data Subjects. This included bringing a civil cause of action against U.S. government officials and the Freedom of Information Act (FOIA).

117. *See* GDPR, *supra* note 22, art. 44 ("All provisions in this Chapter shall be applied in order to ensure that the level of protection of natural persons guaranteed by this Regulation is not undermined."); *Schrems II*, ¶ 92 (discussing how adequate remedies is implicitly required in Article 45 Adequacy Agreements, even though Article 45 does not explicitly require it).

118. Commission Implementing Decision (EU) 2016/1250 of 12 July 2016 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the EU-U.S. Privacy Shield, OJ L 207, 1.8.2016 [hereinafter EU–U.S. Privacy Shield].

119. EU–U.S. Privacy Shield recital 116.

120. EU–U.S. Privacy Shield recital 120.

121. EU–U.S. Privacy Shield recital 68.

122. Executive Order 12333, EPIC (accessed Oct. 26, 2020), https://epic.org/privacy/surveillance/12333/.

123. *See* EU–U.S. Privacy Shield recital 115.

124. EU–U.S. Privacy Shield recital 69.

United States does and does not do during foreign surveillance.[125] The EU Commission found that PPD-28 "capture[d] the essence of the principles of necessity and proportionality."[126] Together, E.O. 12333 and PPD-28 authorizes American intelligence agencies foreign surveillance power.

The third component the EU Commission was concerned with was FISA, an American law establishing foreign surveillance procedures. FISA was originally enacted in 1978 as a response to "abuses of U.S. persons' privacy rights" by the U.S. government.[127] The U.S. government claimed those abuses occurred during national security investigations.[128] Therefore, Congress passed FISA to establish foreign surveillance procedures, specifically to collect information about "agents of foreign powers."[129] Rather than authorizing individual surveillance activities on a case-by-case situation, FISA authorizes authorized surveillance programs (e.g., PRISM[130] and UPSTREAM[131]).[132]

In light of the Ombudsperson Clause, U.S. Executive's power (i.e., E.O. 12333 and PPD-28), and FISA, the EU Commission assessed the EU–U.S. Privacy Shield's safeguards and limitations.[133] Ultimately, the EU Commission adopted the EU–U.S. Privacy Shield, finding that the "United States ensures an adequate level of protection for personal data transferred from the Union" to the United States.[134] Therefore, any entity that accepts the Privacy Shield's provisions can transfer data, relying on an Article 45 Adequacy Agreement.

---

125. *President Obama's Remarks on NSA Program (Transcript)*, POLITICO (Jan. 17, 2014), https://www.politico.com/story/2014/01/barack-obama-nsa-speech-transcript-102315 ("[T]he new presidential directive that I have issued today will clearly prescribe what we do, and do not do, when it comes to our overseas surveillance.").

126. EU–U.S. Privacy Shield recital 76.

127. James G. McAdams III, *Foreign Intelligence Surveillance Act (FISA): An Overview*, FED. L. ENF'T TRAINING CTRS., https://www.fletc.gov/sites/default/files/imported_files/training/programs/legal-division/downloads-articles-and-faqs/research-by-subject/miscellaneous/ForeignIntelligenceSurveillanceAct.pdf.

128. *See id.*

129. *Id.*; *The Foreign Intelligence Surveillance Act - News and Resources*, ACLU, https://www.aclu.org/other/foreign-intelligence-surveillance-act-news-and-resources.

130. "PRISM is a tool used by the US National Security Agency (NSA) to collect private electronic data belonging to users of major internet services like Gmail, Facebook, Outlook, and others." T.C. Sotteck & Janus Kopfstein, *Everything You Need to Know About PRISM*, THE VERGE (July 17, 2013), https://www.theverge.com/2013/7/17/4517480/nsa-spying-prism-surveillance-cheat-sheet.

131. "Upstream Surveillance is a term used by the National Security Agency (NSA) of the United States for intercepting telephone and Internet data from the Internet backbone, i.e., major Internet cables and switches, both domestic and foreign." *Upstream Surveillance*, LDAP WIKI, (last updated Oct. 8, 2018), https://ldapwiki.com/wiki/Upstream%20Surveillance.

132. EU–U.S. Privacy Shield recital 109.

133. EU–U.S. Privacy Shield recital 65.

134. EU–U.S. Privacy Shield art. 1.

Absent an Article 45 Adequacy Agreement (e.g., the Privacy Shield), private entities can transfer data if protected by an Article 46 Safeguard Agreement, the second safeguard provision the GDPR authorizes.[135] One type of Safeguard Agreement is SCCs. SCCs are pre-authorized contracts that can be an EU entity and a non-EU entity can adopt, notably not directly including governments as parties to the SCCs. In the SCCs, the non-EU entity would agree to appropriate safeguards to ensure adequate protection and remedies for EU data users. Once the entities agreed to an SCC, the SCC is subject to review by the EU entity's supervisory authority, an independent public authority that is responsible for monitoring GDPR compliance.[136] Unless the supervisory authority finds that the SCC does not provide an adequate protection, the two entities can transfer the data transnationally, and this protocol applies to all Safeguard Agreements under Article 46.

Together, non-EU entities can legally transfer data under an Article 45 Adequacy Agreement or an Article 46 Safeguard Agreement (e.g., an SCC). If the non-EU entity is in a country with a valid Adequacy Agreement, then the entity only must subscribe to their respective Adequacy Agreement's provisions before transferring data. If the entity is not in a country with an Adequacy Agreement, then the entity can still transfer data with an Article 46 Safeguard Agreement. That is, the non-EU entity does not have to rely on their own government to provide adequate protections. A non-EU entity can provide adequate protections on their own. Together, Adequacy Agreements and Safeguard Agreements provide non-EU entities options to transfer data in compliance with the GDPR.

C. THE CJEU DECISION IN *SCHREMS II*

In *Schrems II*, the CJEU scrutinized the EU–U.S. Privacy Shield Adequacy Agreement and SCC Safeguard Agreements. Essentially, the CJEU addressed whether the EU–U.S. Privacy Shield provided enough protection to EU data users to be a valid Adequacy Agreement, and whether SCCs could ever provide an adequate level of protection mandated by the GDPR. The CJEU concluded negatively to the former but affirmatively to the latter.

135. "In the absence of a decision pursuant to Article 45(3), a controller or processor may transfer personal data to a third country or an international organization only if the controller or processor has provided appropriate safeguards, and on condition that enforceable data subject rights and effective legal remedies for data subjects are available." GDPR, *supra* note 22, art. 46.

136. *Id.* arts. 4, 51.

This Section begins by discussing why the CJEU invalidated the EU–U.S. Privacy Shield. Then this Section discusses why the CJEU affirmed SCCs' validity.

1. The CJEU Privacy Shield Invalidation

The first question the CJEU evaluated was whether the EU–U.S. Privacy Shield was a valid treaty under Article 45.[137] The CJEU invalidated the EU–U.S. Privacy Shield because of its unproportionate limitations and lack of remedies.[138]

The CJEU found the Privacy Shield's inadequacies invalid for two reasons. First, the U.S. foreign surveillance programs were not proportional to what is strictly necessary. The CJEU recognized that data user's privacy rights are not absolute, but considered in connection to personal data's function in society.[139] This includes data transfers necessary for important public interest reasons, like public security.[140] But, that limitation must be proportional to the interest and applied only as strictly necessary.[141] In the Privacy Shield, FISA authorized U.S. surveillance programs in general, not on a case-by-case basis.[142] That means there was no limitation on FISA's powers. Therefore, because FISA lacked proportionality, the Privacy Shield's failure to address those inadequacy rendered it invalid.

Second, the CJEU determined that the Privacy Shield did not provide EU data users adequate remedies. The GDPR requires that those whose rights have been violated shall have the right to an effective tribunal remedy.[143] When deciding an Adequacy Decision, the Commission must consider "effective administrative and judicial redress for the data subjects whose personal data are being transferred."[144]

Furthermore, although some data users can be susceptible to U.S. surveillance, they do not necessarily have legal redress[145] under E.O. 12333 and

137. The fourth, fifth, ninth, and tenth questions concerned whether the Privacy Shield Decision ensures an adequate level of protection and whether data transfers, pursuant to the SCC and Privacy Shield's ombudsperson clause, is compatible with the GDPR and the Charter. In essence, whether the Privacy Shield is valid under EU law.

138. *See* Case C-311/18, Data Protection Commissioner v. Facebook Ireland, Ltd., 2020 E.C.R., ¶ 201.

139. *Id.* ¶ 172.

140. *See* GDPR, *supra* note 22, art. 49.

141. *Schrems II*, ¶¶ 174–76.

142. *Id.* ¶ 179. For a discussion on FISA, see *supra* pp. 1588–1591.

143. GDPR, *supra* note 22, art. 78.

144. *Id.* art. 45.

145. *Schrems II*, ¶ 191.

PPD-28,[146] leaving a gap in EU data users' legal remedies. The Commission relied on the Privacy Shield's Ombudsperson to fill this gap. The Ombudsperson[147] allegedly ensured the required remedial rights.[148] Contrary, the CJEU found that the Ombudsperson did not have binding power over the U.S. intelligence agencies.[149] Furthermore, the Ombudsperson mechanism did not provide any legal safeguards that EU data users could rely on.[150] These two issues made the Ombudsperson mechanism an insufficient remedy.[151] The CJEU found that this gap in legal remedies, and the fact that the Ombudsperson did not fill this gap, the Privacy Shield did not ensure an adequate level of protection. Therefore, the Privacy Shield was invalid.

2. *The CJEU SCC Validation*

The second question the CJEU evaluated was whether SCCs were valid means to internationally transfer data. The CJEU affirmed their validity because SCCs, ostensibly, provide adequate safeguards.

Preliminarily, the CJEU defined what level of protection the GDPR requires to transfer data via an SCC. Addressing Article 46's required level of protection, the CJEU clarified that an adequate level is enough protection to be "essentially equivalent to that guaranteed within the European Union."[152] In order to transfer EU data, non-EU entities must provide an adequate level of protection. To determine whether a non-EU entity is providing an adequate level of protection, an EU member state's supervisory authority assesses both the contractual clauses and the non-EU entity's legal system.[153] Notably, the non-EU government does not have to provide the adequate level of protection; instead, a non-EU entity itself can "compensate for the lack of data protection in a third country."[154] Considering all the above, if a supervisory authority determines there is an adequate level of protection, then the non-EU entity can transfer EU data.

However, if the EU supervisory authority determines that an adequate level of protection is not ensured by the SCC, then the supervisory authority

146. *Id.* ¶ 182–83.

147. This is the "Senior Coordinator for International Technology Diplomacy," or the EU's point of contact.

148. *Schrems II*, ¶ 193.

149. *Id.* ¶ 196.

150. *Id.*

151. *Id.* ¶ 197.

152. *Id.* ¶ 94.

153. *Id.* ¶ 104. When assessing a third-country entity, the CJEU directed Supervisory Authorities to Article 45(2) which enumerates a non-exclusive set of factors to consider.

154. GDPR, *supra* note 22, recital 108.

must take remedial action to ensure EU data user's privacy rights.[155] This includes "impos[ing] a temporary or definitive limitation including a ban on processing . . . to order the suspension of data flows to a recipient in a third country."[156] In addition, supervisory authorities can suspend international data transfers if the supervisory authority determines that "the standard data protection clauses are not or cannot be complied with" and adequate protection cannot be ensured.[157] That is, EU supervisory authorities have some discretion as to which corrective power is necessary. But if a supervisory committee determines that the SCC does not provide an adequate level of protection, then the private parties must suspend data transfers.

In addition to the supervisory authorities, EU private parties also have responsibility to ensure an adequate level of protection before transferring data outside of the EU. Transfers with EU data inevitably involve an EU party before they transfer the data out of the EU. If the EU party cannot guarantee an adequate level of protection, they must suspend data transfers.[158] That is, EU parties have responsibility to ensure EU data-users protection before internationally transferring the data.

In conclusion, the CJEU determined the Privacy Shield was invalid because the U.S. access to EU data was not proportional to what was strictly necessary and failed to provide EU citizens adequate remedies. Nonetheless, the CJEU determined SCCs were still valid because SCCs have backstops to protect data users' privacy rights, so the CJEU determined that SCCs were still a valid means to transfer data.

IV. PARODOXICALLY, *SCHREMS II* PROMOTES INTERNET BALKANIZATION WHILE MAINTAINING AN OPEN INTERNET

By invalidating the EU–U.S. Privacy Shield and upholding SCCs, *Schrems II* requires non-EU entities to adopt the EU's stringent privacy standard (i.e., the GDPR) if they want to participate in the EU market. That leaves some non-EU entities (e.g., American entities, likely, and other close economic allies) to adopt the EU standard. However, the EU standard may fundamentally preclude some non-EU entities because of their own home country regulations.[159] This dichotomy fragments the internet, resulting in Internet

155. *Schrems II*, ¶ 111.
156. GDPR, *supra* note 22, art. 58(2).
157. *Schrems II*, ¶ 113.
158. *Id.* ¶ 135.
159. Joshua P. Meltzer, *The Court of Justice of the European Union in Schrems II: The Impact of GDPR on Data Flows and National Security*, BROOKINGS INST. (Aug. 5, 2020), https://

Balkanization. But, by refusing access to EU data from non-compliant entities, EU data users' information privacy rights are upheld; EU data users have more control over their data without fear that non-compliant parties will inappropriately use their data. Moreover, as a result of having more control over their information, EU data users have more freedom to choose how they present themselves on the internet, how they interact with the internet, and how others interact with them on the internet. That is, they have more autonomy—independence from adverse influence. Having internet autonomy is one of the foundational tenets of the open internet. Thus, in effect, *Schrems II* maintains open internet ideals by promoting Internet Balkanization.

This Part begins by discussing how *Schrems II* promotes Internet Balkanization, the global internet's fragmentation. Next, this Part discusses how the "digital border" the European Union erected in *Schrems II*, the border fragmenting the internet, protects EU data users' information privacy. From that, EU data users maintain digital autonomy, which the third Section of this Part discusses. Finally, this Part concludes by discussing how maintaining digital autonomy maintains one of the open internet's core tenets: the free flow of information.

A. *SCHREMS II* PROMOTES INTERNET BALKANIZATION

Following *Schrems II*, there are three legal remedies entities can adopt to continue transnational data transfers and be compliant with the GDPR. Entities residing in a country with an Article 45 Adequacy Decision can rely on the Adequacy Decision, entities can adopt an Article 46 Safeguard Agreement (e.g., SCC), or entities can localize their data in the European Union. However, limiting entities to these three options promotes Internet Balkanization.

This Section begins by analyzing each option independently. Then, it discusses how limiting non-EU entities to these three options promotes Internet Balkanization.

1. The three options non-EU entities have post-Schrems II

Following the CJEU's decision, there are three option non-EU entities can take to legally transfer EU data.

First, non-EU states can adopt a new Article 45 Adequacy Decision treaty, allowing economic entities to transfer data. For American entities, it is very possible that the European Union and United States will agree to a new EU–

www.brookings.edu/research/the-court-of-justice-of-the-european-union-in-schrems-ii-the-impact-of-gdpr-on-data-flows-and-national-security/; *see infra* pp. 45–48.

U.S. treaty, but this is the second EU–U.S. treaty the CJEU has invalidated.[160] Repeated invalidation insinuates that Article 45 Adequacy Decisions to be less reliable long term. Also, until the EU Commission accepts another Article 45 Adequacy Decision, companies must rely on other lawful means to transfer data.[161] Therefore, because of the lack of reliability and lack of time frame, this remedy is not a viable option in the meantime.

Second, companies could comply with other Safeguard Agreements under Article 46: a more stable choice. Prior to the *Schrems II* decision, some companies, like Microsoft, already adopted standard contractual clauses.[162] Also, "89% of non-government EU firms transfer personal data from the Union to the United States using the SCCs, making it the most relied upon mechanism for such transfers."[163] However, post *Schrems II*, the Commission will have to approve SCCs on a case-by-case basis.[164]

Third, companies can localize EU data users' information in the European Union, a strategy rising in recent years.[165] Data localization is when data controllers or processors move their facilities into the respective jurisdiction, here the European Union, to avoid transnational data transfers laws. This is a way for companies to avoid the transatlantic data transfer issue. However, data localization is not a perfect remedy. For instance, keeping all European data in Europe could be expensive because it impairs the economies of scale and reduces efficiency.[166] Another downside of data localization is the existence of technical problems with data localization.[167] For example, data may need to be

160. *See generally* Case C-362/14, Data Protection Commissioner v. Facebook Ireland, Ltd., 2015 E.C.R. (invalidating the EU-U.S. International Safe Harbor Privacy Principles).

161. The California Consumer Protection Act (CCPA) is the State of California's general statute for privacy rights. It can be argued that since Article 45 only applies to federal statutes, the CCPA fall under Article 45's Adequacy Decision. If so, this would allow California companies to transfer EU data user information without needing to develop their own safeguards.

162. Natasha Lomas, *Europe's Top Court Strikes down Flagship EU-US Data Transfer Mechanism*, TECH CRUNCH (July 16, 2020), https://techcrunch.com/2020/07/16/europes-top-court-strikes-down-flagship-eu-us-data-transfer-mechanism/?guccounter=1#:~:text=A%20highly%20anticipated%20ruling%20by,wrote%20in%20a%20press%20release [https://perma.cc/EYA7-Z39G].

163. Brief for BSA The Software Alliance as Amici Curiae Supporting Respondents Case C-311/18, Data Protection Commissioner v. Facebook Ireland, Ltd., 2020 E.C.R.

164. *Shcrems II*, ¶ 134.

165. Chander & Lê, *supra* note 4 (reviewing recent data localization laws of Australia, Brazil, Canada, China, European Union, France, Germany, India, Indonesia, Malaysia, Nigeria, Russia, South Korea, Vietnam, and others).

166. Reisman, *supra* note 49; Cody Ankeny, *The Costs of Data Localization*, ITI (Aug. 16, 2017), https://www.itic.org/news-events/techwonk-blog/the-costs-of-data-localization.

167. Reisman, *supra* note 49.

accessed by engineers in other countries for maintenance.[168] Furthermore, data localization restricts what companies can do with data. Restricting the flow of information "raises costs for local businesses, reduces access to global services for consumers, hampers local start-ups, and interferes with the use of the latest technological advances."[169] Because of these issues, data localization is not an ideal plan, but it is one that would prevent transnational data transfer issues.

2. These options require entities to adopt an EU-centric internet standard.

Limiting non-EU entities to these options leaves non-EU entities two options to transfer EU data. Non-EU entities can adopt an EU-centric privacy standard, by adopting an Adequacy or Safeguard Agreement, and be included in the EU-centric internet sphere. Or non-EU entities can reject these remedies and exclude themselves from the EU-centric internet sphere. Choosing either agreement results in the entity adopting an EU-centric internet standard and making themselves more EU-centric.

Non-EU countries adopting an Article 45 Adequacy Decision must ensure "a level of protection of fundamental rights essentially equivalent to that guaranteed in the EU legal order."[170] This includes both private entities and the non-EU state. In other words, if a foreign entity wants to have an Adequacy Decision, they will have to adopt and comply with GDPR equivalent standards. By adopting GDPR equivalent standards, non-EU entities adopt an EU-centric internet standard. A handful of countries have an EU Adequacy Decision.[171] These countries can openly flow data between themselves and the European Union, allowing them to be included in the EU-centric internet sphere.

Similarly, if an entity adopts a Safeguard Agreement, that entity is inevitably adopting an EU-centric internet standard. If a non-EU entity resides in a country that does not have a valid Article 45 Adequacy Agreement, then that entity can adopt an Article 46 Safeguard Agreement.[172] When adopting a

168. *Id.*

169. Chander & Lê, *supra* note 4, at 721.

170. *Schrems II*, ¶ 162.

171. *Adequacy Decisions*, EUROPEAN COMM'N, https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en#:~:text=The%20European%20Commission%20has%20so,are%20ongoing%20with%20South%20Korea [https://perma.cc/JM5D-XWZK], (accessed November 3, 2020) (listing Andorra, Argentina, Canada (commercial organizations), Faroe Islands, Guernsey, Israel, Isle of Man, Japan, Jersey, New Zealand, Switzerland, and Uruguay all have Adequacy Decisions).

172. GDPR, *supra* note 22, art. 46 ("In the absence of a decision pursuant to Article 45(3), a controller or processor may transfer personal data to a third country or an international organisation only if the controller or processor has provided appropriate safeguards.").

Safeguard Agreement, entities are required to ensure a level of protection "essentially equivalent" to that guaranteed in the European Union.[173] However, this is not an "identical level of protection," as the CJEU made clear, but it is difficult to distinguish the difference between "essentially equivalent" and "identical."[174] Furthermore, by adopting a Safeguard Agreement, entities ensure protection whether the data will be processed immediately or in the future.[175] This requires entities to ensure an equivalent level of protection perpetually. By adopting an Article 46 Safeguard Agreement, entities adopt the EU-centric internet standard, similar to when third countries adopt an Article 45 Adequacy Agreement.

Entities without Article 45 and Article 46 Agreements can still participate in the EU market if they localize EU user data in the European Union, completely avoiding international data transfers. Under this option, EU data user's information would be collected, processed, and stored in the European Union. Localization creates a confined EU internet sphere.[176] As discussed earlier,[177] data localization is not an ideal solution for several reasons,[178] but it is an option available.

In sum, whether an entity adopts an Article 45, Article 46, or a data localization remedy, the entity is adopting an EU-centric internet standard.

3. The EU-centric internet border divides entities that can adopt an EU internet standard and those that cannot.

Requiring non-EU entities to adopt an EU internet standard creates an internet border around the EU. If a non-EU entity wants to participate, they must adopt one of the three options above. The option is their "passport" to transfer data between different states freely. If an entity fails to adopt one of the three options, they do not have a passport, and they are unable to transfer data. Therefore, there is a dichotomy between entities that adopt the EU

173. *Schrems II*, ¶ 94.

174. Also, supervisory authorities can suspend data transfers, which would disrupt business. This incentivizes entities to ensure compliance. *See Schrems II*, ¶ 113.

175. *Schrems II*, ¶ 89.

176. However, GDPR art. 49 includes exceptions that allow transfer of data if the transfer is necessary for important reasons of public interest," if "the transfer is necessary for the establishment, exercise, or defense of legal claims," "the transfer is necessary for the conclusion or performance of a contract concluded in the interest of the data subject," and other reasons. But those instances are rare. Council of Bars and Law Societies of Europe, *CCBE Assessment of the U.S. CLOUD Act* 7 (Feb. 28, 2019).

177. *See supra* pp. 40–41.

178. *See* Meltzer, *supra* note 159 ("Yet for small companies, the impacts are most pronounced. For many, setting up in the EU is not an option.").

internet standard and those that don't. There are some organizations that fundamentally cannot adopt an EU model.

Beginning with Article 45 Adequacy Agreements, there are only a handful of non-EU states that have valid Adequacy Agreements with the EU.[179] The states that have adopted an Adequacy Agreement are culturally and politically similar to the European Union. The fact that the United States did not have a valid Adequacy Agreement, a country with strong ties and political consistencies with the European Union, highlights how unfeasible an Article 45 Adequacy Agreement can be. There are some states that are very unlikely to adopt an Adequacy Agreement,[180] and not many countries may want to adopt an Adequacy Agreement. Entities outside of these enumerated states have to have an Article 46 Safeguard Agreement or localize their data.

Even though entities can adopt an Article 46 Safeguard Agreement to authorize data transfers, this is not a simple remedy. First, some home state policies fundamentally prevent entities from adopting an adequate Article 46 Safeguard Agreement, especially in more authoritarian states with different privacy rights.[181] Some states *want* private entities to collect information; that way, the state can use the collected behavioral data for their own use.[182] If an EU Supervisory Authority does not find an adequate level of protect, data transfers would be suspended.[183]

Furthermore, companies may be dissuaded from adopting an Article 46 Safeguard Agreement regardless of the home country's policies.[184] This is particularly true for smaller companies. Safeguard Agreements create difficulties for smaller companies by imposing additional costs which smaller companies may feel are not possible or worthwhile to bear.[185] Therefore,

179. Andorra, Argentina, Canada (commercial organizations), Faroe Islands, Guernsey, Israel, Isle of Man, Japan, Jersey, New Zealand, Switzerland, and Uruguay all have Adequacy Decisions.

180. Meltzer, *supra* note 159 ("If the U.S. is still not adequate, then it must be the case that other countries, including China will never be adequate and not only that, but it is hard to see how any Chinese company collecting EU personal data can transfer it back to China consistently with GDPR.").

181. *Id.*

182. Mariano-Florentino Cuéllar & Aziz Z. Huq, *Privacy's Political Economy And The State Of Machine Learning*, 76 NYU ANNUAL SURVEY OF AMERICAN LAW 317, 342 (2021).

183. *Schrems II*, ¶ 113.

184. *See Facebook Will Move UK Users to US Terms, Avoiding EU Privacy Laws*, THE GUARDIAN (Dec. 16, 2020), https://amp.theguardian.com/technology/2020/dec/15/facebook-move-uk-users-california-eu-privacy-laws (discussing how Facebook will move UK data users to U.S. terms because the EU privacy regime is the strictest privacy regime).

185. *See* Meltzer, *supra* note 159 ("The difficulties with SCCs also create additional costs and disincentives for EU companies to develop digital supply chains with SMEs in third countries.").

smaller companies' limited resources may hinder them from adopting the EU-centric internet sphere.

Finally, there is the data localization option. This is the option of last resort for entities that cannot rely on an Article 45 Adequacy Agreement nor comply with an Article 46 Safeguard Agreement. Those entities are left with only localizing their data in the European Union. That data would remain exclusively in the European Union, exclusively in the EU-centric internet sphere.

In order to use EU data at all, non-EU entities must adopt EU privacy standards. However, some non-EU entities are prohibited—fundamentally or functionally—from adopting EU privacy standards. The barriers to entry are too inaccessible. Therefore, a robust legal border confines the EU-centric internet sphere, fragmenting the open internet.

B. FROM THE REGULATIONS THAT WILL FRAGMENT THE INTERNET, EU DATA USERS' INFORMATION PRIVACY RIGHTS REMAIN INTACT

Information privacy is the control a data user has over their information,[186] the "fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data."[187] In an increasing digital, globally connected world, people relinquish greater volumes of data and more private information.[188] Collecting information has turned into an industrial complex.[189] Increasingly, more and more entities want access to data user's information.[190]

However, in *Schrems II* the CJEU's objective was to uphold EU data users' information privacy rights. It succeeded; from the regulations that will inevitably fragment the internet come procedures that will uphold information privacy. *Schrems II* prioritized data users' privacy rights in two ways. First, *Schrems II* limited non-EU state actors' access to EU data users' information by holding states accountable for using private entities data. Second, *Schrems II* limited private third parties' access to data users' information. By separating both non-EU states and private third parties from EU data users, the fragment, *Schrems II* bestowed enforceable privacy right to EU citizens.

---

186. Manheim & Kaplan, *supra* note 16, at 118.

187. GDPR, *supra* note 22, art. 1.

188. *See supra* pp. 5–9; *see also* GDPR, *supra* note 22, recital 6 (illustrating data collection's increasing prominence in society and discussing how "[t]echnology allows both private companies and public authorities to make use of personal data on an unprecedented scale in order to pursue their activities. Natural persons increasingly make personal information available publicly and globally.").

189. THE ECONOMIST, *supra* note 12.

190. *See id.*

1. *Schrems II limits state actors' access to data users' information, prioritizing data users' privacy rights over non-EU states' interests.*

Schrems II recognized that, even when states use private party's data for security reasons, states remain susceptible to the GDPR. However, the CJEU also recognized that private entities themselves can adopt adequate privacy standards regardless of their home state's standards; just because a state does not provide adequate safeguards does not necessarily mean that the private entities themselves cannot provide the required safeguards.[191] Therefore, the CJEU recognized a separation between private entities and their states. But, whether private entities are capable of providing adequate protection depends on whether their home state's policies infringe on the private entities' safeguard measures.[192] If the state infringes on an EU data users' privacy rights (e.g., conducting general surveillance), then the private entity risks having all data transfers suspended.[193] Therefore, states do not get to take advantage of their domiciled companies with EU data, even when they are processing the data for national security concerns.[194] The states' actions have consequences.

Using the EU–U.S. Privacy Shield as an example, the CJEU ensured data users' privacy rights when another party transfers their data out of the European Union via an Adequacy Agreement. The CJEU did that in two ways.

First, when states process data for national security, processing must be proportional to what is strictly necessary.[195] In other words, supervisory authorities do not enjoy carte-blanche access to data and get to claim "security" as its valid reason. Otherwise, this would lead to an indefinite loss of control over one's data.[196] For example, by invalidating the Privacy Shield because of this carte-blanche surveillance power, the CJEU set precedent for data users' privacy in all Adequacy Agreements. Data users can use the web without indefinite surveillance threats. Of course, the supervisory authority can monitor data users when necessary, but that is only in narrow situations, prompted by justified reasons.

Second, the CJEU required Article 45 Adequacy Agreement states to arrange data users with more remedial agency. Providing adequate remedies is

---

191. *Schrems II*, ¶ 136 ("[T]he mere fact that standard data protection clauses . . . do not bind the authorities . . . cannot affect the validity of that decision.").

192. *See id.* ¶ 129.

193. This is another reason why some private entities may not be able to ensure an adequate level of protection. Governments take advantage of tech companies data collection.

194. *See Schrems II*, ¶ 88.

195. *Id.* ¶ 176.

196. Although the supervisory authority may not constantly monitor data users' activities, the threat alone may adversely impact data users' experience.

essential to information privacy. Otherwise, privacy would be toothless. Adverse third parties would be able to interfere with data users without repercussions. Therefore, to practically have privacy rights, adequate remedies are essential. *Schrems II* reinforces the right to adequate remedies.[197] The CJEU recognized how some American surveillance programs do not provide legal redress,[198] which the U.S. government also recognized,[199] and the Ombudsperson did not remedy those gaps.[200] Therefore, the Privacy Shield's lack of adequate remedies left EU data users with toothless privacy rights, privacy rights EU data users could not enforce.

2. Schrems II limits private third parties' access to data users' information.

By recognizing that private entities are not confined by their home state's regulations, the CJEU tasked private entities with the responsibility to ensure an adequate level of protection. This specifically concerns Article 46 Safeguard Agreements. The CJEU recognized that Article 46 Safeguard Agreements are based on private entities, both EU and non-EU private entities, to ensure adequate levels of protection.[201] On a case-by-case basis, the EU entity decides whether Safeguard Agreements provide an adequate level of protection.[202] Also, if the non-EU recipient cannot comply with their Safeguard Agreement, then they are required to suspend data transfers.[203] Therefore, the CJEU delegated responsibility to private entities.[204]

At first impression, giving power to private entities may not seem to give data users more privacy rights, but the way *Schrems II* gives private entities more responsibility indirectly ensures more data user privacy rights. *Schrems II* mandated that economic entities must ensure an adequate level of protection. If not, they risk suspending data transfers. This incentivizes economic entities to ensure adequate protection. Otherwise, data transfer suspensions could drastically harm the company because a suspension of data transfers would disrupt their business operations.

In addition to the original party to whom an EU data user relinquishes their data, the GDPR and EU Charter also require third parties (e.g., economic partners or companies that buy their data) to ensure an adequate level of

197. *Id.* ¶ 197; GDPR, *supra* note 22, art. 45.
198. *Schrems II*, ¶ 191.
199. *Id.* ¶ 181.
200. *Id.* ¶ 196.
201. *Id.* ¶ 134.
202. *Id.*
203. *Id.* ¶ 143 (additionally requiring entities to return or destroy any data that has already been transferred).
204. *Id.* ¶ 137.

protection.[205] If an economic entity transfers user data to a third party, and that party does not provide an adequate level of protection, then the original party risks having their data transfers suspended.[206] In other words, parties are accountable for where they transfer data to. Therefore, EU privacy standards transfer with EU data.

In sum, the CJEU requires third-party entities to comply with the GDPR before the third-party transfers the EU data user's information.

C. FROM INFORMATION PRIVACY, EU DATA USERS WILL HAVE MORE DIGITAL AUTONOMY

When data users have more control over their data, they have more autonomy. As discussed above, *Schrems II* protected data user's control over their information. Now, EU data user's rights transfer along with their data,[207] extending protection past the original recipient. Since data users still have rights when a third party has possession over that data, data users have more control. When data users have more control, they have more influence on who impacts their digital experience. If a data user wants to share their location with friends, an action that may potentially impact where someone goes on a given day, they can. If a data user wants to have curated advertisements, they can. But they are the ones who decide, not an adverse party. In other words, data users control who has what information, information that may influence a data users' actions depending on who has it.

Schrems II embodies digital autonomy. An EU data user may consensually disclose information to a foreign entity. Take, for instance, Facebook U.S. as an extension of Facebook Ireland. An EU citizen discloses information to Facebook U.S. via the Facebook app. However, that data user may not want to disclose that information to the U.S. government via FISA.[208] *Schrems II* prevents that. Even if the foreign entity (Facebook U.S.) shares that information to a third party (e.g., a business partner), that party also has to ensure GDPR compliance.[209] And part of the GDPR grants data users the right

205. *See id.* ¶ 137 (requiring standard data protection clauses to ensure effective measures to provide an adequate level of protection, this can be implicitly read to ensure an adequate level of protection for wherever the data goes).

206. *See id.*

207. *See* Case C-311/18, Data Protection Commissioner v. Facebook Ireland, Ltd., 2020 E.C.R., ¶ 137 (requiring standard data protection clauses to ensure effective measures to provide an adequate level of protection, this can be implicitly read to ensure an adequate level of protection for wherever the data goes).

208. *See supra* p. 34 (discussing how FISA was too over encompassing, allowing U.S. intelligence agencies general surveillance rights over EU citizens).

209. *See Schrems II*, ¶ 137.

to know where their personal information has been transferred,[210] what safeguards that foreign entity has implemented,[211] and the right to "erasure."[212] All these rights aggregate to EU data users having independent control over their information. From this control, data users have more autonomy.

By having digital autonomy, EU citizens can freely express themselves as their individual selves, free from adverse users, manipulation, or political influence. Using the Cambridge Analytica scandal as an example, data users would have had the ability to track where Facebook shared data users' information and how they were using that data. Ideally, this would have preempted Cambridge Analytica from manipulating political opinions by sending targeted propaganda. Therefore, data users' political intentions would have remained unadulterated. They would have remained individualistic.

D. HAVING DIGITAL AUTONOMY MAINTAINS ONE OF THE OPEN INTERNET'S CORE TENETS

By providing more individual autonomy, *Schrems II* promotes the free flow of information an open internet sought to achieve. There are two main reasons why greater accessibility is desirable: the open internet enables information to flow freely, and the open internet stimulates economic advancement.[213]

1. Digital autonomy facilitates the free flow of information.

The open internet has social benefits, arising from the free flow of information, ideas, knowledge, and viewpoints.[214] The free flow of information is beneficial to society because it prevents suppression, and digital autonomy upholds those interest. At the end of this Section, there is a discussion about how the free flow of information could inadvertently harm data users by allowing too much transparency, but this Section resolves those concerns.

As explained above, the open internet is beneficial because it cultivates open communication.[215] It is a vehicle to share information with people,

---

210. GDPR, *supra* note 22, art. 15(1). ("The data subject shall have the right to obtain from the controller confirmation as to whether . . . [their information is] processed, and, . . . access to the personal data and the following information: (c) the recipients . . . to whom the personal data have been or will be disclosed, in particular recipients in third countries or international organisations.").

211. *Id.* art. 15(2) ("Where personal data are transferred to a third country or to an international organisation, the data subject shall have the right to be informed of the appropriate safeguards pursuant to Article 46 relating to the transfer.").

212. *Id.* art. 17.

213. Daskal et al., *supra* note 53, at 26.

214. ECONOMIC AND SOCIAL BENEFITS OF INTERNET OPENNESS, *supra* note 50, at 22–23.

215. *See supra* p. 13.

regardless of distance. Two people, physically present at two different parts of the world, could still connect with each other. From that connection, they can share ideas or information with each other. Or this connection could involve an individual reading the daily news in a different continent. In general, the internet provides open, uncomplicated mediums to communicate globally.

Digital autonomy protects the ability for data users to have this open communication by preventing information suppression. As discussed earlier, digital autonomy provides data users more control over their data.[216] Data users, at least EU data users, will be able to choose whether to communicate and whom to communicate with. At a minimum, EU journalists would be able to communicate what is happening across the world, especially when more traditional methods of communication, like print media, are easier to restrict and censor.[217] In theory, they would be able to disseminate information via social media, news websites, or other mass media vectors.[218] Therefore, suppressive laws would silence their information. A government would not be able to suppress EU data user's right to disclose. EU data users can still publish information about local political movements, human rights crises, or general cultural experiences.

At the same time, free flow of information could have adverse consequences. An entity may have more access to data users' information than intended. As more information becomes accessible, and sophisticated technology can find patterns in seemingly insignificant data, data users could disclose more information than they intended. Furthermore, data users may vary on what kind and how they disclose information.[219] That is, cultural differences may impact what information a data user discloses and who has access to that information.

---

216. *See supra* pp. 56–57.

217. MIKAL HEM, REUTERS INST. FOR THE STUDY OF JOURNALISM, EVADING THE CENSORS: CRITICAL JOURNALISM IN AUTHORITARIAN STATES 20 (2014), https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Evading_the_Censors_Critical_journalism_in_authoritarian_states_0.pdf ("Some journalists use social media and other internet outlets to publish material that is not possible to publish in traditional media, and online newspapers are generally confined by less strict censorship laws. But increasingly, governments in authoritarian countries try to restrict online content.").

218. Of course, if the internet sphere is fragmented due to mechanical barriers, then the internet would be preliminarily inaccessible. However, there are few places with such strict internet barriers (e.g., North Korea).

219. *See generally* James Q. Whitman, *The Two Western Cultures of Privacy: Dignity Versus Liberty*, 113 YALE L.J. 1151 (2004) (exploring how Europeans and Americans have two different cultures of privacy, which exemplifies how different cultures have different definitions of "privacy").

States upholding EU users' autonomy will respect EU data users' different cultural values. Data users may value different information more than others, thus making data users more inclined to disclose certain information. For example, Americans privacy stems from the belief that Americans should be free from intrusions of the state.[220] But Europeans privacy stems from the belief that people have the right to dignity and honor.[221] The different implicit beliefs may result in data users being more willing to disclose certain information than other information. Americans may be more willing to disclose information to companies (i.e., not the state) while Europeans may be more willing to disclose innocuous information to the state.[222]

As a more practical example, data users may vary how much they disclose intimate details depending on which state they are in. For example, queer data users may readily disclose their sexual identity to an American entity to find community. But in countries where it is illegal to be queer,[223] queer data users would be more reluctant to disclose their sexual orientation. In a completely open internet, queer data users may not have that liberty to differentiate which entities have that information. But with digital autonomy, data users get to choose which foreign entities have that information and which do not. Having digital autonomy enables the open internet to respect different data users' cultural differences while still maintaining an open internet.

An increasing digital environment only exacerbates the cultural difference concerns. As discussed earlier, as more information becomes accessible, data processers can discover more intimate details.[224] Data users' digital presence discloses intimate details that they may not want to readily disclose. Therefore, it is pertinent that data users have autonomy in an increasing digital and global environment.

---

220. *Id.* at 1161.

221. *Id.*

222. *Id.*

223. OUTLIFE, https://www.outlife.org.uk/which-countries-criminalise-homosexuality?gclid=CjwKCAiAq8f- BRBtEiwAGr3Dgfod5qF7Y3j0e2sm14QUEj7T2TkFY6bBisoaZmJDZKMTz4xh9xIIuxoCWj0QAvD_BwE [https://perma.cc/8APQ-LN97].

224. Cuéllar & Huq, *supra* note 182, at 328–30 (discussing how AI can also expand privacy intrusions based on making inferences, predicting behavior, drawing on information once unanalyzable (e.g., genetics)).

2. *Arguably, digital autonomy will facilitate an open market.*

An open internet favors economic advancements for two main reasons.[225] First, the open internet, as an open market, facilitates global trade efficiently.[226] Second, the open market facilitates innovation. Internet Balkanization arguably counteracts them both.[227]

But this Note is not focused on remedying Internet Balkanization's economic maladies. In the first place, how Internet Balkanization harms the economy is debatable.[228] Also, empirical research would better address this issue. However, this Note briefly discusses whether digital autonomy will be better or not for the global economy.

One of the major criticisms against Internet Balkanization is that it will ossify large companies.[229] The heightened barriers to entry will prohibit emerging companies from entering various internet spheres. Therefore, since large companies are the only ones capable of handing the various barriers to entry, they will be the only companies with a global reach. However, if states adopt digital autonomy in general, then data users should have more market agency.[230] They would be able to actively choose which companies to solicit. This would create a more pervious market.[231] On the contrary, even if data users have autonomy, the heightened barriers of entry could prevent companies from connecting to data users in the first place, making data users susceptible to market forces. Therefore, without facilitated connection, digital autonomy could be an ill-fated remedy for market access.

In sum, Internet Balkanization has some economic implications. However, how and to what extent they impact the global economy is out of this Note's scope, but future scholars should conduct empirical research on how Internet Balkanization impacts the global economy.

---

225. ECONOMIC AND SOCIAL BENEFITS OF INTERNET OPENNESS, *supra* note 50, at 18–19.

226. *See* U.S. DEP'T OF COM., U.S. SECRETARY OF COMMERCE WILBUR ROSS STATEMENT ON SCHREMS II RULING AND THE IMPORTANCE OF EU–U.S. DATA FLOWS (July 16, 2020). U.S. Secretary of Commerce, Wilbur Ross, expressed his dissatisfaction with the decision because of the vast number of American and European companies that relied on the EU–U.S. Privacy Shield.

227. *See* Daskal et al., *supra* note 53, at 25–26.

228. *Id.*

229. *Id.*

230. *See* Cohen, *supra* note 19, at 1426 (discussing how autonomy results in people having more independent, diverse choice which can occur in the market).

231. ECONOMIC AND SOCIAL BENEFITS OF INTERNET OPENNESS, *supra* note 50, at 35–36 (discussing how the open internet increases trade).

V. CONCLUSION

"Governments of the Industrial World . . . I come from Cyberspace, the new home of Mind. . . . You are not welcome among us. You have no sovereignty where we gather."[232] Here, early internet pioneers declared cyber freedom.

When Maximillian Schrems brought his case against Facebook to the European Court of Justice, his goal was to protect EU data users' information privacy. He succeeded. But to do so, the CJEU adopted principles that would fragment the EU internet away from the global, open internet. Although Internet Balkanization would normally suppress open communication, paradoxically that may not be the case here. Instead, there may be more communication because data users have more autonomy. So, in the end, the governments of the Industrial World still are not welcomed.

232. Barlow, *supra* note 1.

ESSENTIAL FACILITIES, ESSENTIAL PATENTS, AND THE ESSENTIAL OVERSIGHT OF *QUALCOMM*

Emma Hagemann†

I. INTRODUCTION

In the United States today, almost everyone owns a cellphone.1 As the cellular technology behind these devices has improved and smartphones have come to dominate the cellular market, average users rely on their phones for more and more activities—from calling and texting to translating between languages and navigating city streets in real time. But despite this ubiquitous reliance, most users have very little technical understanding of how their cellphones actually do what they do.

One important piece of cellphone functionality is the cellphone's ability to communicate with other devices. Whether it is connecting a call to another telephone or transferring high-speed data to a large server, the phone is rendered useless if it cannot perform this communication function. Interoperability between different networks and phone manufacturers is paramount.

With so many different firms operating in the cellular communications industry (from network owners like AT&T to phone makers like Apple), interoperability has been facilitated by the development of communications standards.[2] To date, there have been four "generations" of cellular communications standards: 1G, 2G, 3G, and 4G.[3] 1G standards supported

---

DOI: https://doi.org/10.15779/Z38000016

† J.D. 2022, University of California, Berkeley, School of Law. I would especially like to thank Professor Talha Syed and Student Advisor Carmen Sobczak for their invaluable suggestions and guidance throughout the research, writing, and editing process. My sincere thanks also to my fellow students in the 2020 Law & Technology Writing Workshop at Berkeley Law and the Berkeley Technology Law Journal editors.

1. In 2019, 96% of Americans owned a cell phone. 81% of these were smartphones. *Mobile Fact Sheet*, PEW RSCH. CTR. (Apr. 7, 2021), https://www.pewresearch.org/internet/fact-sheet/mobile [https://perma.cc/55PL-6HGA].

2. *See* Mark A. Lemley, *Intellectual Property Rights and Standard-Setting Organizations*, 90 CALIF. L. REV. 1889, 1896 (2002) (discussing consumer benefits of standards in many markets); A. Douglas Melamed & Carl Shapiro, *How Antitrust Law Can Make FRAND Commitments More Effective*, 127 YALE L.J. 2110, 2111 (2018) ("[M]any of the benefits generated by the recent and dramatic advances in information technology would have been difficult or impossible to achieve without compatibility standards.").

3. Complaint at 6, Fed. Trade Comm'n v. Qualcomm, Inc., 411 F. Supp. 3d 658 (N.D. Cal. 2019) (No. 5:17-cv-00220). In addition, 5G is right around the corner. Sascha Segan, *What*

only the analog transmission of calls; the currently dominant 4G standard supports the high-speed data transfers so important to smartphones today.[4] Phones must comply with one or more standards to communicate with other devices.

Qualcomm, a multinational technology company, is a major player in two parts of the cellular technology market. First, Qualcomm holds thousands of patents that cover all kinds of important cellular technology.[5] It earns money from these patents by entering into licensing agreements and collecting royalties from parties that want to use the patented technologies. Some of Qualcomm's patents are standard-essential patents (SEPs). This label means that the technology covered by these patents has been deemed an essential component of a cellular communications standard.[6] Qualcomm has SEPs that cover technologies in 2G, 3G, and 4G standards.[7]

Second, Qualcomm is one of the largest manufacturers of the baseband processors ("chips") that go into smartphones.[8] Like smartphones, chips must comply with one or more standards. "While the chips themselves do not enable all of the functionality specified by the standard (e.g., one does not actually speak into a modem chip to make a phone call), these highly complex chips *do* embody the principal technical features of the standard."[9] Thus, Qualcomm's chips contain the 2G, 3G, or 4G technology covered by its SEPs, depending on the type of chip. In that sense, the chips are downstream from Qualcomm's patents. Though there are competing chip manufacturers, Qualcomm has maintained clear dominance in multiple chip markets for over a decade.[10] Qualcomm sells its chips to original equipment manufacturers (OEMs), like Apple, who make smartphones or other smart devices.[11]

Is 5G?, PCMAG (Nov. 24, 2020), https://www.pcmag.com/news/what-is-5g [https://perma.cc/7ZLP-6WN7].

4. Complaint, *supra* note 3, at 6.

5. *Qualcomm Welcomes the Court of Appeals' Complete, Unanimous Reversal of the District Court's Judgment in the FTC Case*, QUALCOMM, https://www.qualcomm.com/ftc [https://perma.cc/9DDD-C46N] (noting that Qualcomm holds 140,000 patents).

6. Fed. Trade Comm'n v. Qualcomm, 969 F.3d 974, 983 (9th Cir. 2020).

7. Complaint, *supra* note 3, at 13–14.

8. *Id.* at 9.

9. Brief of *Amicus Curiae* Professor Jorge L. Contreras in Support of Appellee and Affirmance at 13, Fed. Trade Comm'n v. Qualcomm, 969 F.3d 974 (9th Cir. 2020) (No. 19-16122).

10. *See* Fed. Trade Comm'n v. Qualcomm, Inc., 411 F. Supp. 3d 658, 690, 695 (N.D. Cal. 2019) (concluding that Qualcomm has monopoly power in Code Division Multiple Access (CDMA) and Long-Term Evolution (LTE) modem chips).

11. Qualcomm, 969 F.3d at 985.

It is against this backdrop that the Federal Trade Commission (FTC) filed its antitrust case against Qualcomm. Described as "one of the most important antitrust cases of the twenty-first century," the case was closely watched by the media, industry players, and economists.[12] Throughout the litigation, these actors, plus former government officials, the Antitrust Division of the Department of Justice, and law professors, weighed in as *amici* on both sides.[13]

This Note discusses two of Qualcomm's business practices that the FTC alleged were in violation of antitrust law.[14] First and most relevant was Qualcomm's practice of not offering patent licensing agreements to its rival chipmakers. Second was Qualcomm's "no license, no chips" (NLNC) policy, under which Qualcomm refused to sell chips to an OEM unless that OEM agreed to enter into a patent licensing agreement and pay royalties to Qualcomm based on the price of the final product, the smartphone.

The District Court for the Northern District of California agreed with the FTC that both practices harmed competition and violated antitrust law.[15] In May 2019, the court enjoined Qualcomm from engaging in these practices.[16] On August 11, 2020, the Ninth Circuit summarily reversed the district court.[17] It held that none of Qualcomm's practices violated antitrust law, and it criticized much of the district court's analysis as contrary to accepted legal standards.[18] The two decisions are polarizing, to say the least. Some commentators applaud the Ninth Circuit for correctly reining in an

12. *See, e.g.*, Brief of 46 Amici Curiae Law and Economics Scholars in Support of Petition for Rehearing En Banc at 2, Fed. Trade Comm'n v. Qualcomm, Inc. 969 F.3d 974 (9th Cir. 2020) (No. 19-16122).

13. *See* Docket No. 19-16122, Fed. Trade Comm'n v. Qualcomm, Inc. 969 F.3d 974 (9th Cir. 2020) (listing over twenty amicus briefs submitted to the Ninth Circuit before its ruling).

14. The FTC also challenged Qualcomm's exclusive dealing agreements with Apple. Complaint, *supra* note 3, at 24. This Note does not discuss the exclusive dealing agreements though the district court and the Ninth Circuit disagreed on this front also.

15. Fed. Trade Comm'n v. Qualcomm, 411 F. Supp. 3d 658, 762, 743 (N.D. Cal. 2019) (finding Qualcomm had a duty to license its rivals and anticompetitively harmed OEMs).

16. *Id.* at 818.

17. Qualcomm, 969 F.3d at 987.

18. *Id.* at 987, 992.

overreaching district court;[19] others argue that the Ninth Circuit's reasoning grossly muddied the waters of antitrust law and will lead to future confusion.[20]

While this Note agrees with the outcome in the district court, its central critique is that neither the Ninth Circuit nor the district court properly factored Qualcomm's SEPs into the antitrust analysis.

This Note argues that *Qualcomm*[21] should have utilized the essential facilities doctrine, an antitrust theory of liability not mentioned in either court's decision.[22] The essential facilities doctrine appropriately accounts for Qualcomm's activity in two parts of the cellular market (patents and chips) and provides a framework through which to analyze the power of Qualcomm's SEPs. Had the courts applied this doctrine, the Ninth Circuit may have reached the opposite result and the district court may have reached the same result in a better way. This Note applies the elements of the essential facilities doctrine to the facts of *Qualcomm*. Even though this is a backward-looking exercise,[23] it illustrates how courts and parties should use the essential facilities doctrine as a tool of analysis in future antitrust cases involving SEPs and standards-dominated markets.

---

19. *See, e.g.*, Douglas H. Ginsburg, Joshua D. Wright & Lindsey M. Edwards, *Section 2 Mangled: FTC v. Qualcomm on the Duty to Deal, Price Squeezes, and Exclusive Dealing* 1 (George Mason L. & Econ. Rsch. Paper Series, No. 19-21), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3433564 (describing the district court's ruling as an "inappropriate extension of antitrust liability"); Matthew M. Martino, Tara L. Reinhart, Steven C. Sunshine & Julia K. York, *Ninth Circuit Strikes Down Sweeping Injunction Against Qualcomm and Reins in Expansive Interpretation of Sherman Act*, SKADDEN (Aug. 14, 2020), https://www.skadden.com/insights/publications/2020/08/ninth-circuit-strikes-down-sweeping-injunction [https://perma.cc/2NND-BVHW].

20. *See, e.g.*, Alex Moss, *Throwing Out the FTC's Suit Against Qualcomm Moves Antitrust Law in the Wrong Direction*, ELEC. FRONTIER FOUND. (Aug. 26, 2020), https://www.eff.org/deeplinks/2020/08/throwing-out-ftcs-suit-against-qualcomm-moves-antitrust-law-wrong-direction.

21. This Note uses "*Qualcomm*" to refer to the litigation as a whole, in both the district court and the Ninth Circuit. When referring to either the Ninth Circuit or the district court decision alone, the court will be identified.

22. Based on a non-exhaustive review of the record, it seems the essential facilities doctrine was also not mentioned in the underlying briefing; except for a cursory mention on page 101 of Qualcomm's motion for leave to file an enlarged opening brief. *See* Appellant's Unopposed Motion for Leave to File an Enlarged Opening Brief at 101, Fed. Trade Comm'n v. Qualcomm, 969 F.3d 974 (9th Cir. 2020) (No. 19-16122) ("As to chip sales to OEMs, the Sherman Act would not compel Qualcomm to sell chips to an OEM at all—because Qualcomm is not an essential facility and unquestionably has no antitrust 'duty to deal' with OEMs.").

23. On October 28, 2020, the Ninth Circuit denied the FTC's petition for rehearing en banc, marking the end of the three-year litigation. Order, Fed. Trade Comm'n v. Qualcomm, 969 F.3d 974 (9th Cir. 2020) (No. 1916122).

This Note will proceed as follows. Part II will provide background on Qualcomm's business model, as well as antitrust and patent law concepts relevant for analyzing *Qualcomm*. Part III will discuss and compare the holdings and reasoning of the two *Qualcomm* decisions. Part IV will introduce the essential facilities doctrine and make the case for its application to *Qualcomm* while also addressing criticisms of the doctrine and demonstrating how it might have affected the *Qualcomm* analysis. Part V concludes, noting the future applicability of the essential facilities doctrine in the SEP context.

II. BACKGROUND

The complexity of *Qualcomm* (and its significance) stems from Qualcomm's unique market position, which implicates antitrust law, patent law, and standard setting. This Part clarifies that complexity, laying the groundwork for the discussion and critique of the *Qualcomm* decisions in Parts III and IV. To that end, Section II.A describes in greater detail the two business practices at issue in *Qualcomm*—Qualcomm's "no license, no chips" policy and its refusal to license its SEPs to rival chipmakers. Section II.B turns to the legal background, first introducing the basic antitrust law framework governing *Qualcomm* and then discussing the law on unilateral refusals to deal. Section II.C turns to the patent issues at play in *Qualcomm*. First, it provides a primer on the relationship between antitrust and patent law. Then, it describes the standard-setting landscape in which Qualcomm's patents gained SEP-status, a landscape that both courts drew on in their analysis of Qualcomm's business practices.

A. QUALCOMM'S BUSINESS MODEL

Qualcomm is a San Diego-based company that manufactures semiconductors and software components used in all kinds of wireless technology, from cell phones to smart cars.[24] Qualcomm possesses a large patent portfolio relating to its technologies.[25] Some of these patents are cellular SEPs needed to practice 3G Code Division Multiple Access (CDMA) and 4G Long-Term Evolution (LTE) telecommunications standards.[26]

Two-thirds of Qualcomm's company value comes from its patent licensing business.[27] In addition, it has a successful business manufacturing and selling

24. *About Qualcomm*, QUALCOMM, https://www.qualcomm.com/company/about [https://perma.cc/HCD9-E6LN] (last visited Feb. 26, 2021).

25. Fed. Trade Comm'n v. Qualcomm, 411 F. Supp. 3d 658, 672 (N.D. Cal. 2019).

26. Fed. Trade Comm'n v. Qualcomm, 969 F.3d 974, 983 (9th Cir. 2020). Qualcomm's patents include cellular SEPs (the topic of this Note), non-cellular SEPs, and non-SEPs.

27. *Id.*

cellular modem chips that incorporate its patents.[28] This posture is unique in the industry—most companies with comparable patent portfolios do not manufacture chips, and vice versa.[29] And Qualcomm has considerable market share in both markets. Until 2016, its chips represented 90% of the CDMA chip market and 70% of the LTE chip market.[30] Qualcomm also commands 25% of global patent licensing revenue in the cellular handset space and more than 50% in the modem chip market.[31]

Qualcomm sells its chips to OEMs on the condition that OEMs enter into patent licensing agreements. This is referred to as Qualcomm's "no license, no chips" (NLNC) policy. OEMs pay Qualcomm a running royalty of 5% of the price of the handset, the cellular phone sold by the OEM.[32] It is estimated that Qualcomm makes about twenty dollars on every smartphone sold.[33]

Qualcomm does not offer patent licensing agreements to rival chipmakers. Instead, it enters into "CDMA ASIC Agreements" with these competitors, who practice Qualcomm's SEPs "by necessity."[34] These agreements provide that Qualcomm "promises not to assert its patents in exchange for the company promising not to sell its chips to unlicensed OEMs."[35] In addition, they impose "reporting requirements that allow Qualcomm to know the details of its rivals' chip supply agreements with various OEMs."[36] In exchange, rivals are permitted to practice Qualcomm's SEPs royalty-free.[37] Unlike a traditional license, these agreements constrict rivals' customer base to OEMs licensed with Qualcomm, rather than any OEM willing to purchase a rival chip. Together, the ASIC Agreements and NLNC policy ensure that any OEM who purchases a chip with embedded Qualcomm technologies will pay royalties to Qualcomm, regardless of whether the OEM buys the chip from Qualcomm or a rival chipmaker.

---

28. *Id.*
29. *Id.*
30. *Id.*
31. Qualcomm, 411 F. Supp. 3d at 674.
32. *Id.* at 673.
33. Rick Merritt, *Apple Reveals Qualcomm Patent Fees*, EETIMES (Jan. 14, 2019), https://www.eetimes.com/apple-reveals-qualcomm-patent-fees/ (noting also that the exact fee depends on the particular licensing agreement).
34. *Qualcomm*, 969 F.3d at 984. ASIC stands for "application-specific integrated circuit," and is another term for a modem chip.
35. *Id.*
36. *Id.* at 984–85.
37. *Id.*

The FTC challenged both Qualcomm's NLNC policy towards OEMs and its refusal to license its patents to rivals. The FTC alleged that Qualcomm's behavior was anticompetitive, violating sections 1 and 2 of the Sherman Act.[38]

B. RELEVANT ANTITRUST LAW CONCEPTS

1. *The Sherman Act and the Rule of Reason*

The Sherman Antitrust Act of 1890 was the first U.S. antitrust legislation.[39] Section 1 makes "[e]very contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce" illegal.[40] Section 2 makes it illegal to "monopolize, or attempt to monopolize, or combine or conspire . . . to monopolize any part of the trade or commerce . . . ."[41] In general, section 1 targets "concerted anticompetitive conduct" and section 2 targets "independent anticompetitive conduct,"[42] though sometimes a firm is held independently liable for conduct that violates section 1, as Qualcomm was in the district court decision.[43] Like *Qualcomm*, this Note focuses on the independent conduct (also referred to as unilateral or monopolizing conduct) targeted by section 2.[44]

In part because the Sherman Act's language is so broad, the debate about antitrust enforcement has been a debate over the proper scope and application of the law. That debate continues to this day. No one doubts that antitrust law

38. The FTC also alleged that Qualcomm violated section 5 of the FTC Act. Section 5 allows the FTC to prosecute a wider range of "unfair methods of competition" that may not incur Sherman Act liability but violate the spirit of the law. Because the district court found Qualcomm liable under Sherman Act sections 1 and 2, it did not specifically analyze Qualcomm's liability under the FTC Act. Fed. Trade Comm'n v. Qualcomm, 411 F. Supp. 3d 658, 812 (N.D. Cal. 2019)*; see also* 15 U.S.C. § 45.

39. 15 U.S.C. §§ 1–38. Antitrust law gained a foothold in the United States in the late nineteenth century, in part as a response to the increasing consolidation in the railroads and other industries. *See* Lina Khan, *Amazon's Antitrust Paradox*, 126 YALE L.J. 710, 739 (2017).

40. 15 U.S.C. § 1.

41. 15 U.S.C. § 2.

42. *Qualcomm*, 969 F.3d at 989–90 (emphasis removed).

43. *Qualcomm*, 411 F.3d at 812 (holding Qualcomm alone liable under section 1 for its licensing practices). Note that the district court did not clarify which theories of liability violated which Sherman Act section; it simply found that Qualcomm had violated both sections 1 and 2. *Id.* The *Qualcomm* decisions have been criticized for conflating the section 1 and 2 standards of liability. *See, e.g.*, Thomas F. Cotter, *Two Errors in the Ninth Circuit's* Qualcomm *Opinion*, PATENTLY-O PAT. L.J. (2020).

44. Liability under the essential facilities doctrine is generally thought to fall under Sherman Act section 2. However, as discussed in Part IV.A below, the inquiry under the essential facilities doctrine is distinct from the inquiry under other types of section 2 liability.

should protect and promote healthy competition—rather, the disagreement is over how best to do that and what healthy competition even looks like.[45]

As early as 1911, only twenty-one years after the passage of the Sherman Act, the Supreme Court transitioned from enforcing it literally (finding that *any* agreement that on its face restrained trade or *any* act of monopolization violated the Sherman Act), to using a "rule of reason" framework to analyze alleged antitrust violations.[46] In its simplest expression, this framework means that only *unreasonable* or *unlawful* contracts in restraint of trade or acts of monopolizing will be punished under sections 1 or 2.[47] For alleged unilateral violations, the question then becomes: what is an unreasonable or unlawful act of monopolizing? There is no easy answer, nor is the standard courts use to evaluate unilateral conduct easy to articulate without falling into tautology.[48] In general, courts will look to the circumstances of the case and the effects on the market to decide whether a practice is reasonable or unreasonable under the Sherman Act.[49] The rule of reason framework guides this inquiry.

A "threshold step" for analyzing unilateral conduct under the rule of reason framework is to define the "relevant market" for antitrust purposes.[50] This narrows the field of vision, allowing courts to "focus on anticompetitive effects in the market where competition is allegedly being restrained."[51] Injuries that occur outside the relevant market, even if caused by the allegedly anticompetitive behavior, are "beyond the scope of antitrust law."[52] One way

45. Though beyond the scope of this Note, the debate is currently dominated by two camps—those who advocate adhering to the currently dominant Chicago School-driven consumer welfare approach to antitrust and those who advocate expanding this understanding of consumer welfare and market competition to better account for dynamic competition in modern markets. On the latter side, see, for example, Khan, *supra* note 39, at 710 ("[P]egging competition to 'consumer welfare'. . . is unequipped to capture the architecture of market power in the modern economy."); Maurice E. Stucke, *Reconsidering Antitrust's Goals*, 53 B.C. L. REV. 551, 556 (2012) ("[D]efining the goals of antitrust is paramount: '[e]verything else follows from the answer we give.'" (internal citation omitted)).

46. *See* Standard Oil Co. v. United States, 221 U.S. 1, 62 (1911).

47. *Id.* at 87. This was a departure from the previously capacious interpretation of the Sherman Act. *See id.* at 83 (Harlan, J., concurring) (noting that those who remember the condition of the country in 1890 remember that the concern was the aggregation of capital, and thus the Court previously rejected attempts to limit the Sherman Act with a rule of reason).

48. Stucke, *supra* note 45, at 569 ("Antitrust becomes a tautology. The goal of competition law is 'promoting competition by discouraging anti-competitive behaviour.'").

49. Fed. Trade Comm'n v. Qualcomm, 969 F.3d 974, 989 (9th Cir. 2020) ("The rule of reason requires courts to conduct a fact-specific assessment of market power and market structure to assess the . . . *actual* effect on competition." (internal citations omitted)).

50. *Id.* at 992.

51. *Id.* (internal citations omitted).

52. *Id.* at 993.

of identifying the relevant market is the "hypothetical monopolist" test.[53] Courts ask whether a hypothetical monopolist could profitably impose a small but significant and non-transitory price increase (SSNIP) on the proposed market without losing its monopoly.[54] If the answer is yes, then the proposed market is a relevant market for antitrust purposes. But if the imposition of an SSNIP would cause customers to leave the hypothetical monopolist, the proposed market definition is too narrow.[55]

Once the relevant market is identified using this hypothetical, the court will analyze whether the defendant actually held market power in the relevant market.[56] Market power means the dominant firm has the ability to "profitably raise prices substantially above the competitive level."[57] Courts often use market share as a proxy for market power, though large market share may not always correlate with market power, and vice versa.[58] Other factors such as barriers to entry (the costs for a competitor to enter the market and compete) may be relevant to whether or not a firm possesses market power.[59] In antitrust, market power and monopoly power are often used interchangeably, as both indicate control over price. For a firm to be a monopolist for antitrust purposes, it must have this pricing power (with market share often used as a proxy), but it does not need to be the only player in the market.[60]

After establishing that the defendant had market power in the relevant market, courts engage in a three-part burden-shifting framework to analyze allegations of unilateral anticompetitive conduct.[61] First, the plaintiff must

---

53. HERBERT HOVENKAMP, PRINCIPLES OF ANTITRUST 67–69 (2nd ed. 2017) [hereinafter HOVENKAMP, PRINCIPLES OF ANTITRUST]; Fed. Trade Comm'n v. Qualcomm, Inc., 411 F. Supp. 3d 658, 684 (N.D. Cal. 2019).

54. *Qualcomm*, 411 F. Supp. 3d at 684.

55. *Id.* If customers leave, it suggests that there are available substitutions to the product such that even the hypothetical monopolist could not impose an SSNIP. Therefore, the market definition needs to be expanded to include these substitute products, since the hypothetical monopolist would need a monopoly over the substitutes to impose the SSNIP successfully. *See* Herbert Hovenkamp, *FRAND and Antitrust*, 105 CORNELL L. REV. 1683, 1707 (2020) [hereinafter Hovenkamp, *FRAND*] ("Any factor that limits substitution . . . can result in a narrower market definition.").

56. *Qualcomm*, 411 F. Supp. 3d at 684.

57. *Id.*

58. *Id.*

59. *Id.* at 684–85. The higher the barrier the entry, the more likely a firm already in the market will possess market power.

60. Qualcomm exemplifies this. In chips, Qualcomm has monopoly power for purposes of antitrust law, without fitting into the colloquial idea of a monopolist (a firm completely without competitors in the market).

61. *Qualcomm*, 411 F. Supp. 3d at 695; Fed. Trade Comm'n v. Qualcomm, 969 F.3d 974, 991 (9th Cir. 2020).

show that the defendant's exclusionary conduct (also referred to as anticompetitive conduct) had an anticompetitive effect in the relevant market. Next, if the plaintiff satisfies its burden, the defendant must proffer a non-pretextual procompetitive justification for its exclusionary conduct. Last, the plaintiff must either rebut the procompetitive justification or show that the anticompetitive harms of the exclusionary conduct outweigh its procompetitive benefits.[62]

What qualifies as an anticompetitive effect is up for debate.[63] In *Qualcomm*, the Ninth Circuit and the district court defined an anticompetitive effect as one that "harm[s] the competitive process and thereby harm[s] consumers," not just competitors.[64] The district court noted that possible anticompetitive effects might include "reduced output, increased prices, or decreased quality in the relevant market."[65] Overall, the process of identifying the relevant market, establishing market power, and engaging in a full rule of reason analysis can be an extremely complex, intensive, and burdensome undertaking.

2. *Unilateral Refusals to Deal*

Over time, categories of suspect unilateral conduct have emerged. Unilateral refusals to deal are one such category. While a firm's refusal to deal may be suspect, it only violates antitrust law if it is *anticompetitive*, which courts will assess using the rule of reason framework. This is how the courts in *Qualcomm* analyzed Qualcomm's refusal to license its rivals.

There is no general antitrust duty to deal with rivals "[i]n the absence of any purpose to create or maintain a monopoly."[66] In general, competitors can deal with rivals (or refuse to) as they please. Courts are hesitant to impose a duty to deal, in part because "compelled sharing may actually provide opportunities for collusion, which is the 'supreme evil of antitrust.'"[67]

---

62. *Qualcomm*, 969 F.3d at 991–92.

63. What constitutes "anticompetitive" can change based on one's definition of consumer welfare. *See* Stucke, *supra* note 45, at 576. While there is general agreement (based on Chicago School theories) that higher prices and lower output are anticompetitive effects, other schools of thought would include noneconomic effects or scrutinize other aspects of the overall market. *See, e.g.*, Khan, *supra* note 39, at 803 ("[C]urrent law underappreciates . . . how integration across distinct business lines may prove anticompetitive."). These debates are beyond the scope of this Note but will prove very important for antitrust law in the years to come.

64. *Qualcomm*, 969 F.3d at 990 (emphasis removed); *Qualcomm*, 411 F. Supp. 3d at 797.

65. *Qualcomm*, 411 F. Supp. 3d at 696.

66. United States v. Colgate & Co., 250 U.S. 300, 307 (1919).

67. Aerotec Int'l, Inc. v. Honeywell Int'l, Inc., 836 F.3d 1171, 1183 (9th Cir. 2016) (internal citation omitted).

However, in very limited circumstances, courts will impose a duty to deal on a firm whose refusal to deal is determined to be anticompetitive.

One such limited circumstance is recognized in *Aspen Skiing Co. v. Aspen Highlands Skiing Corp.*[68] In this case, Aspen Skiing operated three ski areas, while Aspen Highlands Skiing operated only one.[69] Though previously the companies collaborated through selling four-mountain ski passes, splitting the profits based on usage, Aspen Skiing terminated this agreement and began selling its own three-area passes.[70] Since three-mountain passes were more desirable for consumers than one-mountain passes, the competitor resort successfully argued that it was being squeezed out of business.[71] From this case came the *Aspen Skiing* exception, which imposes a duty to deal on a competitor who 1) terminates a profitable prior course of dealing and 2) whose only conceivable reason for the termination is to sacrifice short-term benefits for the long-term benefit of excluding competition.[72]

In the past, the Supreme Court also imposed a duty to deal where a vertically integrated[73] firm denied service to a downstream customer with which it competed. Imposing a duty to deal in these circumstances was in part based on leverage theory, which posits that a firm could use "dominance in one line of business to establish dominance in another."[74] *Otter Tail Power* exemplifies this approach.[75] Otter Tail owned the only local power transmission lines and operated as the exclusive retail power seller in 465 towns.[76] When Otter Tail's contracts ended, some towns voted to establish their own municipal retail power systems.[77] Otter Tail refused to sell wholesale power to these municipal systems and engaged in extensive litigation against them to maintain its status in the retail market.[78] The Court found that Otter Tail violated antitrust law by anticompetitively refusing to deal.[79] In so holding, the Court found the Federal Power Act had the authority to compel power companies to interconnect and did not displace antitrust law jurisdiction over

68. 472 U.S. 585 (1985).
69. *Id.* at 593.
70. *Id.* at 604.
71. *Id.* at 608.
72. Fed. Trade Comm'n v. Qualcomm, 969 F.3d 974, 993–94 (9th Cir. 2020).
73. A vertically integrated firm is one that operates in two lines of business that are upstream or downstream of each other. For example, a firm that manufactures tires and cars that use those tires would be vertically integrated.
74. *See* Khan, *supra* note 39, at 731.
75. Otter Tail Power Co. v. United States, 410 U.S. 366 (1973).
76. *Id.* at 370.
77. *Id.* at 371.
78. *Id.* 371–72.
79. *Id.* at 377.

power companies.[80] It also noted that Otter Tail's interest in maintaining its business, though legitimate, did not "immunize otherwise illegal conduct."[81]

A 2004 Supreme Court case, *Trinko*, turned away from the refusal to deal theories adopted in *Otter Tail Power*.[82] In *Trinko*, Verizon, the incumbent telephone service provider, denied access to its wholesale cellular infrastructure to new competitors in the local market. This violated the Telecommunications Act of 1996, which required incumbents to share their networks.[83] Though the Court agreed Verizon had violated its duty to deal under the Telecommunications Act, it found that Verizon had no antitrust duty to deal and did not fit into the *Aspen Skiing* exception.[84] Thus, Verizon had not violated antitrust law.

For the Court, the existence of the Telecommunications Act weighed against finding an antitrust duty to deal, even though the Act imposed on Verizon a duty to share its network. The Court noted that "the existence of a regulatory structure designed to deter and remedy anticompetitive harm" cautioned against using antitrust law as a remedy.[85] It emphasized that "[n]o court should impose a duty to deal that it cannot explain or adequately and reasonably supervise."[86] Thus, when "compulsory access" would "require[] the court to assume the day-to-day controls characteristic of a regulatory agency," the "problem should be deemed irremediable by antitrust law."[87] Overall, the Court in *Trinko* "expressed great skepticism about unilateral, unconditional refusal-to-deal cases," perhaps leaving only the *Aspen Skiing* exception.[88]

C. RELEVANT PATENT LAW CONCEPTS

Though *Qualcomm* is an antitrust case, patent law is an important backdrop informing its analysis. Since the FTC alleged that Qualcomm's patent licensing practices were anticompetitive, the two legal regimes collided. This Section

80. *Id.* at 374–75.

81. *Id.* at 380.

82. Verizon Commc'ns Inc. v. Law Offices of Curtis V. Trinko, 540 U.S. 398 (2004). *See* HOVENKAMP, PRINCIPLES OF ANTITRUST, *supra* note 53, at 311 (noting that *Trinko* "very likely ended the long run of a controversial doctrine [leverage theory]").

83. *Trinko*, 540 U.S. at 402–03. In addition, the FCC had already ordered Verizon to share its network. *Id.* at 413.

84. *Id.* at 409. The Court characterized *Aspen Skiing* as "at or near the outer boundary of § 2 liability." *Id.*

85. *Id.* at 412. *But see* Brett Frischmann & Spencer Weber Waller, *Revitalizing Essential Facilities*, 75 ANTITRUST L.J. 1, 25 (2008) ("Verizon was prepared to incur litigation expenses far in excess of this modest fine and reporting obligations to avoid the one set of penalties that actually would be effective in mandating nondiscriminatory access.").

86. *Trinko*, 540 U.S. at 415.

87. *Id.* (internal alterations omitted).

88. Melamed & Shapiro, *supra* note 2, at 2125.

first discusses the conceptual relationship between antitrust and patent law and then provides background on the specific patents at issue in this Note—SEPs and the standard-setting context in which those patent rights are created.

1. *Relationship Between Antitrust and Patent Law*

Patent law's constitutionally mandated mission is to "promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right" to their discoveries.[89] Patent law seeks to promote innovation by incentivizing inventors, rewarding their investments with the exclusive right to profit from an idea and control its use for a limited time.[90] Patent holders can choose to use their ideas exclusively, or to license them and collect royalty payments from licensees.[91] These exclusive rights prevent "free riders," who made no investment to develop the technology, from profiting off of it without cost.[92] However, patent protections are not granted in perpetuity, as widespread access to and application of patented ideas can be important sparks for further discovery.[93] Thus, always with the goal of promoting innovation, patent law operates along a continuum of exclusion and access, seeking to find the perfect equilibrium between the two.[94]

Though antitrust law also seeks to promote innovation, its means to this end look very different. While patent rights necessarily foreclose some level of competition, antitrust seeks to maximize competition, premised on the idea that unfettered competition leads to more innovation and a healthier economy (with happier consumers). In this respect, the two legal regimes can be at odds.

When the two regimes collide, one often simply gives way to the other[95]—however, patent rights should factor into the antitrust market analysis. Patent

---

89. U.S. CONST. art. 1 § 8, cl. 8.

90. *See* Oren Bracha & Talha Syed, *Beyond the Incentive-Access Paradigm-Product Differentiation & Copyright Revisited*, 92 TEX. L. REV. 1841, 1850 (2014) (noting that the "basic economic function of . . . patent protection is to enable the creator of an informational work to charge a price for assessing that work that recoups some of the sunk costs incurred in developing it"); *see also* Melamed & Shapiro, *supra* note 2, at 2121 ("The patent laws are intended to limit, not maximize, the royalties to which patent holders are entitled.").

91. *See* Moss, *supra* note 20 ("[P]atent owners have no obligation to let anyone use technology their patent covers.").

92. Frischmann, *supra* note 85, at 2 ("[A]t the core of intellectual property is the right to exclude, without which some producers would abandon their efforts for fear of free riding (unlicensed sharing) by competitors.").

93. *See* Melamed & Shapiro, *supra* note 2, at 2122 ("[T]he purpose of patent law is to promote innovation, not to maximize the returns to patent holders.").

94. Frischmann, *supra* note 85, at 2.

95. HERBERT HOVENKAMP, MARK D. JANIS, MARK A. LEMLEY, CHRISTOPHER R. LESLIE & MICHAEL A. CARRIER, IP AND ANTITRUST: AN ANALYSIS OF ANTITRUST PRINCIPLES APPLIED TO INTELLECTUAL PROPERTY LAW § 1.03, 18 (3rd ed. 2017) [hereinafter

rights confer market power.[96] A patent is a barrier to entry for competitors, and, if valuable, it allows a patent holder to set a supracompetitive price.[97] Of course, market power alone is not enough to generate antitrust scrutiny—a firm with market power must also *monopolize*, not just exist as a monopoly. But for antitrust purposes, a patent right does not create a presumption of market power or monopoly.[98] In general, then, the "exercise of intellectual property rights is . . . neither particularly free from scrutiny under the antitrust laws, nor particularly suspect under them."[99]

Ideally, the two legal regimes should work in tandem, so that more powerful patent rights are balanced by more detailed antitrust scrutiny.[100] Technological standard-setting, at least where those standards incorporate patented technology, is a particular area in which a balance must be struck between the competitive benefits of sharing resources and the innovative benefits of patent protection.

2. *Alphabet Soup: Standard-Setting Organizations (SSOs), Standard-Essential Patents (SEPs), and Fair, Reasonable, and Nondiscriminatory (FRAND) Commitments*

Standards "specify design features that enable products manufactured by different vendors to work together in a manner that is largely invisible to the consumer."[101] Standards and standardization are nothing new—older examples include electrical plug and drill bit specifications—but they have

---

HOVENKAMP, IP AND ANTITRUST] ("[T]he history of the intersection [of IP and antitrust] has been characterized by cycles of over- and under-enforcement, in which first antitrust and then intellectual property is on the ascendancy while the other recedes into the background."). Currently, in the United States, patent law seems to be on top. *See* Spencer Weber Waller, *The Omega Man or the Isolation of U.S. Antitrust Law*, 52 CONN. L. REV. 123, 161 (noting that former head of DOJ Antitrust, Makan Delrahim, is a member of the patent bar who "has been on record against the use of antitrust law to weaken IP protection for over twenty years").

96. *See* Bracha & Syed, *supra* note 90, at 1852 ("[I]t *is* necessary for IP protection, if it is to achieve its incentive function at all, to confer some supramarginal pricing power. In the absence of any degree of pricing power, there will be no added ability to recoup the fixed costs of development and no added incentive.").

97. *See id.*

98. It used to, but this precedent was overruled in *Illinois Tool Works v. Independent Ink*, 547 U.S. 28, 31 (2006).

99. DOJ & FTC, ANTITRUST GUIDELINES FOR THE LICENSING OF INTELLECTUAL PROPERTY 3 (Jan. 12, 2017) https://www.justice.gov/atr/IPguidelines/download.

100. *See* HOVENKAMP, IP AND ANTITRUST, *supra* note 95, at 18–19.

101. Jorge L. Contreras, *Technical Standards, Standards-Setting Organizations and Intellectual Property: A Survey of the Literature (With an Emphasis on Empirical Approaches) in* RESEARCH HANDBOOK ON THE ECONOMICS OF INTELLECTUAL PROPERTY LAW, Vol. II Analytical Methods 4 (Peter S. Menell & David Schwartz, eds., 2019), https://ssrn.com/abstract=2900540.

increased in prevalence and importance with the rise of network and telecommunications technologies, becoming "essential to the global technology infrastructure."[102] Standards in technology provide the important benefits of interoperability and "compatibility between products made by different manufacturers."[103] Today, standards "are pervasive in technology markets."[104]

Increasingly over the last two decades, "interoperability standards have been developed by groups of market participants that collaborate within voluntary associations known as . . . standards-setting organizations (SSOs)."[105] SSOs, generally comprised of industry participants, establish standards in a given industry. At issue in *Qualcomm* are 2G, 3G, and 4G telecommunications standards, all determined by SSOs. One such standard, in which Qualcomm technology dominates, is CDMA, a 2G and 3G standard.[106] Qualcomm's technologies also play a significant role in LTE, the primary 4G standard,[107] though to a lesser extent than in CDMA.[108] Importantly, LTE builds upon CDMA and other 3G standards.[109]

If an SSO adopts a standard that includes an already-patented technology, the patent on that technology becomes a standard-essential patent (SEP).[110] Once a standard is adopted, any firm that wishes to operate using the standard

102. *Id.*

103. Lemley, *supra* note 2, at 1893.

104. Contreras, *supra* note 101, at 7. Well-known examples of standards include USB, Wi-Fi, and Bluetooth. *Id.* at 4.

105. *Id.* at 4. It is an open question whether SSOs themselves present antitrust concerns, since they involve competitors collaborating, but that question is beyond the scope of this Note. *See* Lemley, *supra* note 2, at 1993; Hovenkamp, *FRAND*, *supra* note 55, at 1695 (describing SSOs as "bona fide joint ventures" that should not be undermined by antitrust law).

106. Complaint, *supra* note 3, at 6, 13.

107. *Id.* at 6.

108. *Id.* at 13–14.

109. *Id.* at 7. Because the standards build on each other, an LTE chip generally needs to comply with LTE, 2G, and 3G.

110. An SEP covers patented technology that has been "adopted as an 'essential' component" of a network standard. HOVENKAMP, PRINCIPLES OF ANTITRUST, *supra* note 53, at 327. Firms are generally required by SSOs to disclose to the SSO that they believe they hold patents covering technology essential to the standard. Contreras, *supra* note 101, at 23. Their judgments of essentiality are not tested by the SSO, except by subsequent litigation. *Id.* at 28 ("[T]he determination whether a particular patent is indeed essential to a particular standard is typically left to the patent holder."). One standard can contain zero, a few, or hundreds of patented technologies, all of which would become SEPs. *Id.* at 9.

"[has] no choice but to use the patented technology" embedded in that standard.[111] "[T]he patent holder owns an asset that is essential . . . ."[112]

Patent-holding members of SSOs generally volunteer to have their technology incorporated into standards because holding an SEP for technology that is part of an important standard is likely to increase demand for patent licenses.[113] Once an SEP is conferred by an SSO, an SEP holder may have an incentive to engage in patent holdup,[114] that is, to "'hold-up' the market by demanding excessive royalty rates after a standard has been widely adopted and manufacturers which have made investments in the standardized technology have become 'locked-in.'"[115]

To combat this risk and the significant power that comes with owning an SEP, the "price for [SEP] status" is generally an obligation to "license the patent to network participants at 'fair, reasonable, and nondiscriminatory' royalties," known as FRAND commitments.[116] However, "[v]irtually no SSO specifies" what FRAND commitments require in practice.[117] Further, whether and how courts will enforce FRAND commitments is relatively untested. Heated debate exists as to whether enforcing FRAND commitments would discourage innovation or act as a necessary limit to possible monopolistic opportunism.[118] Regardless, FRAND commitments represent an attempt to balance the innovative benefits of exclusion and the competitive benefits of access.

---

111. Melamed & Shapiro, *supra* note 2, at 2113.

112. *Id.*

113. Erik Hovenkamp & Timothy S. Simcoe, *Tying and Exclusion in FRAND Licensing: Evaluating Qualcomm* 7 (Boston Univ. Questrom Sch. of Bus. Rsch. Paper No. 3523797) (Jan. 22, 2020), https://ssrn.com/abstract=3523797.

114. Brief of *Amicus Curiae* Timothy J. Muris in Support of Appellee at 18, Fed. Trade Comm'n v. Qualcomm, 969 F.3d 974 (9th Cir. 2020), No. 19-16122 ("SEP holders have the incentive and ability to engage in holdup.").

115. Contreras, *supra* note 101, at 17.

116. HOVENKAMP, PRINCIPLES OF ANTITRUST, *supra* note 53, at 327.

117. Lemley, *supra* note 2, at 1964–65.

118. *Compare* Melamed & Shapiro, *supra* note 2, at 2116 [parenthetical needed], *with* Alden Abbott, *The New Madison Approach and the Harmonization of Antitrust and Patent Law: A Retrospective Summary*, IPWATCHDOG (Feb. 15, 2021), https://www.ipwatchdog.com/2021/02/15/new-madison-approach-harmonization-antitrust-patent-law-retrospective-summary/id=129912/ [https://perma.cc/A8BW-PXEA] (describing the New Madison Approach, which posits that holdup is not an antitrust problem and therefore antitrust law should not enforce FRAND).

III. THE QUALCOMM CASE

By way of refresher, Qualcomm holds numerous SEPs covering technology embedded in 2G, 3G, and 4G standards. Qualcomm is also the leading manufacturer of CDMA (2G, 3G) and LTE (4G) chips, all of which contain the SEP technology needed to make the chips standard-compliant. In addition, Qualcomm committed to at least two SSOs to license its SEPs on FRAND terms. Qualcomm engaged in two allegedly anticompetitive practices regarding its SEPs: (1) it refused to license its SEPs to rivals, instead offering them CDMA ASIC Agreements and (2) it required OEMs who wanted to purchase Qualcomm chips to license Qualcomm's SEPs ("no license, no chips").

In January of 2017, the FTC sued Qualcomm, challenging both practices as anticompetitive and in violation of antitrust law.[119] Specifically, it alleged that Qualcomm had engaged in both practices while holding a monopoly in the modem chip market (2006–2016 for CDMA chips and 2011–2016 for premium LTE chips). This, the FTC contended, meant Qualcomm had violated sections 1 and 2 of the Sherman Act.

This Part discusses the district court and Ninth Circuit opinions in *Qualcomm*. It illuminates the divisions between the two courts and grounds Part IV's analysis of how the essential facilities doctrine could have applied in this case.

A. THE DISTRICT COURT DECISION

In May 2019, Judge Lucy H. Koh of the Northern District of California issued a lengthy opinion finding that Qualcomm violated sections 1 and 2 of the Sherman Act. This decision rested on extensive factual findings[120] made at the conclusion of a ten-day bench trial.[121] The district court agreed with the FTC that both challenged practices were anticompetitive and illegal. The court issued a nationwide injunction against Qualcomm, which, among other things, prohibited Qualcomm from conditioning chip purchases on license status and required it to license its SEPs to rival chip suppliers on FRAND terms.[122]

As a threshold matter, the district court used the hypothetical monopolist test to determine that the relevant antitrust markets were the markets for

119. Fed. Trade Comm'n v. Qualcomm, 411 F. Supp. 3d 658, 675 (N.D. Cal. 2019).

120. Judge Koh cited to extensive documentary evidence, noted at length that she found Qualcomm's witnesses and documentary evidence prepared for trial unreliable, and then spent over fifty pages discussing Qualcomm's conduct towards individual OEMs and rivals. *Id.* at 676–80 (credibility issues); *id.* at 698–751 (conduct toward OEMs and rivals).

121. *Id.* at 669.

122. *Id.* at 820–21.

CDMA and LTE modem chips.[123] The court then found that Qualcomm held monopoly power in those markets, regardless of the company's assertions that rivals were "steadily . . . eating away" at Qualcomm's market share.[124] Qualcomm held almost 100% of market share in the CDMA chip market from 2006–2016, and still held almost 80% of market share in 2018.[125]

The district court then used the rule of reason framework to analyze the alleged anticompetitive conduct. Ultimately, it found that the FTC proved anticompetitive conduct and harm, and that Qualcomm's procompetitive justifications were pretextual and inadequate to justify the anticompetitive harms.[126]

1. *Qualcomm's Refusal to License Rivals*

The district court held that Qualcomm's refusal to license its rival chipmakers violated the Sherman Act. It stressed that Qualcomm's license refusals had promoted rival exit from the market, prevented rival entry, and delayed or hampered the success of other rivals.[127] The court found Qualcomm had a duty to deal with (license) its rivals because (1) Qualcomm's FRAND commitments required it to license its SEPs to rivals, and (2) even without the FRAND commitments, Qualcomm's conduct fit into the *Aspen Skiing* exception.

First, in an order granting partial summary judgment for the FTC, the district court held that Qualcomm's FRAND commitments to two SSOs required it to provide licenses to its rivals.[128] The court applied California contract law in reaching this conclusion.[129] Both the FTC and Qualcomm conceded that Qualcomm's written assurances that it would license on FRAND terms constituted binding contracts with the SSOs, but Qualcomm argued that these contracts only required it to provide licenses to OEMs.[130] The district court disagreed, finding that the text of the agreements and precedent made it clear that an "SEP holder that commits to license its SEPs on FRAND terms must license those SEPs to all applicants."[131] The district

123. *Id.* at 684.

124. *Id.* at 690.

125. *Id.* at 690–91. In LTE, Qualcomm's market share was almost 90% in 2014 and around 65% in 2017. *Id.* at 694.

126. *Id.* at 756, 778.

127. *Id.* at 744–51.

128. *Id.* at 751.

129. Order Granting FTC's Motion for Partial Summary Judgment, Fed. Trade Comm'n v. Qualcomm, 411 F. Supp. 3d 658 (N.D. Cal. 2019) (No. 17-cv-00220), 2018 WL 5848999, at *8.

130. *Id.* at *7, *9.

131. *Id.* at *11.

court revisited this reasoning in its final disposition of the case.[132] It recognized that standards, while important for interoperability, conferred disproportionate market power.[133] The court acknowledged that licensing to OEMs was more lucrative for Qualcomm (since royalties were based on the handset price), but did not accept this profit motive as a procompetitive justification for Qualcomm's behavior, dismissing it as "self-serving and pretextual."[134]

Second, the district court also held that Qualcomm fit into the *Aspen Skiing* exception because it had terminated a voluntary and profitable course of dealing with an anticompetitive intent and, therefore, had an antitrust duty to deal.[135] The court found that at one point Qualcomm had licensed its SEPs to rivals, but that it had stopped doing so because of changes to patent exhaustion law.[136] These changes meant that Qualcomm would no longer be able to assert its patents against OEMs who purchased chips from licensed rival chipmakers. The court did not accept Qualcomm's proffered procompetitive justifications (profit, reduced transaction costs) for ceasing what it deemed to be a voluntary and profitable course of dealing, finding that its change in policy was motivated by anticompetitive malice.[137]

2. *No License, No Chips Policy*

In addition, the district court found that Qualcomm used its chip monopolies to coerce OEMs into signing patent license agreements.[138] By forcing OEMs to become licensees before allowing them to buy Qualcomm's chips (which many OEMs relied on as their primary chip supply), Qualcomm ensured it could receive patent royalties based on the price of the smartphone.[139] In over forty pages of its opinion, the district court meticulously analyzed Qualcomm's conduct during licensing negotiations with OEMs and the resulting license agreements.[140] It found that, in multiple negotiations, Qualcomm had threatened to cut off chip supply in order to obtain more favorable licensing terms.[141] In other negotiations, Qualcomm had

132. *Qualcomm*, 411 F. Supp. 3d at 751.
133. *Id.*
134. *Id.* at 756.
135. *Id.* at 762.
136. *Id.* at 754. Patent exhaustion is a doctrine establishing that the sale of an item embedding patented technology terminates the patent rights in that technology for the purposes of that one item ("exhausts" them). *Id.* at 698.
137. *Id.* at 760.
138. *Id.* at 698.
139. *Id.* at 773.
140. *Id.* at 697–744.
141. *Id.* at 743.

offered significant royalty rebates if an OEM agreed to source its chips exclusively from Qualcomm.[142]

Overall, the district court found, this conduct resulted in significant anticompetitive harms to OEMs and rivals.[143] OEMs were harmed because they were forced to pay Qualcomm's unreasonably high royalty rates based on the handset price.[144] Moreover, OEMs hesitated to challenge Qualcomm's rates with patent litigation because of possible threats to their chip supply.[145] The court reasoned that because OEMs had to pay royalties to Qualcomm no matter whose chips they purchased (since rivals with CDMA ASIC Agreements could not sell to unlicensed OEMs), they simply decided to stay with Qualcomm.[146] In turn, rival chipmakers were harmed by the lost OEM business.[147]

Throughout its opinion, the district court repeatedly emphasized that Qualcomm's practices operated in tandem, reinforcing one another and creating insurmountable barriers for rivals.[148]

B. THE NINTH CIRCUIT DECISION

Qualcomm appealed and, in August 2020, a three-judge panel of the Ninth Circuit summarily reversed all of the district court's findings and its injunction, in a far shorter opinion.[149] Though the Ninth Circuit agreed with the district court on the relevant antitrust market (the modem chip market) and the applicable legal standard (the rule of reason three-part burden-shifting test), their agreement stopped there.[150] The Ninth Circuit criticized the district court for focusing on licensing harm to OEMs which, it concluded, were outside the relevant antitrust market and therefore outside the scope of antitrust law.[151] It also criticized the district court for looking beyond the modem chip market to the effects that Qualcomm's conduct had on the larger cellular market.[152] Finally, it described the district court's analysis of the interrelatedness of Qualcomm's practices as inappropriate for antitrust.[153]

---

142. *Id.* at 744.
143. *Id.* at 743–44.
144. *Id.* at 773.
145. *Id.* at 786–88.
146. *Id.* at 791.
147. *Id.*
148. *Id.* at 795–97.
149. Fed. Trade Comm'n v. Qualcomm, 969 F.3d 974, 987 (9th Cir. 2020).
150. *Id.* at 991–92.
151. *Id.* at 993.
152. *Id.* at 992–93.
153. *Id.* at 993.

Overall, the Ninth Circuit did not find any of Qualcomm's practices anticompetitive, only "hypercompetitive," which, it emphasized, made them perfectly legal.[154] Throughout its opinion, the court preached deference to technology markets, patent and contract law, and monopolists.[155] Since, in the Ninth Circuit's view, the FTC did not meet its burden of proving anticompetitive harm, it did not deeply analyze Qualcomm's procompetitive justifications, though it suggested it would have likely accepted them.[156]

1. *Qualcomm's Refusal to License Rivals*

The Ninth Circuit overturned the district court's holding that Qualcomm had a duty to license its SEPs to its competitors.

First, it found that Qualcomm's conduct did not fit into the *Aspen Skiing* exception, which it stated was the "one, limited exception" to the general rule that there is no duty to deal.[157] It did not fit *Aspen Skiing* because, in the Ninth Circuit's view, Qualcomm had not wrongfully terminated a previous course of dealing in which it had granted licenses to rival chipmakers.[158] It had previously granted non-exhaustive licenses, but had ceased this practice in response to changes in patent exhaustion law.[159] This, the court reasoned, was a legitimate reason for terminating the practice, not one motivated by "sacrific[ing] short-term benefits in order to obtain higher profits in the long run from the exclusion of competition."[160] The Ninth Circuit concluded that Qualcomm simply "[chose] the path that was 'far more lucrative,' both in the short term *and* the long term, regardless of any impacts on competition."[161]

The Ninth Circuit also applauded the CDMA ASIC Agreements. It repeatedly noted that rivals benefited from being able to practice Qualcomm's patents royalty-free and found these agreements acted as "de facto licenses" (since Qualcomm promised not to assert its patents).[162] The court

154. *Id.* at 982, 1005.

155. *Id.* at 990 (noting the court is especially hesitant to declare novel business practices unreasonable in technology markets); *id.* at 997 ("[T]he antitrust laws are not well suited to govern contract disputes between private parties in light of remedies available under contract or patent law." (internal citation omitted)); *id.* at 990 ("The mere possession of monopoly power, and the concomitant charging of monopoly prices, is not itself unlawful; instead, it is an important element of the free-market system." (internal quotations removed) (citing Verizon Commc'ns Inc. v. Law Offices of Curtis V. Trinko, 540 U.S. 398, 407 (2004))).

156. *Id.* at 996.

157. *Id.* at 993. The FTC had conceded error on the district court's *Aspen Skiing* conclusion. *Id.* at 995.

158. *Id.* at 994.

159. *Id.*

160. *Id.*

161. *Id.* (internal citation omitted).

162. *Id.* at 996.

characterized the agreements as Qualcomm's "no license, no problem" policy.[163]

Second, the Ninth Circuit did not find a duty to deal based on Qualcomm's FRAND commitments. The court did not decide the issue of whether Qualcomm's refusal to license rival chipmakers was a breach of its FRAND commitments to SSOs, vacating the district court's holding to that effect.[164] It leaned on the statements of *amici* and scholars to note that contract law, rather than antitrust, may be the appropriate forum for FRAND disputes.[165]

However, it did note that *even if* Qualcomm had breached its FRAND commitments, the FTC did not "satisfactorily explain" how that breach caused anticompetitive harm, since competitors could still practice the patents royalty-free.[166] And, even if Qualcomm breaching its FRAND commitments had harmed competitors, that was not enough to show harm to *competition*, which the FTC needed to prove.[167] The Ninth Circuit emphasized that a few competitors had entered the modem chip market and dismissed the claims that Qualcomm's refusal to deal had harmed rivals' ability to develop and retain OEM customers, limited their growth, and delayed or prevented their market entry.[168] It was also relevant for the court that other companies with significant SEP portfolios had followed Qualcomm's lead in adopting OEM-level licensing policies.[169] Ultimately, the Ninth Circuit decided that the FTC had not met its burden of proving anticompetitive harm under the rule of reason framework.

2. *No License, No Chips Policy*

The Ninth Circuit also found that Qualcomm's NLNC policy did not run afoul of antitrust law. The court, in upholding the policy, repeatedly emphasized that it was "chip-supplier neutral."[170] OEMs were required to pay royalties to Qualcomm no matter who they sourced their chips from, and all chips contained Qualcomm's patents, to which Qualcomm was entitled to

---

163. *Id.* at 995. For an interesting look at this issue, *see* Jorge L. Contreras, *"No License, No Problem"–Is Qualcomm's Antitrust Victory a Patent Exhaustion Defeat?*, PATENTLY-O BLOG (Sept. 1, 2020), https://patentlyo.com/patent/2020/09/qualcomms-antitrust-exhaustion.html [https://perma.cc/YUM5-NGZA] (arguing that the ASIC Agreements look "suspiciously like licensing," and could be construed as such by future courts and thus be subject to patent exhaustion).

164. *Qualcomm*, 969 F.3d at 987, 995.

165. *Id.* at 997.

166. *Id.* at 995–96.

167. *Id.* at 996.

168. *Id.* (noting that Intel and Mediatek entered the chip market in 2015–2016).

169. *Id.*

170. *Id.* at 996, 1005.

royalties; therefore, in the Ninth Circuit's view, no one chipmaker was at a disadvantage.

In addition, the Ninth Circuit did not accept the district court's finding that Qualcomm's royalty rates were unreasonably high. Even assuming the rates were unreasonable, the Ninth Circuit dismissed this as an issue for patent law.[171] It further noted that the unreasonably high royalty rates, if they existed, would harm OEMs, and such harm was beyond the scope of antitrust since OEMs were customers, not competitors, in the relevant antitrust market.[172]

IV. QUALCOMM AND THE ESSENTIAL FACILITIES DOCTRINE

The two opinions in *Qualcomm* have one important commonality: neither adequately addresses the impact of Qualcomm's status as an SEP holder on the antitrust analysis. Though the two courts agreed that the relevant market for antitrust purposes was the modem chip market (and certainly, the chip market is one relevant market), both ignored the ways in which Qualcomm's SEPs affected the economic realities of that market and therefore should have affected the antitrust analysis. This Part argues that the essential facilities doctrine (EFD), a doctrine that can impose a duty to deal on a firm that owns an essential facility, provides an effective lens through which to analyze Qualcomm's SEPs, a lens that is conceptually clearer and appropriately cognizant of Qualcomm's dual monopolies (in chips and in SEPs). Applying the essential facilities doctrine to the facts of *Qualcomm* illuminates the holes in the Ninth Circuit's approach, leading to the same result as the district court (that a duty to deal should be imposed on Qualcomm), but providing a better, simpler way of getting there.

This Part is broken down into four sections. Section IV.A introduces the essential facilities doctrine and addresses critiques against it. Section IV.B explains why the EFD should apply in the SEP context. Section IV.C applies the EFD to the facts of *Qualcomm*, showing that Qualcomm's conduct likely met the traditional criteria for liability under an EFD theory. Section IV.D discusses the advantages of this approach over the three-part burden-shifting test used by the district court and the Ninth Circuit.

171. *Id.* at 999.

172. *Id.* at 999–1000. This reasoning has been widely criticized as contrary to settled antitrust law. *See, e.g.*, FTC Petition for Rehearing *En Banc* at 3–4, Fed. Trade Comm'n v. Qualcomm, 969 F.3d 974 (9th Cir. 2020) (No. 19-16122).

A. THE ESSENTIAL FACILITIES DOCTRINE

The essential facilities doctrine is an antitrust doctrine that imposes a duty to deal with rivals on a firm that controls an "essential facility." Because the essential facilities doctrine imposes a duty to deal, it is often referred to as a variation on refusal to deal analysis. However, the essential facilities doctrine does not require the same relevant market inquiry or utilize the three-part burden-shifting test. Though the Supreme Court has never officially adopted the doctrine,[173] various scholars and lower courts that use the doctrine find support for it in Supreme Court precedent.[174] For example, *Otter Tail Power* is often cited as an essential facilities case, though the case never mentions the term.[175]

The Supreme Court's decision in *Terminal Railroad*[176] is regularly cited as the origin of the EFD, though again, the case itself does not use the term.[177] In *Terminal Railroad*, the Court, after finding that the Terminal Railroad Association had illegally monopolized the St. Louis railroad industry, declined to apply the traditional antitrust remedy for monopolizing—dissolution—and instead required the Terminal Railroad Association to give its competitors access to its tracks and bridges, its essential facilities.[178] This creative remedy was prompted by the Court's recognition that the railroads might be a "natural monopoly" with high barriers to entry, such that breaking up the Terminal Railroad Association would create socially undesirable inefficiencies.[179] Since *Terminal Railroad*, courts have applied the EFD or its analytical equivalents to telephone infrastructure, sports stadiums, radio stations, and more.[180]

---

173. *See* Verizon Commc'ns Inc. v. Law Offices of Curtis V. Trinko, 540 U.S. 398, 410 (2004) (noting that the Court does not consider the essential facilities doctrine "crafted by some lower courts" established law).

174. *See, e.g.*, Frischmann, *supra* note 85, at 6 (noting that *Otter Tail* looks like an essential facilities case); MCI Commc'ns Corp. v. Am. Tel. and Tel. Co., 708 F.2d 1081, 1132 (7th Cir. 1983) (applying the essential facilities doctrine and citing to *Otter Tail* and *Terminal Railroad* for support).

175. Phillip Areeda, *Essential Facilities: An Epithet in Need of Limiting Principles*, 58 ANTITRUST L.J. 841, 847 (1990).

176. United States v. Terminal R.R. Ass'n of St. Louis, 224 U.S. 383 (1912).

177. *See, e.g.*, Intergraph Corp. v. Intel Corp., 195 F.3d 1346, 1356 (Fed. Cir. 1999) (noting the EFD stems from *Terminal Railroad*); Teague I. Donahey, *Terminal Railroad Revisited: Using the Essential Facilities Doctrine to Ensure Accessibility to Internet Software Standards*, 25 AIPLA Q.J. 277, 279 (1997) (describing the EFD as a "creative antitrust remedy emanating from *Terminal Railroad*").

178. Donahey, *supra* note 177, at 281.

179. *Id.* at 281–82.

180. MCI Commc'ns Corp. v. Am. Tel. and Tel. Co., 708 F.2d 1081, 1132 (7th Cir. 1983) (telephone infrastructure); Hecht v. Pro-Football, Inc., 570 F.2d 982, 992–93 (D.C. Cir. 1977)

The EFD does not require a firm to give rivals free access, only reasonable and nondiscriminatory access.[181] Like *Aspen Skiing*, the EFD is a qualification on the general right to refuse to deal, one that can form the basis for monopolist liability.[182] *Trinko* arguably limited the EFD's viability in dicta;[183] however, the Ninth Circuit and other lower courts continue to recognize and utilize the doctrine post-*Trinko*.[184]

There is no perfectly settled articulation of the elements of the essential facilities doctrine. The threshold inquiry, however, is common: in order to succeed on an essential facilities claim, a plaintiff must first prove that access to the facility at issue is truly essential—an input necessary for other firms to compete.[185] This implies that the facility cannot reasonably be duplicated by competitors.[186] Often, the essential facility is some type of infrastructure that would be prohibitively expensive or inefficient for a competitor to recreate.[187] In the Ninth Circuit, a firm's facility will only be considered essential if control of the facility carries with it the power to eliminate, or at least severely handicap, competition in the downstream market.[188]

Another (often implicit) requirement for an essential facilities claim is that the owner of the essential facility be in downstream competition with those who seek access to that facility. In other words, the essential facility owner is vertically integrated, operating the essential facility upstream and competing in a downstream market that requires the facility as an input.[189] This requirement makes sense—without competition between the facility owner and those who seek to use the facility, the facility owner would have no incentive to deny

---

(football stadium); Nobody in Particular Presents, Inc. v. Clear Channel Commc'ns, Inc., 311 F. Supp. 2d 1048, 1109 (D. Colo. 2004) (rock radio station).

181. *See* Frischmann, *supra* note 85, at 11.

182. *See* Robert Pitofsky, Donna Patterson & Jonathan Hooks, *The Essential Facilities Doctrine Under U.S. Antitrust Law*, 70 ANTITRUST L.J. 443, 446 (2002).

183. *Trinko*'s dicta and this Note's response to it will be discussed in Part IV.D.

184. *See, e.g.*, Aerotec Int'l, Inc. v. Honeywell Int'l, Inc., 836 F.3d 1171, 1184 (9th Cir. 2016) ("Although the Supreme Court has never recognized the doctrine, we have continued to treat it as having a basis in § 2 of the Sherman Act.").

185. *See* Pitofsky et al., *supra* note 182, at 449.

186. If competitors could easily duplicate it, they would not need access.

187. *See, e.g.*, MCI Commc'ns Corp. v. Am. Tel. and Tel. Co., 708 F.2d 1081, 1132 (7th Cir. 1983) (telephone infrastructure).

188. Alaska Airlines v. United Airlines, 948 F.2d 536, 544 (9th Cir. 1991).

189. *See* Paul D. Marquardt & Mark Leddy, *The Essential Facilities Doctrine and Intellectual Property Rights: A Response to Pitofsky, Patterson, and Hooks*, 70 ANTITRUST L.J. 847, 849–52 (2003) (describing the competitive element as the "leveraging element" of the essential facilities doctrine and noting that judicial opinions sometimes "[take] that element for granted"). Alaska Airlines v. United Airlines, 948 F.2d 536, 544 (9th Cir. 1991).

access.[190] *Otter Tail Power* is again illustrative: Otter Tail was the only regional distributor of power, and it also competed in the retail power markets downstream.

Once the plaintiff establishes essentiality and competition, courts will analyze whether access to the facility was actually denied and whether access could have feasibly been given.[191] Offering access only on terms that are unreasonable can amount to a practical denial of access.[192] However, it is not necessary to grant access on whatever terms the rival firm would prefer.[193] Denial of access is also not *per se* unlawful; if the firm can provide a legitimate business reason for the denial, the court may not require it to give access.[194]

Though evidence of bad intent or anticompetitive effects can supplement an EFD claim, it is not required. Once a plaintiff has cleared the high bar of proving that access to a facility is truly essential, such that denying access eliminates downstream competition, the danger is clear and there is "little need to engage in the usual lengthy analysis of factors such as intent."[195]

The essential facilities doctrine has been criticized on numerous fronts. Six of these criticisms, and their rejoinders, are worth addressing now.

First, the doctrine has been characterized as overbroad.[196] Critics who take this view posit that essentiality is too easy to allege, such that any successful product could be deemed essential under the EFD. However, "the vast majority of essential facilities claims are rejected by the courts," suggesting that essentiality is a sufficiently high bar to protect against the EFD's overuse.[197]

Second, some suggest that the doctrine is superfluous, simply recreating a duty to deal analysis.[198] This argument, too, ignores the significance of the essentiality requirement, which not only creates a high barrier to entry but also simplifies the overall analysis once that barrier is cleared. Unlike in a traditional

190. *See id.* Note that this means leverage theory, the idea that a vertically integrated firm can leverage dominance in one market for power in another, is implicit in the EFD.

191. *MCI*, 708 F.2d at 1132–33.

192. Delaware & Hudson Ry. Co. v. Consol. Rail Corp., 902 F.2d 174, 179–80 (2d Cir. 1990) ("[T]here need not be an outright refusal to deal in order to find that denial of an essential facility occurred. It is sufficient if the terms of the offer to deal are unreasonable.").

193. MetroNet Servs. Corp. v. Qwest Corp., 383 F.3d 1124, 1130 (9th Cir. 2004).

194. *See* Areeda, *supra* note 175, at 852.

195. *See* Alaska Airlines v. United Airlines, 948 F.2d 536, 546 (9th Cir. 1991); *see also* Donahey, *supra* note 177, at 308 (noting that the EFD is more about remedy than conduct).

196. Mark A. Lemley, *Antitrust and the Internet Standardization Problem*, 28 CONN. L. REV. 1041, 1084 (1996).

197. *See id.*; *see, e.g.*, Aerotec Int'l, Inc. v. Honeywell Int'l, Inc., 836 F.3d 1171, 1185 (9th Cir. 2016) (rejecting the claim where the plaintiff failed to prove that the defendant's repair parts were essential).

198. HOVENKAMP, PRINCIPLES OF ANTITRUST, *supra* note 53, at 287.

refusal to deal case, courts need not scrutinize intent or anticompetitive effects to find liability. Thus, the essential facilities doctrine is a twist on duty to deal analysis that almost amounts to strict liability for the denial of access to essential facilities, once essentiality and denial of access are proven.

A third salient concern is that applying the essential facilities doctrine creates harmful incentives such as de-incentivizing incumbents from investing in essential infrastructure and incentivizing latecomers to free-ride.[199] This argument, however, is "overblown."[200] Incumbents still have incentives to develop essential facilities since being first to market generally has significant benefits.[201] In addition, even if an incumbent is required to give access to its essential facility, this does not preclude it from "continuing to receive a return, even a monopoly return, on that investment."[202] Nothing in the essential facilities doctrine prohibits a firm from charging rivals a monopoly price for access, so long as the monopoly price is reasonable.[203]

Fourth, some take the view that requiring access to essential facilities does not improve consumer welfare.[204] This critique rests on the premise that access to essential facilities will not affect prices for consumers, since essential facilities owners will just charge their competitors the monopoly price and that price will trickle down to consumers.[205] However, even if this occurs, access to essential facilities may have benefits to consumers outside of pricing. Namely, increased access may lead to more competition in downstream markets, which could lead to increased product quality.[206]

Fifth, some argue that forcing competitor firms to deal with one another poses a risk for collusion, the prime evil of antitrust. These risks are possible, but unlikely: if collusion were desirable, the essential facility owner would likely have granted access of its own accord.[207] And in some markets, the benefits of cooperation, such as increased innovation by new entrants and downstream users in the market, may outweigh the risk of potential collusion.[208]

Lastly, critics of EFD argue that it is not administrable, either because courts will struggle to decide when access should be granted or because courts

199. Frischmann, *supra* note 85, at 31–32.
200. *Id.* at 32.
201. *Id.* at 33.
202. *Id.* at 34.
203. *Id.*
204. *Id.* at 28.
205. *Id.*
206. *Id.* at 29–31.
207. *Id.* at 40.
208. *See id.*

are not well-equipped to monitor the terms of access.[209] However, this underestimates the conceptual clarity of the doctrine. "The essential facilities doctrine is one of the circumstances in which plain English and antitrust lingo converge."[210] Courts are as well positioned as anyone to ask simply: Was this facility essential and, if so, was access to it denied?[211] The remedy, compulsory access on reasonable and nondiscriminatory terms, can also be straightforward. Courts need not decide whether access standards are "correct in some cosmic sense."[212] Moreover, in many cases, antitrust courts may be the only recourse for downstream competitors seeking access to an essential facility.

In sum, the criticisms of the essential facilities doctrine do not persuasively undermine its utility. Accordingly, the doctrine should be recognized as an important analytical tool for antitrust law.

B. EFD AND SEPS

How might the essential facilities doctrine apply to intangible resources? In today's world, open and easy communication, made possible by technology compatibility standards, is indisputably essential. Yet there is considerable debate about the wisdom of applying the EFD to intellectual property rights and categorizing an IP right as an essential facility. While some strongly support applying essential facilities analysis to these intangible assets,[213] others fiercely oppose it, seeing the EFD as too great a threat to intellectual property rights.[214] When its elements are satisfied, the EFD regime mandates access. Ultimately, the debate over the application of EFD to patent rights reflects the larger conflict between antitrust and patent law—how best to balance exclusion and access to protect innovation and competition. In large part, opposition to using the EFD in the intellectual property context stems from a concern that painting IP rights with too broad a brush (deeming too many

209. *Id.*; Areeda, *supra* note 175, at 853 ("No court should impose a duty to deal that it cannot explain or adequately and reasonably supervise.").

210. Aerotec Int'l, Inc. v. Honeywell Int'l, Inc., 836 F.3d 1171, 1184 (9th Cir. 2016).

211. *See* Frischmann, *supra* note 85, at 41–42 ("These [essential facilities cases] are dime-a-dozen types of decisions that are a far cry from the polycentric multivariate balancing-type of cases that courts are comparatively poorer at deciding. If one concludes that courts cannot handle this kind of dispute, then most of the federal docket should be discarded . . . .").

212. *Id.* at 44.

213. *See, e.g.*, Donahey, *supra* note 177, at 279 (arguing for applying EFD to software interface standards).

214. *See, e.g.*, Marquardt, *supra* note 189, at 859 ("Casual and unfettered application of essential facilities principles to intellectual property threatens to penalize the most successful innovations for their success, using the antitrust laws to eviscerate core intellectual property protections and incentives.").

patent rights essential) will cut away at patent law too greatly.[215] This Note takes the position that the EFD is not too broad a brush and that it is a necessary tool in the standard-essential patent context.

This Section lays the conceptual groundwork for why the EFD can and should apply to SEPs. It argues that an SEP can be an essential facility, confer a monopoly, and constitute a relevant antitrust market. However, though this Note argues that the essential facilities doctrine should always be available as a tool for analyzing SEP holder conduct, it does not suggest that every SEP will qualify as an essential facility under the EFD, or that every denial of an SEP license will qualify for antitrust liability under the EFD.

1. An SEP is an essential facility. Or, at least, it can be.

The essential facilities doctrine's threshold inquiry is whether or not the facility at issue is "essential." Courts define "essential" for purposes of the doctrine as an "input without which a firm cannot compete with the monopolist."[216] Accordingly, the facility must be impractical or unreasonable for competitors to duplicate.[217] This is a high bar, and courts will dismiss essential facilities claims in which essentiality is not proven.[218]

SEPs can meet this essentiality requirement, though a case-by-case inquiry is necessary since in some circumstances an SEP, despite its name, might not be truly essential.[219] Courts and parties must analyze the essentiality of a given SEP like they would any other alleged essential facility. Though the standard-essential patent has already been deemed "essential" by its holder and the SSO, further proof beyond SEP status will help screen out non-meritorious EFD cases.

SEPs, like railroad tracks or power lines, are impractical to duplicate. Even assuming that rivals could theoretically develop identical technologies (an endeavor that would cost significant time and resources), doing so without permission from the patent holder would be illegal, leading to patent

215. *See, e.g.*, *id.*

216. Pitofsky et al., *supra* note 182, at 449.

217. Aerotec Int'l, Inc. v. Honeywell Int'l, Inc., 836 F.3d 1171, 1185 (9th Cir. 2016).

218. *See, e.g.*, *id.* (dismissing the claim where the competitor had access to the alleged essential facility from another source); *see also* Pitofsky et al., *supra* note 182, at 449 (noting that courts rarely impose liability under EFD, in part because of the essentiality requirement).

219. It is possible that a declared SEP was improperly disclosed to the SSO (and therefore is not actually necessary to the standard). *See* Contreras, *supra* note 101, at 28. It is also possible that, in some standard-setting industries, multiple standards compete with each other, such that a rival could avoid one SEP by switching to another standard. Or a standard might have been declared but not utilized widely in practice. In short, the context will matter, and courts and parties should be willing to dive into the circumstances to determine whether a particular SEP is essential.

infringement liability. These realities of the patent landscape illustrate that SEPs and the technologies they cover are impractical (and illegal) for rivals to duplicate. In other words, the barriers to entry are insurmountable. Thus, where an SEP's technology is truly essential, an SEP can qualify as an essential facility.

2. *An SEP holder is a monopolist and the market for SEPs is a relevant antitrust market.*

The owner of the essential facility must also be a monopolist in order to incur possible liability under the EFD. In some ways, this requirement restates the essentiality requirement, and the two will often travel together. However, it is a slightly separate inquiry since, in theory, there could be competition and multiple equivalent inputs at the essential facilities level that would make the essential facilities doctrine inapplicable.

There is debate over whether intellectual property rights can properly form the basis of a monopoly in a relevant antitrust market.[220] Those who argue a patent right should not be considered a monopoly for purposes of antitrust take the position that the knowledge that forms the basis of the right is not a product or market, and so it is improper to view it as granting market power.[221] However, this argument ignores economic reality. Certainly, not *all* intellectual property rights confer monopolies over a relevant market just because they afford a creator exclusive rights to her idea. Nevertheless, it is possible for the exclusive rights inherent in a patent grant to confer monopoly power.[222] When there are no alternatives to a given patented technology and demand for that patent is consistently high, such that the patent holder can engage in monopoly pricing of royalties, it is appropriate to speak about a patent monopoly over a relevant antitrust market.

Therefore, where an SEP is truly essential and there are no substitutes for the technology covered by the SEP, the market for SEP licenses is also a relevant market for antitrust purposes. Using the hypothetical monopolist test,

---

220. Bracha & Syed, *supra* note 90, at 1850 (noting that there is a "somewhat misguided debate over whether IP rights create 'monopoly power' or are 'merely' property"); s*ee also* HOVENKAMP, PRINCIPLES OF ANTITRUST, *supra* note 53, at 317 ("[I]t would be better not to speak of a patent 'monopoly' at all.").

221. *See, e.g.*, *id.*

222. *See* Bracha & Syed, *supra* note 90, at 1851 ("The reply from commentators . . . is to concede that IP rights do not always create monopolies, but then rejoin that monopoly power is a matter of degree and that in some cases, the exclusionary power created by IP rights does rise to the level of monopoly."); Hovenkamp, *FRAND*, *supra* note 55, at 1690 ("Once the standard is adopted and implementers have incorporated it into their own technologies, a standard essential patent is likely to be in a much stronger position, approaching monopoly in some cases.").

a hypothetical SEP monopolist in this scenario would likely be able to impose a small but significant and non-transitory price increase on its royalty rate without losing customers. This is in part because implementers are locked into a particular standard and the SEP technologies that come along with it. These locked-in customers will pay the increased SEP price by necessity. Thus, an SEP right can create both a monopoly and a relevant antitrust market where the SEP is truly essential, evincing the applicability of the EFD to SEPs.

C. APPLYING THE EFD TEST TO QUALCOMM

There is no perfectly settled articulation of the elements of the essential facilities doctrine, but the elements laid out in *MCI v. AT&T*[223] are often referred to as the leading version of the doctrine.[224] In *MCI*, the Seventh Circuit put forth four elements of the essential facilities doctrine: (1) control of the essential facility by a monopolist, (2) a competitor's inability practically or reasonably to duplicate the essential facility, (3) the denial of the use of the facility to a competitor, and (4) the feasibility of providing the facility.[225]

These elements, though useful for guidance, are both incomplete and redundant. The elements are incomplete in large part because they do not make explicit the implicit threshold inquiry: whether the facility at issue is truly essential for competition.[226] Another implicit element of the EFD inquiry that is missing from the *MCI* articulation is that the firm who controls the essential facility generally must be in competition with the firm seeking access.[227] This competition element provides the incentive for the firm to deny access to the essential facility. The second *MCI* element is redundant since in determining whether a facility is truly essential, courts will look to whether it is duplicable.[228]

Thus, this Note frames the first two EFD elements in a slightly different way than the *MCI* court. In order for the EFD to apply, (1) the facility at issue must be essential and the facility owner must control access to it, (2) the facility owner must be in competition with the firm seeking access, (3) the facility

223. 708 F.2d 1081 (7th Cir. 1983). In this case, MCI, a new entrant to the telecommunications business, charged that AT&T, the incumbent, did not provide the necessary network interconnection required by the Communications Act. The Seventh Circuit upheld the jury's finding of liability under the essential facilities doctrine. *Id.* at 1133.

224. *See, e.g.*, Frischmann, *supra* note 85, at 6 (describing *MCI* as the modern version of the EFD).

225. *MCI*, 708 F.2d at 1132–33.

226. Marquardt & Leddy, *supra* note 189, at 850.

227. *See id.*

228. *See, e.g.*, *MCI*, 708 F.2d at 1133 ("The facilities in question met the criteria of 'essential facilities' in that MCI could not duplicate Bell's local facilities.").

owner must have denied access to the essential facility, and (4) it must be feasible for the facility owner to provide access.

1. Qualcomm's SEPs are essential facilities to which Qualcomm controls access.

In the telecommunications context of *Qualcomm*, dominant standards like CDMA and LTE are essential components of cellular technology. No chipmaker would be able to sell a chip that was not compatible with a standard. Though there are two standards that are 2G- and 3G-compliant (CDMA and GSM),[229] chips must generally comply with both (in addition to LTE in a 4G chip), since the standards build upon one another in terms of functionality.[230] Although not all SEPs are truly essential to their standard, Qualcomm's SEPs cover truly essential CDMA and LTE technologies.[231] Since complying with CDMA and LTE standards is essential for anyone competing in the downstream modem chip market, "implementers have no choice but to use" Qualcomm's "patented technology."[232] Qualcomm's SEPs are therefore essential facilities.

In addition, Qualcomm alone controls access to its SEPs. Unlike in *Trinko* or *Otter Tail*, no federal or state agency has the power to force Qualcomm to license its patents.[233] Only Qualcomm can choose whether and on what terms to provide or deny access to its property. Though granting a chipmaker access to a patent looks different than granting a train access to a station, the end result is the same: without access, one cannot make productive use of the essential facility.

2. Qualcomm competes with those who seek access to the essential facility.

Here, both the district court and Ninth Circuit found that Qualcomm is not only competitive in chips, but also holds a monopoly in two downstream modem chip markets, 3G CDMA and 4G LTE. Like Otter Tail Power, Qualcomm dominates two pieces of the market and has every incentive to leverage its upstream power to curtail impending downstream competition.[234] Where this anticompetitive incentive exists, essential facilities analysis is appropriate.

---

229. Complaint, *supra* note 3, at 6.

230. *Id.* at 7–8.

231. On a brief review of the record, it does not appear that the essentiality of Qualcomm's SEPs was disputed.

232. *See* Melamed & Shapiro, *supra* note 2, at 2113.

233. SSOs, unlike federal agencies, do not have the power to compel Qualcomm to grant patent licenses. SSOs are private parties and can only bring contract litigation to get a court to compel Qualcomm to license its patents.

234. *See* Otter Tail Power Co. v. United States, 410 U.S. 366, 383 (1973).

Were Qualcomm to completely deny rival chip manufacturers access to its SEPs, downstream chip competition would be eliminated. This is because, without Qualcomm SEPs, rivals cannot legally make standard-compliant chips. If only Qualcomm's chips were standard-compliant, the competitive market for chips would be eliminated and Qualcomm would become the sole chipmaker in dominant telecommunications markets.[235]

3. Qualcomm denied its rivals access to its SEPs.

Qualcomm's conduct also meets the third element of the EFD. Qualcomm has historically denied its rivals access to its SEP licenses. Even though it allowed rivals to practice the technology, this "access" was access on unreasonable terms or an effective denial of access.[236] Moreover, Qualcomm's practices explicitly went against its FRAND commitments, and these commitments should be used to define "access" in this instance.

Though the CDMA ASIC Agreements allowed Qualcomm's rivals to practice the technology covered by its patents, the rivals were still denied access to the SEP licenses themselves. The CDMA ASIC Agreements cut off rivals' potential customer bases, restricting their ability to successfully make use of the technology—the ability to practice the patents, then, was not really access at all, but access on unreasonable terms or an effective denial of access.[237]

Rivals had no choice but to accept Qualcomm's terms.[238] If a rival rejected the CDMA ASIC Agreement, it would open itself up to patent infringement liability or be forced to make worthless products that did not comply with standards. Rival chipmakers had little negotiating power, if any, since they required Qualcomm's SEP technologies to manufacture standard-compliant chips. Accordingly, Qualcomm was able to extract significant concessions from rivals, in the form of data reporting requirements and limits on rivals' customer bases to licensed OEMs. Because of the "onerous" reporting requirements of the CDMA ASIC Agreements, Qualcomm gained sensitive business information about rival chipmakers' sales and contracts with

235. *Cf.* Alaska Airlines v. United Airlines, 948 F.2d 536, 545 (9th Cir. 1991) (airline control of the computer system did not carry with it the ability to eliminate downstream competition, in part because other computer systems existed).

236. *See* Delaware & Hudson Ry. Co. v. Consol. Rail Corp., 902 F.2d 174, 179–80 (2nd Cir. 1990) ("[T]here need not be an outright refusal to deal in order to find that denial of an essential facility occurred. It is sufficient if the terms of the offer to deal are unreasonable.").

237. *See id.*

238. *See* Melamed & Shapiro, *supra* note 2, at 2123–24 (noting the "pressure on implementers to accept offers that they regard as noncompliant [with FRAND], for fear that a court will reach a different conclusion at a later time and expose them to patent holdup").

OEMs.[239] Moreover, "[w]ithout a license . . . a rival [could not] sell modem chips with any assurance that Qualcomm [would] not sue the rival and its customers for patent infringement."[240] Thus, Qualcomm denying its rivals access to its SEP licenses hampered rival ability to gain traction in the modem chip market, to innovate in chips, or to invest in the development of next-generation standards that might compete with Qualcomm's. In short, the CDMA ASIC Agreements did not provide adequate SEP access to rivals and thus did not allow rivals to compete on a level playing field.[241]

Qualcomm's FRAND commitments should be used to help define "access." Industry terms like FRAND, voluntarily accepted, can and should factor into what "access" means in a given EFD case.[242] Where FRAND commitments define "reasonable access" as providing a license, as in *Qualcomm*, that definition should create a strong presumption that refusing to provide a license was an unreasonable form of access for purposes of the essential facilities inquiry. To ignore the presence of FRAND commitments would be to ignore the competitive realities of Qualcomm's situation, as the Ninth Circuit did.[243]

4. *Qualcomm could feasibly have granted access to its SEPs.*

Qualcomm could feasibly have granted rival chipmakers reasonable access (licenses) to its SEPs, and it did not have legitimate justifications supporting its denial of access. Though Qualcomm might have made more money licensing only to OEMs, pure profit-maximizing motives do not defeat an

239. Fed. Trade Comm'n v. Qualcomm, 411 F. Supp. 3d 658, 745 (N.D. Cal. 2019).

240. *Id.* at 744; *see also* Contreras, Amicus Brief, *supra* note 9, at 12 ("[W]ith no license from Qualcomm, the suppliers of chips that embody Qualcomm's patented technology remain vulnerable to suit, subject only to Qualcomm's unilateral discretion not to sue them.").

241. *See* United States v. Am. Tel. and Tel. Co., 524 F. Supp. 1336, 1352–53 (D.D.C. 1981) ("[A]ccess must be afforded 'upon such just and reasonable terms and regulations as will, in respect of use, character, and cost of services, place every such company upon as nearly as equal plane as may be.'" (citing United States v. Terminal R.R. Ass'n of St. Louis, 224 U.S. 383, 411 (1912))).

242. *See* Hovenkamp, *FRAND*, *supra* note 55, at 1684 ("Antitrust best achieves its purpose when it takes markets as it finds them, and then protects them from threats to competition."); *id.* at 1695 (describing SSOs and FRAND commitments as joint ventures in which "the purpose of the antitrust laws is not to destroy the venture or undermine its purpose, but rather to evaluate how the challenged restraint operates within the venture").

243. *See* Moss, *supra* note 20 (arguing that the Ninth Circuit's framing of Qualcomm's duty to license "requires ignoring the promises Qualcomm made to license its SEPs on reasonable and non-discriminatory terms—promises that courts in this country and around the world have consistently enforced").

essential facilities claim.[244] This should not change in the patent royalty context.[245] In addition, as the district court found, avoiding patent exhaustion was not an adequate justification for Qualcomm's refusal to license.[246] Were it to license to rivals, Qualcomm could still charge significant (even monopoly) patent royalties to rivals, making a profit and maintaining its market power. As in *Otter Tail*, where the court rejected Otter Tail's argument that its refusal to deal was justified as a mechanism to maintain the strength of its business, Qualcomm's motive here should not be given much weight.[247] It cannot claim that licensing to rivals would prevent it from effectively serving its customers or be impractical.[248] Thus, granting access to rivals was feasible for Qualcomm, and it did not have a sufficient justification for denying access.

In conclusion, Qualcomm could have and should have been found liable under the EFD for refusing to license its SEPs to rivals. This would have imposed upon Qualcomm a duty to license its rivals on reasonable and nondiscriminatory terms.

D. ADVANTAGES OF USING THE EFD APPROACH

Using the EFD approach rather than the three-part burden-shifting framework has both conceptual and policy advantages.

1. *Conceptual Advantages*

Using the essential facilities doctrine in *Qualcomm* makes for a simpler analysis than the burden-shifting anticompetitive conduct approach used by the district court and the Ninth Circuit. Rather than a roundabout inquiry into what is anticompetitive and who needs to be harmed for antitrust to kick in (Competitors? Consumers? Competition? But only in the relevant market?), the essential facilities doctrine consists of a series of clear factual inquiries (in "plain English"[249]), each one building on the last. The inquiry stops if the

244. *See, e.g.*, Hecht v. Pro-Football, Inc., 570 F.2d 982, 992–93 (D.C. Cir. 1977) (an essential facilities owner can deny access where such sharing would "inhibit the defendant's ability to serve its customers adequately").

245. *See* Brief of *Amicus Curiae* Mediatek Inc. in Support of Appellee and Affirmance at 16, Fed. Trade Comm'n v. Qualcomm, 969 F.3d 974 (9th Cir. 2020) (No. 19-16122) ("Immunizing unlawful conduct because it is profitable as well as exclusionary would be perverse."); Melamed & Shapiro, *supra* note 2, at 2121 ("The patent laws are intended to limit, not maximize, the royalties to which patent holders are entitled.").

246. *See* Moss, *supra* note 20 (noting that patent exhaustion is not optional, though the Ninth Circuit's decision made it out to be).

247. *See* Otter Tail Power Co. v. United States, 410 U.S. 366, 380 (1973).

248. *See* United States v. Am. Tel. & Tel. Co., 524 F. Supp. 1336, 1360 (D.D.C. 1981) (noting that antitrust law does not require an essential facility to be shared if sharing would be impractical or would inhibit the facility owner's ability to serve its customers).

249. Aerotec Int'l, Inc. v. Honeywell Int'l, Inc., 836 F.3d 1171, 1184 (9th Cir. 2016).

answer to any of those questions is no. Instead of hundreds of pages of discussion, the district court could have decided the issue quickly and concisely, without delving into the relevant market analysis, the three-part burden-shifting test, or anticompetitive effects.

The EFD also has substantive advantages as applied to this case, in that it adequately accounts for Qualcomm's dual monopolies—in SEPs and in chips. Though the district court understood that Qualcomm's refusal to deal worked in tandem with its NLNC policy, its refusal to deal analysis suffered because it focused only on the modem chip market. Had the district court imposed, and the Ninth Circuit upheld, a duty to deal under the EFD, Qualcomm's NLNC policy would have naturally been abandoned, at least with regard to those OEMs that purchased rival chips.[250]

One hurdle for the EFD to overcome is the impression that the Supreme Court has dealt the doctrine "death by dicta."[251] This dicta primarily comes from *Trinko*, in which the Court briefly stated that its conclusions about Verizon's liability "would be unchanged even if [it] considered to be established law the 'essential facilities' doctrine crafted by some lower courts."[252] It further noted that the "existence of sharing duties under the 1996 [Telecommunications] Act . . . [made] it unnecessary to impose a judicial doctrine of forced access."[253] The impact of this dicta on the EFD has generated significant scholarly attention.[254] There is a general consensus that,

250. In a world where Qualcomm licensed its SEPs to rivals, Qualcomm could not force OEMs to pay royalties or enter into licensing agreements for smartphones using rival chips because of patent exhaustion. It could still pursue NLNC against OEMs who bought Qualcomm chips (which would likely be most, if not all, because of Qualcomm's chip dominance) but these OEMs would have a stronger bargaining chip (pun intended) against Qualcomm. They could possibly negotiate for lower royalties or sue under patent law to determine whether the royalty rates were reasonable since they would have the option of buying rival chips without a royalty. This is of course a speculative argument. More research would need to be done on the operation of the NLNC policy as a tying arrangement and how Qualcomm might still be able to use it to leverage its chip power into licensing royalties, even if it was required to license to rivals. For an interesting look at this issue, see Hovenkamp & Simcoe, *supra* note 113, at 3–6 (discussing the NLNC policy as a tying arrangement in which licenses are the tying good and chips are the tied good).

251. *See* Frischmann, *supra* note 85, at 3. This precedent is one possible reason the FTC did not utilize the EFD, despite the fact that the Ninth Circuit has endorsed the EFD post-*Trinko*.

252. Verizon Commc'ns Inc. v. Law Offices of Curtis V. Trinko, 540 U.S. 398, 410 (2004).

253. *Id.* at 411. This conclusion largely seems to rest on arguments and assumptions about the administrability of the EFD, since conceptually there is no reason not to have two legal schemes mandating access. Whether *Trinko* was rightly or wrongly decided, however, is beyond the scope of this Note.

254. *See, e.g.*, Frischmann, *supra* note 85, at 24 ("*Trinko* has it precisely backwards in its views on the relationship between the essential facilities doctrine and regulation."); Weber

to the extent the EFD survives *Trinko*, it applies only where there is no regulatory agency supervising the industry.[255] Thus, even if *Trinko* limits the EFD overall, it does not limit its application in the SEP context, where there is no regulatory oversight of SSOs and FRAND commitments.

2. *Policy Advantages*

Courts should apply the EFD to SEPs because (1) antitrust law, more than patent or contract law, is best suited to address the patent holdup issues in the SEP context and (2) the EFD, more than the three-part burden-shifting framework, is best suited to balance the antitrust and patent law concerns.

Antitrust law is the appropriate forum for addressing SEP disputes like those at issue in *Qualcomm*—it need not be left to contract and patent law, as the Ninth Circuit suggested. As discussed by the district court and several *amici*, because of Qualcomm's dual monopolies, rivals and OEMs had little incentive to challenge Qualcomm's practices in contract or in patent law, where the damages are lower and the price of litigation might be a patent infringement suit or the loss of chip supply.[256] Contract and patent law are thus not "sufficient to guard against patent holdup" in the SEP context, though they may be important tools factoring into the overall analysis.[257] Moreover, the fact that Qualcomm's conduct *could* be addressed in contract or in patent law does not mean it should be immunized from antitrust liability.[258]

In addition, the EFD appropriately balances antitrust law's concern for competition and patent law's concern for innovation in the SEP context. The complex structure of SSOs, SEPs, and FRAND commitments is an attempt to strike a balance between the severe competitive risks of SEPs and the innovative benefits of standards. An SSO's establishment of an SEP right carries the risk that the SEP-holder will use its grant of market power to engage in holdup or other harmful practices. SSOs combat those risks not by doing away with beneficial standards, but by imposing FRAND commitments.[259]

Waller, *supra* note 95, at 159; Hovenkamp, *FRAND*, *supra* note 55, at 1718–19 (distinguishing *Trinko* and *Otter Tail* on the basis of EFD).

255. *See, e.g.*, HOVENKAMP, PRINCIPLES OF ANTITRUST, *supra* note 53, at 284.

256. *See* Fed. Trade Comm'n v. Qualcomm, 411 F. Supp. 3d 658, 786 (N.D. Cal. 2019; *see also* Class Plaintiffs' Brief as *Amicus Curiae* in Support of the Federal Trade Commission's Petition for Rehearing *En Banc* at 6, Fed. Trade Comm'n v. Qualcomm, 969 F.3d 974 (9th Cir. 2020) (No. 19-16122) (noting that some OEMs were forced to sign away their litigation rights).

257. *See* Melamed & Shapiro, *supra* note 2, at 2123.

258. *See* Hovenkamp, *FRAND*, *supra* note 55, at 1727 ("[N]o principle calls for antitrust deference to a private contractual regime.").

259. *See id.* at 1703 ("The FRAND process ensures that all participating firms have access to a common technology so that they can operate on a network where interconnection is

The EFD is the antitrust law complement to those FRAND commitments—like FRAND, it simply requires reasonable and nondiscriminatory access. When FRAND commitments fail to protect against SEP holdup, the EFD steps in as a necessary (but appropriately measured) check in the SEP context. Where patent rights are at their strongest, and the potential for abuse at its highest, antitrust can use the essential facilities doctrine to provide balance in favor of granting access. Antitrust law should not ignore the reality that, especially in a world where patented technology dominates economies and is essential to consumers, the behavior of powerful patent holders can have widespread impacts on competition, consumers, and the economy as a whole.[260]

V. CONCLUSION

Qualcomm is over, but the issues it raises are not going anywhere. With the impending rise of self-driving cars and other "smart" inventions, interoperability will continue to be a paramount concern for technology developers. In other words, communications standards are here to stay. This is ultimately a good thing. Technological standards make our lives as consumers easier—they allow us to plug our devices in to any outlet in the country and to call our grandmother's flip phone from an iPhone across the country. While any one standard might fade with time (e.g., 1G), standard-setting and standards in general are likely fixed features of our dynamic technological landscape. Where standards are accessible to all, competition and innovation thrive.

The essential facilities doctrine plays a key role in protecting this utopic vision of standards because it ensures access, without which innovation would be stifled. SEP holders, when they compete in markets that implement standards technology, have powerful incentives to deny reasonable access to that technology. When they succumb to those incentives, antitrust law must come in to protect competition in standard-implementing industries. Just as patent law seeks to balance exclusion and access, antitrust, too, must balance the patent right to exclude against the competitive benefits of access. In the standards context, the benefits of access generally far outweigh the benefits of exclusion. The essential facilities doctrine, as applied to SEPs, properly

essential. As a result, foreclosure can be much more harmful in a networked industry than elsewhere.").

260. *See id.* at 1685 ("Other firms will very likely follow Qualcomm's lead. If that happens the FRAND system will fall apart, doing irreparable injury to the modern wireless telecommunications network or, at the very least, diminishing the leadership role of the United States in preserving effective network competition.").

balances these concerns. Through the essential facilities doctrine, antitrust law can provide a limited but vital check on patent rights in standardized industries, one that will no doubt be necessary in the years to come.

UNCLOUDING THE CRYSTAL BALL: HOW TO DEMYSTIFY AND REFOCUS ANTITRUST LAW FOR DYNAMIC MARKETS

Aditi Ghatlia†

I. INTRODUCTION

The Fourth Industrial Revolution[1] has brought about a rise in dynamic and complex markets, defined by frequent and disruptive innovation.[2] These markets—such as telecommunications, augmented reality, and artificial intelligence—are becoming essential to and integrated into society. While dynamic markets benefit consumers through increasingly innovative products and services, they pose unique challenges to the existing framework of American antitrust law.

From its inception in the late 1800s, antitrust law in the United States has sought to balance the positive and negative effects of market consolidation to protect market competition.[3] One opportunity for regulators to monitor consolidation is through the review of horizontal mergers. But horizontal merger analysis—which examines a proposed union of two market

---

DOI: https://doi.org/10.15779/Z383J3923X

† J.D., 2022, University of California, Berkeley School of Law. I would like to acknowledge and give thanks to Professor Talha Syed for sharing his insight and expertise, as well as my parents, Naresh and Priti Ghatlia for their support and encouragement throughout this process.

1. *See* Klaus Schwab, *The Fourth Industrial Revolution*, BRITANNICA, https://www.britannica.com/topic/The-Fourth-Industrial-Revolution-2119734 (last visited Nov. 9, 2020) [https://perma.cc/3JLJ-RRE4]. This Fourth Industrial Revolution will include the "convergence of digital, biological, and physical innovations" and will rapidly change the products and services offered to consumers. These markets include science, artificial intelligence, telecommunications, and augmented reality. *Id.*

2. *See* Gregory Sidak & David Teece, *Dynamic Competition in Antitrust Law*, 5 J. COMP. L. & ECON. 581, 581 (2009); Gregory Sidak & David Teece, *Rewriting the Horizontal Merger Guidelines in the Name of Dynamic Competition*, 16 GEO. MASON L. REV. 885, 885 (2009) ("[S]uch competition does more than bring price competition – it tends to overturn the existing order.").

3. *The Antitrust Laws*, FED. TRADE COMM'N, https://www.ftc.gov/tips-advice/competition-guidance/guide-antitrust-laws/antitrust-laws [https://perma.cc/C5XB-QVME] (last visited Sept. 12, 2020); David S. Evans & Keith N. Hylton, *The Lawful Acquisition and Exercise of Monopoly Power and its Implications for the Objectives of Antitrust*, 4 COMPETITION POL'Y INT'L 203, 206 (2008).

competitors—is a daunting task. Courts are asked to do the near-impossible: ascertain the effects of a *potential* merger on a market and determine if those effects will be anticompetitive.[4] In essence, courts take on the role of "fortuneteller[s]" and must "peer[] into a crystal ball" to predict the future.[5]

Analyzing the impact of a merger on an entire market could easily lead judges down an unadministrable rabbit hole. There are hundreds of potential effects and several different market players, including consumers, producers, and state and federal governments. Each player has their own objectives within a market, and these objectives may conflict. A court facing this rabbit hole must narrow its analysis if it hopes to reach a decision. But this raises difficult questions: Which effects matter? Whose interests should be protected, and how?

Courts and regulators face added difficulties when addressing these questions in dynamic markets. Unlike non-dynamic (static) markets, where the overarching structure of the market remains stable from year to year, dynamic markets are consistently upset by rapid technological innovations.[6] For example, the milk market is static: while there might be new entrants, such as producers of almond and oat milk, the overall structures in which milk is made, bought, and sold have not significantly changed in the last hundred years. In contrast, the telecommunications market, a dynamic market, has seen constant, dramatic shifts in the underlying technologies, deployment, and types of services provided in the last thirty years, and it has become essential to consumers' everyday lives.[7] The unpredictable nature of dynamic markets adds difficulty to the jobs of regulators in protecting the market and its players.[8]

The predominant Chicago School framework of antitrust has narrow answers to the questions of which effects and whose interests the law should consider: it focuses only on short-term economic market effects, including price and innovation, and their impact on consumers.[9] Practitioners and scholars have dubbed this focus as "consumer welfare."[10] In a horizontal merger analysis, the consumer welfare approach directs courts and regulators to assess the potential anticompetitive effects of a merger, as well as its

4. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 187–89 (S.D.N.Y. 2020).

5. *Id.* at 186, 188.

6. Gregory Sidak & David Teece, *Dynamic Competition in Antitrust Law*, 5 J. COMP. L. & ECON. 581, 581 (2009).

7. *Deutsche Telekom*, 439 F. Supp. 3d at 189–190.

8. *See id.*

9. *See generally* Douglas H. Ginsburg, *Bork's "Legislative Intent" and the Courts*, 79 ANTITRUST L.J. 941 (2014).

10. *Id.* at 942–43.

potential efficiencies.[11] Generally, consumer welfare increases when efficiencies in the market allow producers and sellers to offer better products and services at cheaper prices, and when these efficiencies outweigh any reduction in competition.[12]

However, the Chicago School framework is ill-equipped to predict the effects of mergers in dynamic markets, which are increasingly commonplace in the twenty-first century. The recent case *New York v. Deutsche Telekom AG*,[13] litigated in the Southern District of New York, shows why. In this case, the court found that the proposed merger of two telecommunications services, Sprint and T-Mobile, would not lead to anticompetitive effects for market consumers and thus did not violate antitrust law. Adhering in large part to the Chicago School framework, the court concluded that the short-term efficiencies and increases in network abilities outweighed any potential harms to competition from the merger. This Note argues that two errors in *Deutsche Telekom* represent deficiencies in modern antitrust law that must be corrected to accurately regulate mergers in dynamic markets like the telecommunications industry. Namely, merger analysis in these markets should account for both the effects of the merger in the long term, and on the economy as a whole.

Part II of this Note discusses the basics of antitrust law and provides a primer on horizontal mergers. Part III presents an overview of *Deutsche Telekom* and highlights evidence used by the court in its efforts to predict the future of the telecommunications market. Part IV analyzes the merits of the court's decision in two sections. First, it discusses three errors made in the case. The court: (1) focused only on short-term efficiencies, (2) primarily studied consumer effects, ignoring overall effects on the economy, and (3) incorrectly prioritized witness testimony over documentation. Second, it examines how the first two of these errors reflect broader inadequacies in antitrust law. To accurately predict the future of dynamic markets, courts must broaden their efficiency analyses to account for the long-term effects of a merger and expand their goals to examine effects on the overall economy. In broadening their efficiency analyses, courts should balance potential short-term gains in efficiency with the long-term harms of consolidation of capital and power in dynamic markets, which can lead to higher prices and less innovation. Courts focusing on broader economic effects will better protect consumer, business, and government interests, as well as economic success. Both regulators and

11. U.S. DEP'T OF JUSTICE & FED. TRADE COMM'N, HORIZONTAL MERGER GUIDELINES 4 (2010).

12. Herbert Hovenkamp, *Implementing Antitrust's Welfare Goals*, 81 FORDHAM L. REV. 2471, 2471 (2013).

13. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 187–89 (S.D.N.Y. 2020).

judges should leverage the deficiencies in this case and the renewed political focus on antitrust law today to promote larger structural antitrust reform.

II. BACKGROUND

Reviewing antitrust law in the United States illuminates the policies and practices at play in *Deutsche Telekom*. This Part provides relevant background on antitrust law in the United States. First, it presents a brief history of the laws that form the cornerstone of antitrust enforcement. Second, it summarizes the prevailing analytical framework of antitrust law, the Chicago School, which heavily emphasizes consumer welfare and which the court relied upon in *Deutsche Telekom*. Third, it provides an overview of the horizontal merger doctrine and reviews courts' rationales in deciding previous merger cases. This background will help inform the policies and practices at play in *Deutsche Telekom*.

A. A BRIEF HISTORY OF ANTITRUST LAW

The late 1800s and early 1900s ushered in a consolidation of power in U.S. markets through the creation of monopolistic trusts.[14] The most famous example is Standard Oil, a trust owned by only one businessman that controlled almost all oil production and transportation in the country.[15] Standard Oil amassed its large market concentration through anticompetitive behaviors such as buying out other firms.[16] These practices gave Standard Oil great leverage in setting prices above the competitive level.

In response to the rise of trusts and their anticompetitive practices, Congress enacted the first major antitrust laws. Congress passed the Sherman Act in 1890 to curb attempted or successful monopolization and market collusion.[17] In 1914, Congress reinforced the Sherman Act with the passage of the Clayton Act.[18] Provisions of the Clayton Act advanced the Sherman Act by "prevent[ing] competitive conditions from deteriorating even when competition [is] not clearly problematic at the time."[19] Specifically, § 7 of the Clayton Act outlaws horizontal mergers that would "substantially . . . lessen

14. *See generally* John Morley, Essay, *The Common Law Corporation: The Power of the Trust in Anglo-American Business History*, 116 COLUM. L. REV. 2145 (2016).

15. *See generally* Standard Oil Co. of New Jersey v. United States, 221 U.S. 1 (1911).

16. *Id.*

17. FED. TRADE COMM'N, *supra* note 3; Evans & Hylton, *supra* note 3, at 217.

18. FED. TRADE COMM'N, *supra* note 3.

19. William J. Baer, Assistant Attorney General, Antitrust Division, Origin of the Species: The 100 Year Evolution of the Clayton Act, Remarks Before the ABA Clayton Act 100th Anniversary Symposium (Dec. 4, 2014), *in* 6 HEALTH CARE & ANTITRUST L., 2020, at App. E237.

competition, or . . . tend to create a monopoly."[20] Together, these laws enable courts to remedy anticompetitive conduct[21] and, and along with the FTC Act, are to this day the "core federal antitrust laws."[22]

Each act promotes the goal of antitrust: protecting "economic freedom and opportunity by promoting free and fair competition in the marketplace."[23] Free and fair competition results in lower prices, better quality, more quantity and choice, and greater innovation.[24] These benefits are achieved by balancing the positive and negative effects of market consolidation.[25] On the one hand, consolidation of power and limited competition can increase investment and innovation through economies of scale.[26] On the other hand, they can also lead to monopolistic behaviors that increase prices and reduce supply.[27] While monopolies themselves are not inherently unlawful, a monopoly's behavior can be unlawful if it has undue anticompetitive market effects. Therefore, antitrust law depends heavily on the distinction between pro-competitive and anticompetitive behavior to protect consumers.[28]

B. FRAMEWORKS OF ANTITRUST LAW

This Section briefly summarizes the current conceptual framework for antitrust analysis and provides background to the later discussion of its deficiencies in dynamic markets.

The predominant framework for antitrust analysis is the Chicago School framework. Named after the various lawyers and economists at the University of Chicago, the Chicago School's approach to the goals of antitrust is purely economic.[29] Courts and regulators ignore potential political and moral implications of mergers (or horizontal agreements, alleged monopolization, etc.); instead, they assess only the economic harms and gains in a market, such

---

20. Clayton Act § 7, 15 U.S.C. § 18.

21. Most notably, in 1911 the Supreme Court ordered the split of Standard Oil into various smaller companies that would compete against each other. *See* Standard Oil Co. of New Jersey v. United States, 221 U.S. 1 (1911).

22. FED. TRADE COMM'N, *supra* note 3.

23. *Mission*, U.S. DEP'T OF JUSTICE, https://www.justice.gov/atr/mission [https://perma.cc/K3SC-3CSZ] (last visited Nov. 1, 2020).

24. *Id.*

25. *Id.*

26. *See* Hovenkamp, *supra* note 12, at 2486. Selling a larger quantity of goods usually leads to a reduction in price of each good, a principle called economies of scale. *See* Economies of Scale, CAMBRIDGE DICTIONARY, https://dictionary.cambridge.org/us/dictionary/english/economies-of-scale (last visited Apr. 3, 2021).

27. *See* Evans & Hylton, *supra* note 3, at 203.

28. *Id.* at 206.

29. *See* Bruce Kobayashi & Timothy Muris, *Chicago, Post-Chicago, and Beyond: Time to Let Go of the 20th Century*, 78 ANTITRUST L.J. 147, 149–152 (2012).

as higher prices or higher-quality products.[30] This reliance on economics has "rendered antitrust a more mathematically rigorous and technically demanding field" than other areas of law.[31]

Economic analysis in the Chicago School focuses on "consumer welfare," a term made famous by Judge Robert Bork.[32] While scholars disagree on the exact definition of consumer welfare, courts in merger cases often conduct consumer welfare analysis by measuring the short-term effects of a merger on consumers in the aggregate, examining whether they will receive better prices, quality, quantity, and innovation from a merger.[33] One of the most important ways that antitrust regulators assess the effects of a merger on consumer welfare is by examining the potential realization of efficiencies,[34] which can positively impact consumers in a market.[35] Economic efficiency gains occur when a merger allows the merged firm to compete more effectively—for instance, by consolidating resources or combining human capital—resulting in lower prices and higher quality goods.[36] But efficiency analysis is never exact, as determining potential efficiencies is a forward-looking exercise and judges

---

30. *See* Thomas J. Horton, *Rediscovering Antitrust's Lost Values*, 16 U.N.H. L. REV. 179, 180–181 (2018).

31. *See* Michael Baye & Joshua Wright, *Is Antitrust Too Complicated for Generalist Judges? The Impact of Economic Complexity and Judicial Training on Appeals*, 54 J. L. & ECON. 1, 2 (2011).

32. *See* Hovenkamp, *supra* note 12, at 2472.

33. *See id.* Economists generally have two market concerns. First, there is a concern that some consumers will be priced out of the market because they will not pay higher prices for the goods. This loss is called the deadweight loss. *See* Herbert Hovenkamp, *Antitrust's Protected Classes*, 88 MICH. L. REV. 1, 6 (1989). The second concern is while the total surplus (or the total of both consumer and producer welfare) might remain the same or even grow in monopolistic markets, producer surplus will grow more than consumer surplus. This imbalance in surplus is caused by producer profits from selling higher cost products to remaining consumers. In general, economists look for growth in overall total surplus as this number indicates the growing wealth of an economy. *See id.* At 6. The Chicago School's misleading emphasis on consumer welfare refers only to total surplus, and not to specific consumer surplus. *See* Ginsburg, *supra* note 9, at 942–43. In championing a focus on total surplus, a Chicago view analysis of antitrust only considers the first concern of deadweight loss caused by lower output and higher prices in the market. *See* Hovenkamp, *supra* note 12, at 2473. Chicago School followers do not think it is relevant to discern producer profits from consumer profits, as this wealth shift "had no discernible net impact on consumer welfare" since both producers and consumers can adjust their spending based on their preferences. Daniel A. Crane, *The Tempting of Antitrust: Robert Bork and the Goals of Antitrust Policy*, 79 ANTITRUST L.J. 835, 846 (2014).

34. U.S. DEP'T OF JUSTICE & FED. TRADE COMM'N, *supra* note 11, at 4.

35. *See* John B. Kirkwood & Robert Lande, *The Fundamental Goal of Antitrust: Protecting Consumers, Not Increasing Efficiency*, 84 NOTRE DAME L. REV. 191, 193–95 (2008).

36. *See* DEP'T OF JUSTICE & FED. TRADE COMM'N, *supra* note 11, at 4; Lina Khan, *Amazon's Antitrust Paradox*, YALE L. J. 710, 716 (2017).

must make predictions based only on information available to them in the present.

While the Chicago School remains the dominant antitrust framework, a new approach to antitrust, the Neo-Brandeisian School, has gained popularity.[37] Neo-Brandeisians criticize the Chicago School for permitting companies to consolidate capital and power so long as prices remain low,[38] arguing that this consolidation can create a cyclical effect of propping up powerful individuals at the expense of the working class.[39] In contrast, Neo-Brandeisians want antitrust law to focus on the welfare of all individuals and the general society, not just direct consumers.[40] To that end, Neo-Brandeisians argue that instead of focusing solely on short-term economic effects, antitrust regulators should consider the aggregation of political power stemming from concentration, the protection of small businesses, and other fairness concerns unrelated to direct economic indicators.[41] Further, Neo-Brandeisians believe that the Chicago School is unequipped to deal with dynamic markets, many of which are structured in a way that traditional economics cannot fairly evaluate. For example, the well-known Neo-Brandeisian scholar Lina Khan has asserted that Amazon, a large dynamic player in the supply chain and home goods market, cannot be adequately regulated under the Chicago School framework because Amazon strives for low prices for direct consumers—the basic criteria for the Chicago School—at the expense of their workers and potential competitors in the internet sales market.[42]

C. HORIZONTAL MERGER ENFORCEMENT: THE SEMANTICS

While the New Brandeis School is gaining steam in academic circles, the Chicago School still dominates the current judicial regulation of mergers. This Section examines how courts evaluate horizontal mergers under the Chicago School's consumer welfare framework. This context is necessary to understand the analysis and critique of *Deutsche Telekom*, discussed in Parts III and IV.

---

37. *See* Jonathan Sallet, *Louis Brandeis: A Man for This Season*, 16 COLO. TECH. L.J. 365, 368 (2018); David Streitfeld, *Amazon's Antitrust Antagonist Has a Breakthrough Idea,* N.Y. TIMES, Sept. 7, 2018, https://www.nytimes.com/2018/09/07/technology/monopoly-antitrust-lina-khan-amazon.html [https://perma.cc/7GK7-PWDP].

38. *See* Stucke and Ezrachi, *The Rise, Fall, and Rebirth of the U.S. Antitrust Movement*, HARV. BUS. REV. (Dec. 15, 2017), https://hbr.org/2017/12/the-rise-fall-and-rebirth-of-the-u-s-antitrust-movement [https://perma.cc/LB4N-RKBC].

39. *See* Khan, *supra* note 36 at 740.

40. *See id.* at 720.

41. *See* Herbert Hovenkamp, *Is Antitrust's Consumer Welfare Principle Imperiled?*, 45 J. CORP. L. 101, 120 (2019).

42. *See generally* Khan, *supra* note 36.

Horizontal mergers refer to the combination of two potential or actual competitors in a market and are regulated under § 7 of the Clayton Act.[43] In a § 7 analysis, courts must establish a "reasonable probability," not just a possibility, of anticompetitive behavior by the merged entity to block a merger.[44] Courts must essentially predict the competitive effects of the merger, including economic efficiencies and gains or losses in consumer welfare. While merger analysis is quite case- and fact-specific, courts follow general guidelines and use a variety of tools and economic indicators to aid their predictions.[45]

To comport with the goals of antitrust law and allow each side a fair hearing, courts utilize a three-step burden-shifting test during horizontal merger analyses.[46] As a threshold step before applying this test, courts establish the relevant market in terms of geography and products to help determine where their focus should be.[47]

In the first step, plaintiffs must establish a prima facie case showing that the merger will significantly decrease competition through evidence of consolidation and lack of viable competitors in the defined market.[48] Plaintiffs can use one of two common economic methods to establish a presumption of anticompetitiveness:[49] they can show that the newly merged company will have greater than thirty percent of the market share, or use the Herfindahl-Hirschman Index (HHI) to indicate that the merged company's market share will reach a certain threshold that the Department of Justice (DOJ) and Federal Trade Commission (FTC) have deemed "presumptively anticompetitive."[50] Each method uses market concentration as a proxy for market power. This proxy assumes that as a firm gains market share, it will wield more control and

43. DEP'T OF JUSTICE & FED. TRADE COMM'N, *supra* note 11, at 1.

44. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 198 (S.D.N.Y. 2020).

45. *See id.*

46. Chi. Bridge & Iron Co. N.V. v. Fed. Trade Comm'n, 534 F.3d 410, 423 (5th Cir. 2008).

47. *See, e.g.*, *Deutsche Telekom*, 439 F. Supp. 3d at 189–90; *Brown Shoe Co. v. United States*, 370 U.S. 294, 336 (1962).

48. *See, e.g.*, *Deutsche Telekom*, 439 F. Supp. 3d at 189–90; *Brown Shoe Co.*, 370 U.S. at 321.

49. *See Deutsche Telekom*, 439 F. Supp. 3d at 199.

50. *Id.* at 206. The HHI is "calculated by adding the squares of [all] the individual firms' market shares" in the market. *Id.* Moderately concentrated markets have an HHI between 1500 and 2500, while highly concentrated markets have an HHI above 2500. *Id.* If a merger increases HHI by 200 points or more and is above 2500, the merger is presumed to be anticompetitive and plaintiffs will have met their burden of proof. DEP'T OF JUSTICE & FED. TRADE COMM'N, *supra* note 11, at 19. Hence, this is a fairly objective and economic parameter by which one can assess the potential level of concentration in the market.

can exercise its power through anticompetitive behaviors.[51] If met, the plaintiff's prima facie case creates a presumption that the merger will be anticompetitive and should be enjoined.[52]

In the second step, defendants may bring evidence rebutting this presumption and showing the court that the merger will, on balance, not have anticompetitive effects.[53] This evidence should "cast doubt on the accuracy" of the prima facie case.[54] Defendants can successfully rebut a plaintiff's case with different sources of evidence, including unique market circumstances that show the merger will not harm market competition.[55] Defendants can also present indicators showing ease of market entry—which increases the likelihood of competition—and evidence of market trends towards competition.[56]

Finally, in the third step, the burden shifts back to the plaintiff, who can bring additional evidence to the court, such as market trends or documents showing that the market would suffer from the merged firm's likely anticompetitive behaviors.[57] Courts will weigh all the evidence to make a final decision.[58]

D. A WINDOW INTO MERGER ANALYSIS

This Section highlights the various types of evidence that courts can examine within a horizontal merger analysis. Current antitrust law values economic measures of consumer welfare and uses efficiencies as a measure of this welfare. The cases below show how courts have used different efficiency measures, including past behavior, barriers to entry, and market circumstances, to predict merger effects on consumer welfare.[59]

In *FTC v. Heinz Co.*, the D.C. Circuit ruled that the merger of two baby food producers would be anticompetitive because of high market

---

51. Monopoly Power and Market Power in Antitrust Law, DEP'T OF JUSTICE (last updated June 25, 2015), https://www.justice.gov/atr/monopoly-power-and-market-power-antitrust-law [https://perma.cc/6978-L5GQ].

52. *Deutsche Telekom*, 439 F. Supp. 3d at 206.

53. Herbert Hovenkamp & Carl Shapiro, *Horizontal Mergers, Market Structure, and Burdens of Proof*, 127 YALE L.J. 1996, 1996 (2018).

54. Chi. Bridge & Iron Co. N.V. v. FTC, 534 F.3d 410, 423 (5th Cir. 2008).

55. FTC v. Univ. Health, Inc., 938 F.2d 1206, 1218 (11th Cir. 1991).

56. *Id.* One example of market trends towards competition would be more competitors entering the market in recent years.

57. *Chi. Bridge & Iron Co.*, 534 F.3d at 423.

58. *See id.*

59. Market barriers to entry are defined as "any characteristic that impedes, discourages, or delays, for a significant period of time, entry into a market." Horizontal mergers—Other factors—Rebuttal evidence—Entry, 1 HEALTH CARE & ANTITRUST L. § 6A:9 (2020).

concentration and minimal efficiency gains.[60] The FTC easily established a prima facie case using an HHI calculation indicating that the market share exceeded the threshold for presumptive anticompetitiveness.[61] Defendants attempted to rebut this presumption by arguing that there would be no significant competitive loss from the merger.[62] Specifically, defendants argued that consumers did not view the products as substitutes and that the merger would increase efficiencies.[63] The court noted that the large increase in HHI (an increase of 510 points to a total of 5,286) meant that it needed to complete a "rigorous analysis" to ensure that the efficiencies were not false promises and would actually offset the negative effects of market concentration.[64] Defendants argued that the merger would allow the companies to aggregate resources, use resources more efficiently, and compete aggressively against the market leader, all leading to a reduction in prices.[65] However, the court concluded that this argument was speculative and that high barriers to entry in the baby food market would only increase the anticompetitive effects of the merger, as fewer companies could join the market and compete.[66] After weighing the evidence, including documents from Heinz that discussed ending competition, the court ruled in favor of the FTC.[67]

Similarly, in *FTC v. University Health, Inc.*, the Eleventh Circuit ruled against the potential merger of two hospitals because of anticompetitive effects and the defendants' inability to prove concrete efficiencies in the market.[68] Defendants argued that the efficiencies of the merger would include streamlining capital and sharing software between the two hospitals.[69] While they estimated a dollar value the merger would save in operational costs, the

60. 246 F.3d 708, 7116–18 (D.C. Cir. 2001).

61. *Id.* at 715–16.

62. *Id.* at 718. Courts have commonly taken into account potential efficiencies from a merger as a successful way for defendants to rebut a presumption of anticompetitive behavior. *See, e.g.*, FTC v. Wilh. Wilhelmsen Holding ASA, 341 F. Supp. 3d 27, 73 (D.D.C. 2018); United States v. H&R Block, Inc., 833 F. Supp. 2d 36, 89 (D.D.C. 2011).

63. *Heinz Co.*, 246 F.3d at 718.

64. *Id.* at 721.

65. *Id.* at 722.

66. *Id.* at 722, 718.

67. *Id* at 717, 727. Other courts have also taken into consideration internal documents to establish the intent of market players when analyzing the potential effects of a merger. *See* United States v. Oracle Corp., 331 F. Supp. 2d 1098, 1166–68 (N.D. Cal. 2004) (reviewing documents where companies who were merging noted each other as "closest competitors"). Courts have also relied on barriers to entry; if barriers to entry are low, then the presumption of anticompetitive behavior is likely be rebutted as new players could easily enter the market. *H&R Block, Inc,*, 833 F. Supp. 2d at 73.

68. 938 F.2d 1206, 1222–24 (11th Cir. 1991).

69. *Id.* at 1223.

court was unconvinced, as defendants provided no specifics on how these efficiencies would be "passed on to consumers."[70] The defendants also argued that their status as non-profit hospitals and their past procompetitive behavior should persuade the court that they would not engage in anticompetitive behavior.[71] The court was not persuaded by this evidence, which it deemed not probative in value.[72] It declared that no business was bound by its prior actions, and therefore, the court would not rely on past behaviors to predict future behaviors.[73] The court ruled the merger to be anticompetitive.[74]

These two cases exemplify how courts rely on a variety of economic indicators to determine the potential short-term effects and market efficiencies of a merger. In each case, the court's ruling was ultimately determined by its prediction of the merger's effects on consumers and their welfare.

III. NEW YORK V. DEUTSCHE TELEKOM

In *Deutsche Telekom*, multiple states brought a lawsuit under the Clayton Act against the merger of two telecommunications providers, Sprint and T-Mobile, claiming that the merger would decrease market competition and lead to anticompetitive behavior.[75] This Part first provides background on the telecommunications market and then presents an overview of the case's facts, the court's holding, and the evidence considered.

A. THE TELECOMMUNICATIONS MARKET

The modern telecommunications services market (namely, mobile devices) is an integral part of society. What began as a simple "method of voice communication" is now a practical tool for individuals to "manage countless facets" of their lives, including banking, transportation, social media, and audio and visual streaming.[76] As of 2019, over ninety-six percent of Americans owned at least one cellphone, and this growth has facilitated a fast-paced and "on-the-go" lifestyle.[77] Retail mobile wireless telecommunications services (RMWTS, or telecommunications companies), provide the products and services to consumers in this market (cell phones and data plans).[78] Generally,

70. *Id.*
71. *Id.* at 1224.
72. *Id.*
73. *Id.*
74. *FTC v. University Health, Inc.*, 938 F.2d 1206, 1225–26 (11th Cir. 1991).
75. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179 (S.D.N.Y. 2020).
76. *Id.* at 189–90.
77. Pew Research Center, *Mobile Fact Sheet*, https://www.pewresearch.org/internet/fact-sheet/mobile/ [https://perma.cc/QM5C-L4HL] (last visited Oct. 4, 2020).
78. *Deutsche Telekom*, 439 F. Supp 3d at 190.

consumers choose their provider based on factors including price, options to bundle devices, and the speed and consistency of networks.[79]

Telecommunications companies provide service through spectrum, which is the range of radio waves used for the transmission of cellular data.[80] Spectrum is limited in quantity, highly regulated by the federal government, and sold to companies through auction by the Federal Communications Commission (FCC).[81] Both limited supply and high demand make spectrum "the most critical resource" for providers.[82]

The telecommunications market is constantly innovating, in part due to the ever-growing uses for spectrum (e.g., phone calls, data sharing, and artificial intelligence), making the telecommunications market dynamic.[83] Dynamic markets are defined by disruptive innovations that create constant churn and can "overturn the existing order" of a market.[84]

Disruptive innovation in the telecommunications market comes primarily from the generations of telecommunications technology, more commonly known as "G's."[85] G's improve the speed, capabilities, and efficiencies of networks with each iteration, and are being developed at an increasingly rapid rate.[86] For example, the first generation, developed in the 1980s, allowed for voice calls to be transmitted over spectrum.[87] 2G was developed a decade later and allowed for text messages, and 3G provided access to the internet in the 2000s.[88] 4G has enabled application development and high-speed data travel.[89] 5G rollout is predicted to bring even faster speeds and the advent of virtual

79. *Id.*
80. *Id.*
81. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 190 (S.D.N.Y. 2020).
82. *Id.*
83. *Id.* at 192.
84. Sidak & Teece, *supra* note 2, at 582–83; Pedro Gonzaga, *Merger Enforcement in Dynamic and Innovative Markets*, 4 EUR. COMPETITION & REG. L. REV., V, 6 (2020). While no product or service market is completely static, some markets are more structurally static than others. Markets such as the milk or shoe market, for example, are structurally static; while new products might be introduced into these markets, the way in which the market operates has not changed from decade to decade. Legislators and courts have been focused on antitrust law through the lens of static competition for the last fifty years through their economic heavy framework.; Sidak & Teece, *supra* note 2, at 581.
85. *Deutsche Telekom AG*, 439 F. Supp. 3d at 192.
86. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 192 (S.D.N.Y. 2020).
87. *Id.*
88. *Id.*
89. *Id.*

reality applications.[90] Current projections indicate that consumers will use almost five times as much data for 5G than is currently being used, incentivizing a successful 5G rollout for providers.[91] However, 5G will need its own exclusive and unshared spectrum, making the rollout more difficult and expensive than previous generations.[92] Providers will have to reshape their services to meet 5G standards.

The telecommunications market is not only dynamic but also capital-intensive. To deploy spectrum and send information to mobile devices, telecommunications providers require a robust infrastructure network.[93] This network includes physical cell towers that transmit radio waves, fiber cables connecting these structures, and a "core network" of computers to direct cell traffic.[94] A provider's coverage and capacity is determined by this network, referred to as Radio Access Network (RAN).[95] Strong networks will provide coverage regardless of geographical location but face data limitations when many consumers are using the network at once.[96]

There are currently two types of service providers in this market:[97] Mobile Network Operators (MNOs) and Mobile Virtual Network Operators (MVNOs).[98] MVNOs lease RAN access from MVOs.[99] Examples of MVNOs include Altice and TracFone.[100] MNOs own their own RAN.[101] The four MVOs in the United States—AT&T, Verizon, T-Mobile, and Sprint—are common household names, but they differ in size and capacity.[102] Verizon and AT&T are the largest of the MVOs; each provider has sizable spectrum holdings, close to 100 million subscribers, upwards of four billion dollars in yearly revenue, and reputations for high quality services.[103] T-Mobile, the third largest MVO, has grown rapidly in the past decade after an influx of cash from a previous merger proposition and a bold creative strategy from CEO John

90. *Id.* at 192–93. The 5G rollout will likely create new jobs. These jobs will include construction of the physical network infrastructure as well as development of new 5G applications and tools.

91. *See* New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 192 (S.D.N.Y. 2020).

92. *Id.* at 209.

93. *Id.*

94. *Id.* at 191.

95. *Id.* at 191, 191 n.4.

96. *See id.* at 191–92.

97. *Id.* at 193.

98. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 193 (S.D.N.Y. 2020).

99. *Id.*

100. *Id.* at 195. MVNOs make up less than two percent of the national market share. *Id.*

101. *Id.* at 193.

102. *See id* at 193–95.

103. *Id.* at 193.

Legere.[104] This strategy included successful investments in spectrum and marketing services at lower prices to contrast their products with those of AT&T and Verizon.[105] T-Mobile now counts approximately eighty million subscribers and three billion dollars in yearly revenue.[106] Sprint, the smallest MVO, has seen a decline in quality of service over the past ten years, due in part to unsuccessful investments in spectrum holdings.[107] Sprint serves forty million subscribers and records two billion in yearly revenue.[108]

B. THE DUETSCHE TELEKOM CASE: SETTING THE STAGE

In 2018, T-Mobile and Sprint announced that they planned to merge and would retain the name T-Mobile (this Note will refer to the merged company as New T-Mobile).[109] The companies filed the proposed merger with the DOJ and the FCC. After negotiations and concessions by Sprint and T-Mobile, both agencies approved the merger in 2019.[110] After this approval, ten state attorneys general, including those of New York and the District of Columbia (hereinafter "Plaintiffs"), filed a lawsuit in the Southern District of New York to prevent this consolidation.[111]

Plaintiffs claimed that the proposed merger of T-Mobile and Sprint (hereinafter "Defendants") violated § 7 of the Clayton Act because it would reduce competition in the RMWTS market and result in increased costs to consumers, estimated at 4.5 billion dollars a year.[112] However, after a lengthy analysis of the evidence, the court found that the merger would not lead to anticompetitive behavior in the market and allowed it to proceed.[113]

---

104. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 193–94 (S.D.N.Y. 2020).

105. *Id.* at 194.

106. *Id.* at 193.

107. *Id.* at 194.

108. *Id.*

109. Andrew Limbong, *T-Mobile Completes Takeover of Rival Company Sprint*, NPR (Apr. 1, 2020, 5:57pm ET), https://www.npr.org/2020/04/01/825523250/t-mobile-completes-takeover-of-rival-company-sprint [https://perma.cc/CY7C-DHV7].

110. Yuki Noguchi, *T-Mobile and Sprint Merger Finally Wins Justice Department's Blessing*, NPR (July 26, 2019, 11:30 AM ET), https://www.npr.org/2019/07/26/745544033/t-mobile-and-sprint-merger-finally-wins-justice-departments-blessing [https://perma.cc/VT7S-328F]. These concessions include agreements not to raise prices for three years and to deploy a working 5G network as quickly as possible. Merrit Kennedy, *10 State Attorneys General Sue to Block T-Mobile, Sprint Merger*, NPR (June 11, 2019, 2:15 PM ET), https://www.npr.org/2019/06/11/731695613/10-state-attorneys-general-sue-to-block-t-mobile-sprint-merger [https://perma.cc/V587-NTSY].

111. Kennedy, *supra* note 110.

112. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 186 (S.D.N.Y. 2020); Noguchi, *supra* note 117.

113. *Deutsche Telekom*, 439 F. Supp. 3d at 189.

Before discussing the case itself, Judge Marrero of the Southern District of New York provided his own thoughts on the difficulties of antitrust cases.[114] Judge Marrero wrote that "antitrust disputes virtually turn[] the judge into a fortuneteller" because judges must predict the potential effect of mergers on competition and on consumers.[115] Both parties spend millions of dollars on expert witnesses, economic research, and documentation, all of which speaks favorably to their position. However, in most cases the data for each side will "cancel each other out."[116] This cancellation leaves judges with nothing but "competing crystal balls."[117] Instead of juggling these competing interests, judges rely on their own "tried and tested version of peering into a crystal ball"—looking at past behavior, weighing the credibility of witnesses, and using experience to employ "behavioral measures" that numbers fail to establish.[118]

C. THE HORIZONTAL MERGER ANALYSIS

After conducting a complete horizontal merger analysis, Judge Marrero concluded that Defendants' evidence and witnesses successfully rebutted Plaintiffs' presumption of anticompetitive behavior and allowed the merger to move forward.[119]

As a threshold matter, the court established the relevant product market (products that consumers use interchangeably) to include only MVOs[120] (larger telecommunications companies that own their own infrastructure and can set their own rates).[121] Regarding the geographic market (where the product is marketed), the court found both a large national market and local markets, each using different marketing strategies.[122]

1. *Plaintiffs' Prima Facie Case*

As discussed *supra*, plaintiffs must first establish a prima facie presumption that the merger will "result in undue market concentration in an area of effective competition" in the product and geographic markets, thereby

114. *See id.* at 186–90.
115. *Id.* at 186.
116. *Id.* at 187.
117. *Id.*
118. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 188 (S.D.N.Y. 2020).
119. *Id.* at 189.
120. *Id.* at 199–202. The court did not include MVNOs because it found that MVNOs were quite restrained by the MVOs they lease from and had no price setting power; therefore, the market shares of MVNOs were attributed to the MVOs from which they leased. *Id.*
121. *Id.*
122. *Id.* at 203.

harming competition.[123] In *Deutsche Telekom*, both a calculation of overall market share and the HHI index suggested the merger would significantly increase market concentration: Plaintiffs showed that New T-Mobile would have over thirty percent of the market share and that the HHI would increase by 679 points, totaling 3,186.[124] The court found that this satisfied Plaintiffs' burden.[125]

2. Defendants' Rebuttal

The court then analyzed Defendants' evidence and found that it rebutted Plaintiffs' prima facie case.[126] Defendants brought three arguments at this stage: (1) the merger would bring efficiencies to the market through a new 5G network and lower costs, (2) Sprint was a weakened competitor, and (3) DISH Network's entrance as a new competitor, paired with additional government remedies, would increase competition in the market.[127]

a) Efficiencies

The court found that efficiencies from the merger would reduce prices and improve the quality of telecommunications services, thus increasing consumer welfare.[128] Defendants predicted that consumers would benefit from decreased prices because New T-Mobile would operate at reduced costs due to streamlined networks and infrastructure.[129] New T-Mobile would continue to offer aggressive rates to compete with Verizon and AT&T and siphon off those competitors' customers.[130] Further, New T-Mobile would run on a network consisting of only the strongest portions of both T-Mobile and Sprint's networks, enhancing the quality of customers' service.[131] Finally, New T-Mobile's increased spectrum capacity would also lead to improved product offerings and a faster 5G rollout.[132]

The court accepted all of Defendants' economic efficiency claims and noted that previous mergers in this market had led to greater efficiencies,

123. *Id.* at 199.

124. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 206 (S.D.N.Y. 2020).

125. *Id.*

126. *Id.* at 207.

127. *Id.*

128. *Id.* at 208. Previous courts, as well as the DOJ, have established that "evidence of efficiencies" in the market can rebut the plaintiffs' presumption of competition. *Id.* at 208–09.

129. *Id.* at 208.

130. *Id.*

131. *Id.*

132. *Id.* at 208–09. As discussed *infra*, the merger included stipulations mandated by the DOJ and FTC, including that New T-Mobile meet strict deadlines for rolling out 5G (with associated fines if deadlines were not met). *Id.* at 225.

thereby boosting the credibility of Defendants' arguments.[133] The court found that the predicted efficiencies were merger-specific, meaning that neither company by itself could achieve these results.[134] For instance, while it was theoretically possible for Sprint and T-Mobile to buy spectrum through FCC auctions, increasing their individual capacity, no auctions were scheduled for the near future.[135] Plaintiffs disagreed that the efficiencies were merger-specific. They argued that a new technology, Dynamic Spectrum Sharing, would allow providers to use the same spectrum for both 4G and 5G simultaneously.[136] Through Dynamic Spectrum Sharing, Sprint and T-Mobile could individually gain efficiencies in spectrum that would strengthen their 5G services, rendering the merger efficiencies unnecessary.[137] However, because this technology was still experimental, the court found that it was not reliable.[138]

The court also went further to acknowledge additional benefits of the merger beyond efficiencies. It wrote in a footnote that the speedy development of 5G in the United States provided a broader advantage to the public interest because it would allow the United States to remain a top innovator and out-compete other technologically advanced countries such as China and Korea.[139]

b) Sprint was a Weakened Competitor

Defendants successfully argued that Sprint was a weakened competitor and, if it continued as an individual business, it would not "compete effectively in the future."[140] The court discussed various reasons for this weakened status, including Sprint's neglect in network investment and its financial difficulties.[141] Because Sprint's decreasing viability in the telecommunications market meant that it could allegedly not continue as a competitive standalone firm, the court stated that Sprint's merger with T-Mobile would not hurt competition.[142] While there is specific case law and guidance on the viability of a weakened

133. *Id.* at 216–17.
134. *Id.* at 211.
135. *Id.*
136. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 212 (S.D.N.Y. 2020); *see also Breakthrough 5G data call using dynamic spectrum sharing to accelerate nationwide 5G deployments* (Sept. 4, 2019) https://www.ericsson.com/en/news/2019/9/ericsson-spectrum-sharing [https://perma.cc/KDM9-ZUF8] (providing more information on Dynamic Spectrum Sharing).
137. *Deutsche Telekom*, 439 F. Supp. 3d at 212.
138. *Id.*
139. *Id.* at 209 n.12.
140. *Id.* at 217.
141. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 219–20 (S.D.N.Y. 2020).
142. *Id.* at 189.

competitor defense, the implications of this defense are beyond the scope of this Note.[143]

c) DISH and Other Market Remedies

Defendants argued that the entry of DISH as a fourth competitor in the telecommunications market would mitigate any residual anticompetitive effects of the merger. Before this litigation, both the DOJ and the FCC reviewed the merger and added stipulations to reduce anticompetitive effects, including requirements for DISH to join the market.[144] These stipulations were considered part of the merger by the court.

Among other requirements, the stipulations mandated that New T-Mobile divest one of Sprint's MVNOs, Boost, to DISH.[145] DISH is currently a TV network provider, but it has a low and mid-band spectrum portfolio equivalent in size to Verizon and has previously indicated interest in entering the telecommunications market.[146] This large spectrum holding made DISH a strong candidate for entering the market,[147] which it would do by obtaining Boost and entering a licensing deal to use New T-Mobile's network while DISH started building out its own MVO network.[148] The court believed that DISH's entry as an MVO "maverick" to the market would be timely, likely, and sufficient to mitigate any anticompetitive effects of the merger as it would add new competition in the market.[149]

While DISH would face a capital-intensive process to build out a competitive network, the court believed that two technologies would help lower DISH's barriers to entry. First, Dish claimed it would cut costs by building their virtualized network on top of software and cloud services, rather than building their software on hardware that it builds from scratch.[150] Second, DISH planned to build an Open Radio Access Network (ORAN) that would enable use of multiple vendors' hardware and software, thereby reducing the

143. DEP'T OF JUSTICE & FED. TRADE COMM'N, *supra* note 11, at 32 ("Failure and Exiting Assets").

144. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 224 (S.D.N.Y. 2020).

145. *Id.* at 225. The court established that DISH's attainment of Boost, with its near-ten million customers and high satisfaction rates, would help the network's market entrance. *Id.* at 227.

146. DISH, https://www.dish.com/[https://perma.cc/Z292-YTKU]; *Deutsche Telekom*, 439 F. Supp. 3d at 224–26.

147. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 224–27 (S.D.N.Y. 2020).

148. *Id.* This lease agreement would last seven years at low wholesale rates and would have no cap on usage for the first three years.

149. *Id.* at 225-37.

150. *Id.* at 229.

costs of relying on one vendor to build out a RAN.[151] The court accepted Defendants' argument that these two technologies, along with access to some of Sprint's decommissioned infrastructure, would reduce market barriers to entry.[152]

3. *Plaintiffs' Last Burden*

Plaintiffs brought additional evidence of anticompetitive effects to invalidate Defendants' rebuttal. First, Plaintiffs submitted evidence indicating that New T-Mobile might engage in anticompetitive behavior post-merger.[153] Documents from both Sprint and T-Mobile showed that executives had "considered the prospect of anticompetitive coordination."[154] These documents also indicated that executives believed a "consolidated market would be more profitable" and that price signaling had started between the companies.[155] However, the court noted that since the creation of those documents, price signaling had stopped.[156] Disregarding these documents, the court instead considered the history of the market.[157] It claimed that because T-Mobile had always been a "maverick" and had disrupted prices, New T-Mobile would continue to do so.[158] The court highlighted that it had spent "two full weeks assessing the credibility of each witness and their claims."[159] It relied on testimony from CEOs stating that collusion would not fit with T-Mobile's previous strategies and therefore believed that New T-Mobile would not behave in an anticompetitive manner.[160]

Second, statements from executives indicated that DISH had overstated its commitment to entering the market.[161] Plaintiffs entered documents showing that Defendants believed DISH's bid to enter the market was a "stupid bluff[]" and the company would just "build a meaningless thin network" to avoid fines with the FCC.[162] An executive stated that the network would be "something the lawyers can use, but not something customers can use."[163] Additionally, Plaintiffs brought evidence that DISH's spectrum

151. *Id.*
152. *Id.*
153. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 235 (S.D.N.Y. 2020).
154. *Id.*
155. *Id.* at 235–36.
156. *Id.* at 236.
157. *Id.* at 235–36.
158. *Id.* at 237.
159. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 236 (S.D.N.Y. 2020).
160. *Id.* at 236–37.
161. *Id.* at 230.
162. *Id.* (internal quotation mark omitted).
163. *Id.*

portfolio was actually "speculative hoarding": many people believed DISH was holding spectrum so it could resell it once other providers became desperate enough to pay higher prices.[164] However, the court did not believe that these statements accurately reflected DISH's intent.[165] Instead, it relied on the FCC and DOJ's commitment that DISH would enter the market.[166]

Lastly, Plaintiffs countered Defendants' assertion that DISH's use of New T-Mobile's RAN would help it compete in the market.[167] They argued that New T-Mobile would have strong incentives for DISH to fail at becoming a viable competitor and would hurt their chances of market entry.[168] New T-Mobile would hold a powerful position because it would be leasing DISH the RAN needed to enter the market.[169] New T-Mobile could manipulate prices or slow down speeds to hurt DISH.[170] However, the court believed the appointment of an independent monitor would remedy this problem.[171] This monitor would oversee New T-Mobile's relationship with DISH and create fixed formulas to calculate pricing.[172] These formulas would ensure DISH a fair chance to compete and secure the network against anticompetitive prices.[173]

Overall, the court found that Defendants' evidence weighed heavily in favor of the merger, and Plaintiffs' evidence was not sufficient to prove likely anticompetitive effects. The court believed that efficiencies from the merger and the addition of a new competitor would increase quality and decrease prices for consumers.

IV. DYNAMIC MARKETS AS UNDERSTOOD THROUGH THE LENS OF NEW YORK V. DEUTSCHE TELEKOM

This Note argues that because the court in *Deutsche Telekom* did not properly account for the unique characteristics of dynamic markets, its analysis was incomplete and incorrect. If the court had used an analysis specific to dynamic markets, it may have found the merger to be anticompetitive. But regardless of the outcome, considering the specifics of dynamic markets would

164. *Id.* at 231.
165. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 230–31 (S.D.N.Y. 2020).
166. *Id.* The court also noted that the DOJ and FCC would levy fines against DISH if it did not enter the market or meet its other commitments. *Id.*
167. *Id.* at 228.
168. *Id.*
169. *See id.*
170. *See id.*
171. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 228 (S.D.N.Y. 2020).
172. *Id.*
173. *Id.*

have ensured that the court fully assessed the true implications of the merger. Lessons learned from this case should be carried forward to ensure antitrust analysis is more accurate for future mergers in dynamic markets.

While the *Deutsche Telekom* court did not subscribe to a strict Chicago School analysis of antitrust, the issues in its holding are reflective of broader deficiencies in the framework. The court veered from the usual Chicago School analysis in two ways. First, it criticized antitrust law's focus on competing economic measures and briefly mentioned other public benefits of the merger, neither of which would be a concern in a strict Chicago analysis.[174] Second, the court recognized the dynamic structure and rapid innovations of the telecommunications market.[175] Importantly, however, these acknowledgments did not result in a more expansive analysis. The court instead focused on the same traditional Chicago School metrics of innovation, quantity, quality, and low prices.[176] Therefore, the weaknesses in this court's analysis and lessons learned from this case are still relevant to the Chicago School at large. When applied to dynamic markets, the Chicago School's focus on short-term consumer welfare measured through prices and quality is insufficient.

This Part proceeds in two sections. Section A critiques three specific problems in the analysis of *Deutsche* Telekom and discuss how these problems exist in analyses of all dynamic markets. Section B explores how to implement changes in antitrust law to avoid two of these problems in the future. It argues that, as currently applied, the Chicago School framework is incomplete. To accurately assess the effects of mergers in dynamic and integrated markets, the approach to promoting the current goals of the Chicago School must be altered, and the goals themselves must be expanded. This new approach must consider both short and long term effects of efficiency and the impacts of a merger on individuals outside of direct consumers, including laborers, retailers, and other global players.

A. A CRITIQUE OF DEUTSCHE TELEKOM

The court's analysis in *Deutsche Telekom* was flawed on two grounds.[177] The court: 1) conducted only a narrow and short-term economic efficiency analysis;

174. *Id.* at 193, 232.

175. *Id.* at 242–43.

176. *See generally* New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 242–43 (S.D.N.Y. 2020); DEP'T OF JUSTICE & FED. TRADE COMM'N, *supra* note 11.

177. The court was also flawed in its evidentiary prioritization of witness statements over documentation. While witnesses pledge an oath to tell the truth, they can be motivated by different factors, such as professional and career goals and personal success. Additionally, there is a healthy churn of executives (who act as witnesses in these trials) in corporate

and 2) focused on the effects of the merger for consumers without considering global-scale effects.

1. *Economic Efficiency Analysis*

Because *Deutsche Telekom*'s efficiency analysis did not account for disruptive innovation in dynamic markets, the court incorrectly concluded that short-term efficiencies from the merger would benefit consumers and promote better service and quality.[178] The court's analysis was incomplete in two ways. First, the court should have examined the long-term effects of efficiencies, specifically effects on barriers to entry. Second, it should not have relied overwhelmingly on past firm behavior as a predictor for future behavior.

a) An Overview of Deutsche Telekom's Incomplete Short-Term Efficiency Analysis

In *Deutsche Telekom*, the court accurately noted the potential short-term efficiencies resulting from the merger. These efficiencies are particularly important to study in capital-intensive markets such as the telecommunications market, where firms must invest in expensive physical infrastructure to compete.[179] In this case, the court highlighted the positive effects that efficiencies from combined spectrum and consolidated capital investment might have on consumers, including lower prices and better service.[180]

While the court's efficiency analysis was appropriate under the Chicago School framework, it was incomplete in this context because it did not evaluate the long-term tension between investment and innovation in dynamic markets.[181] More specifically, it failed to consider that efficiencies may have pronounced long-term effects in dynamic markets that ultimately undermine competition. While efficiencies such as consolidation of infrastructure can

---

management structures. Even if a witness is reliable on the stand, she could be replaced by someone who holds disparate views. As a germane example, T-Mobile's own CEO, John Legere, announced he would be leaving the company in early 2020, soon after the *Deutsche Telekom* trial ended. *See* Chris Weich, *John Legere Abruptly Resigns from T-Mobile Board of Directors "To Pursue Other Options,"* THE VERGE (Apr. 24, 2020, 6:01 PM ET), https://www.theverge.com/2020/4/24/21235226/john-legere-resigns-tmobile-board-directors [https://perma.cc/S5NX-8JFC]. Even if the court's reliance on Legere's testimony and his dedication to keeping T-Mobile competitive was well founded, New-T-Mobile will now be run by new management that could easily change strategies. *See id.*

178. *See* New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 207–10 (S.D.N.Y. 2020).

179. *See id.* at 191–92.

180. *Id.* at 207–10.

181. *See id.* at 167–68; *see also* John Kwoka, *The Effects of Mergers on Innovation: Economic Framework and Empirical Evidence*, THE ROLES OF INNOVATION IN COMPETITION LAW ANALYSIS 13, 13 (Paul Nihoul and Peter Van Cleyenbruegel eds., 2018).

lower prices in the short term by reducing overhead costs, such consolidation can limit dynamic innovation in the long term by reducing incentives to entirely disrupt the industry. As discussed in Part III, dynamic markets rely on disruptive innovation to bring better and often transformative services and products to the market.[182] However, if service providers such as New T-Mobile amass enough capital investment through merger efficiencies, they will no longer have the incentive to promote disruptive innovation, as this disruption would likely render many of the company's capital investments obsolete.[183] The same investment efficiencies that created short-term price reductions could therefore lead to long-term innovation stagnation in the market and reduce the quality of services.[184] For example, if in ten years 6G were developed[185] and based on an entirely new type of spectrum or transmission infrastructure, New T-Mobile would face little incentive to embrace 6G as it would leave the company's current spectrum holdings useless. Even though 6G would benefit consumers, New T-Mobile and other providers, equipped with market concentration and domination, could slow the rollout of this hypothetical 6G.

Because the efficiency analysis is different in dynamic and static markets, courts must more carefully protect long-run competition in dynamic markets to incentivize firms to continually embrace disruptive technologies and developments. Long-term efficiency consideration is rarely needed in static markets, even if the markets are capital intensive, because firms are incentivized to innovate within static market structures to differentiate themselves from competitors and attract new customers.[186] Consumers benefit from this innovation by receiving higher-quality products. But courts must take extra care in allowing disruptive innovation to flourish in capital intensive, dynamic markets.

Specifically, in *Deutsche Telekom*, the court should have not just studied short-term efficiencies, but balanced these efficiencies with the long-term effects on innovation and competition. A more effective analysis would have

182. Sidak & Teece, *supra* note 2, at 581.

183. Kwoka, *supra* note 181, at 16–17.

184. *Id.* at 16–17; *see also* Paul Krugman, *Monopoly Capitalism Is Killing US Economy*, THE IRISH TIMES, (April 19, 2016, 12:00), https://www.irishtimes.com/business/economy/paul-krugman-monopoly-capitalism-is-killing-us-economy-1.2615956 [https://perma.cc/C2UB-KZMH] (explaining that it did not make sense for Verizon's profit margin to expand its own high speed network even though there was strong demand from consumers).

185. *See* Klint Finley, *Trump Shouldn't Plan to Tweet From a 6G Phone Anytime Soon*, WIRED (Feb. 21, 2019, 8:44 PM), https://www.wired.com/story/trump-shouldnt-plan-tweet-from-6g-phone-soon [https://perma.cc/E8B3-TL68].

186. *See* Kwoka, *supra* note 181, at 16–17.

carefully examined barriers to market entry and would not have relied on past behavior to predict future behavior. The next two Sections address the balancing of short- and long-term efficiencies.

b) Analysis of Barriers to Entry

To effectively analyze long-run competition in *Deutsche Telekom*, the court should have conducted a more thorough analysis of DISH Network's potential entry as a new fourth competitor in the telecommunications market and its likelihood of mitigating the merger's anticompetitive effects. Although the court devoted nearly ten pages to this analysis,[187] it placed undue weight on DISH's claims that it would build a "virtualized network" and operate an Open Radio Access Network (ORAN), and incorrectly dismissed the plaintiff's contradicting evidence.[188] The court misunderstood—or failed to examine—key technologies that were unproven in the market and incorrectly concluded that DISH would be a viable market entrant because it assumed with certainty that DISH could rely on these technologies. Instead, it should have looked only to currently viable technologies to assess DISH's likelihood of entry.

The court accepted DISH's arguments that virtual network technologies would decrease construction and network costs by reducing the need for physical infrastructure.[189] But virtualized networks were a nascent concept at the time with the "hype around [them] exceed[ing] the reality at the moment."[190] These underdeveloped networks could prevent DISH from becoming a successful competitor.[191] The theoretical cost benefits of virtual networks depended on the fact that their price would be lower than traditional RAN; however, if implementation issues arose, such as problems developing a network strong enough to support a large customer base, the price would skyrocket, causing major financial troubles for DISH.[192] The court acknowledged that even with its current spectrum holdings, DISH would require substantial time and money to build a full MVO network and mitigate anticompetitive merger effects.[193] Despite this, it still relied on unproven technologies to find that DISH would be a viable future competitor.

187. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 224–33 (S.D.N.Y. 2020).

188. *See id.* at 229.

189. *Id.*

190. Matt Kapko, *Dish Faces Challenges on Virtualized Network Vision*, SDX CENTRAL (July 30, 2019, 1:45 PM), https://www.sdxcentral.com/articles/news/dish-faces-challenges-on-virtualized-network-vision/2019/07 [https://perma.cc/8RZR-TZEY].

191. *See Deutsche Telekom*, 439 F. Supp. 3d at 228.

192. *See id.*

193. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 228 (S.D.N.Y. 2020).

In so concluding, the court ignored expert testimony rebuking these technologies. Experts testified that DISH would likely respond to problems with virtual networks by reverting to the more expensive and time-intensive RAN network, reducing the likelihood of market entry success.[194] Additionally, DISH's own CEO admitted outside of trial that a virtual network could cost billions of dollars to implement and recognized the possibility that it might fail.[195] Telecommunications executives also acknowledged this possibility in documents.[196]

Instead of relying on unproven technologies, the court should have analyzed the potential of entry using only currently viable technology. Innovations in dynamic markets can occur so rapidly that it is imprudent to depend on the successful implementation of a specific nascent technology. In other words, in these markets, courts should "ignore innovation possibilities that are too remote to see."[197] In the case of the telecommunications market, neither ORAN nor virtualized networks had been fully endorsed by a majority of experts as viable options.[198] If the court had relied on current technologies, it would have concluded that the barriers to entry—namely the capital costs of developing a RAN system in a timely period—would be impossible for DISH to successfully overcome, even with DISH's agreement with T-Mobile. DISH's inability to enter the market would reduce competition in the long run, inhibiting innovation and hurting consumers and producers in the market.

c) Analysis of Past Behavior

The court's balancing of economic efficiencies with potential anticompetitive effects was also tainted by its reliance on past behavior. Throughout *Deutsche Telekom*, the court assumed that market players' past competitive practices predicted strong post-merger market competition.[199] But past behavior is an ineffective and inaccurate indicator of future behavior in dynamic markets, where rapid innovation can quickly render past practices obsolete.

---

194. Dan Meyer, *Dish Networks Views Virtualization as 5G's "Real Paradigm Shift*," SDX CENTRAL (Apr. 11, 2019, 9:37 AM), https://www.sdxcentral.com/articles/news/dish-networks-views-virtualization-as-5gs-real-paradigm-shift/2019/04/ [https://perma.cc/8RZR-TZEY].

195. *Id.*

196. *Deutsche Telekom*, 439 F. Supp. 3d at 198.

197. Herbert J. Hovenkamp, *Schumpeterian Competition and Antitrust*, FACULTY SCHOLARSHIP AT PENN L. 10–11 (Oct. 15, 2008), https://scholarship.law.upenn.edu/faculty_scholarship/1788/.

198. Kapko, *supra* note 190.

199. *See* New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 236 (S.D.N.Y. 2020).

The nature of dynamic markets—new technologies or innovations dramatically altering the market landscape—can make previous assumptions irrelevant and necessitate a shift in market strategies.[200] As discussed above, once a firm dominates a market, its "interest in creative destruction becomes greatly diminished" because it will not want its new market position to "erod[e]."[201] Accordingly, the likelihood of the firm continuing to innovate will decrease, meaning that past behavior will not predict future behavior.[202] As antitrust scholar Professor Hovenkamp aptly stated, "the small firm seeking entry must shake up the pot" but the "dominant firms are well established and . . . tend to profit from stable growth."[203]

While in past years, T-Mobile was the underdog seeking market power and therefore relied heavily on innovation and creative thinking, this mindset changed as it gained market power.[204] T-Mobile's former position as third-largest competitor allowed it to become a maverick, and T-Mobile accordingly "built its identity and business strategy on . . . challenging AT&T and Verizon" to capture their customer base.[205] Relying on testimony from T-Mobile CEO Legere, the court concluded that New T-Mobile would not stop these practices post-merger, but instead continue to aggressively compete.[206] However, the merger could change the company's priorities, making it no longer profitable to upturn the market. For example, if New T-Mobile invested in a 5G network that enabled it to build a customer base of similar size to AT&T or Verizon, it might no longer invest in continuously improving its network and shaking up the standards for telecommunications providers, but instead spend its money on protecting its customer base.

2. *Consumer and Global Impacts*

In addition to focusing solely on short-term efficiency at the expense of fully understanding long-run effects on competition, the court in *Deutsche Telekom* confined its review to individual-level effects on consumers and did not consider the effects of the merger on a global scale. The Chicago School usually defines a consumer as "a person or entity that engages in

200. Gonzaga, *supra* note 84, at 5.

201. Hovenkamp, *supra* note 197, at 6–7.

202. *Deutsche Telekom*, 439 F. Supp. 3d at 193–94.

203. Hovenkamp, *supra* note 197, at 6. While DISH's entry might not be small in spectrum size, it would be small in that DISH would have to build its customer base and market reputation from scratch.

204. *Deutsche Telekom*, 439 F. Supp. 3d at 193–94.

205. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 237 (S.D.N.Y. 2020).

206. *Id.*

consumption."[207] Consumers are directly involved in the market through their purchases of goods and services. In line with this definition, the court exclusively focused on the merger's effects on the prices and quality of services in the telecommunications market (including types of cellular networks) because those were the primary ways in which market consumers would be impacted.[208] However, unlike static markets that can be studied in isolation—such as the milk, shoe, or grocery store markets—the telecommunications market is highly integrated in today's global economy.[209] Services that the market provides help fuel the gig economy, retail shopping, banking, development, and various other markets.[210] Therefore, a merger not only affects consumers, but also other individuals and institutions throughout the economy. While studying consumer effects is essential, courts must expand their analyses to also consider global economic effects that impact all entities the market indirectly touches. In the case of the telecommunications market, these indirect markets include labor, global trade, and national security. In focusing only on direct markets, courts risk neglecting important consequences of the merger.

Deutsche Telekom's failure to thoroughly consider the importance of building an American 5G network illustrates this concern. The race to develop the first fully functional 5G network is the single largest issue in the telecommunications market.[211] As discussed in greater detail below, it affects the labor market, technological developments in numerous sectors, and national security. Because it is so important to the market and the global economy, the court should have considered the effects of the merger in light of this race.[212] While the court acknowledged that the merger would allow a quicker rollout of 5G and briefly noted these impacts, they were relegated almost exclusively to a single footnote.[213]

The U.S. economy, and almost all individuals, would greatly benefit from the development of a robust 5G network. In building out the first 5G network, the United States would hold its place as a top global innovator and

207. Caron Beaton-Wells, *Antitrust's Neglected Question: Who Is "The Consumer,"* 65 ANTITRUST BULLETIN 173, 175 (2020).

208. *Deutsche Telekom*, 439 F. Supp. 3d at 208–09.

209. *See id.* at 189–90.

210. *See id.*

211. *See* Klint Finley & Joanna Pearlstien, *The WIRED Guide to* 5G, WIRED (Sept. 10, 2020, 7:00 AM), https://www.wired.com/story/wired-guide-5g [https://perma.cc/K9B4-VW6P].

212. *See* New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 209 n.12 (S.D.N.Y. 2020).

213. *Id.*

outcompete other technologically advanced countries such as China and Korea.[214] The United States would also set telecommunications standards and solidify global dependence on, and sales of, American products and services.[215] More importantly, a 5G network would create new economic opportunities for many industries, including those involving self-driving cars, telemedicine, automated factories, and smart appliances.[216] In turn, these opportunities would have large impacts on employment and economic growth.[217]

Additionally, the race to 5G has serious national security implications. The federal government has already banned U.S. companies from purchasing 5G products or technology from certain Chinese companies, including Huawei, because of fears that these products will allow the Chinese government to conduct cyber espionage on the American government.[218] There are also fears that Huawei will allow Chinese officials and businesses to steal American intellectual property.[219] While the impacts of national security on the economy are less direct than on job growth, they can be crucial to the safety of individuals and businesses.[220]

The court should have carefully discussed each of these considerations, as the merger of Sprint and T-Mobile could significantly change the rollout of 5G. The promise of New T-Mobile developing a successful 5G network might have supported the court's decision to let the merger proceed, as it would mitigate national security concerns and stimulate the overall economy. If New

---

214. *Id.*

215. *See* Jeremy Hsu, *How the U.S. Can Prepare to Live in China's 5G World*, IEEE SPECTRUM, (Apr. 23, 2019), https://spectrum.ieee.org/tech-talk/telecom/standards/how-america-can-prepare-to-live-in-chinas-5g-world [https://perma.cc/2NPF-7ZM5].

216. *See* Brian Fung, *The Race to 5G Wireless Tech is on. A Report Finds Americans May Have an Early Lead*, WASH. POST, (Feb. 19, 2019, 11:44 AM PST), https://www.washingtonpost.com/technology/2019/02/19/race-g-wireless-tech-is-report-finds-americans-have-an-early-lead [https://perma.cc/G2QY-ZLYE, https://perma.cc/6QS3-YZJY].

217. *See id.*

218. *See* David McCabe, *F.C.C. Designates Huawei and ZTE as National Security Threats*, N.Y. TIMES (June 30, 2020), https://www.nytimes.com/2020/06/30/technology/fcc-huawei-zte-national-security.html [https://perma.cc/P3GQ-CPY3].

219. *See* Lindsay Maizland & Andrew Chatzky, *Hwawei: China's Controversial Tech Giant*, COUNCIL ON FOREIGN RELATIONS (August 6, 2020 8:00 AM EST), https://www.cfr.org/backgrounder/huawei-chinas-controversial-tech-giant [https://perma.cc/S79Z-UUM7].

220. In the fall of 2020, President Trump announced he would take actions to ban WeChat, the popular messaging app, in the United States because he believed that the Chinese app infringed on the data security and privacy of its American users. Geoffrey Gertz, *Why is the Trump Administration Banning TikTok and WeChat?*, BROOKINGS (Aug. 7, 2020), https://www.brookings.edu/blog/up-front/2020/08/07/why-is-the-trump-administration-banning-tiktok-and-wechat [https://perma.cc/Q6LZ-3PS8]. While this ban did not go through, it exemplifies potential national security issues that can affect lives of ordinary citizens. *See id.*

T-Mobile is not able to build out this network, the United States might not see a strong 5G network at a fast timeline. However, if they successfully and rapidly build out a network, other individuals might suffer the consequences of this prioritization of a 5G network. A focus on 5G prices for consumers alone cannot capture these larger effects.

The court should have analyzed these important international and broader economic and labor considerations in its merger analysis. While overarching economic considerations are not part of a traditional Chicago School analysis, their importance demonstrates why the Chicago School must be updated.

B. DEUTSCHE TELEKOM AND THE FUTURE OF ANTITRUST LAW

Antitrust law must adapt to reflect the emergence of dynamic markets and to avoid the missteps in *Deutsche Telekom*. As the court discussed in the case, the job of predicting a merger's effects is difficult and onerous.[221] A court faced with this inquiry must somehow make sense of conflicting facts and unclear evidence to decide what will affect competitors, consumers, and the market.[222] It must do so without the guidance of straightforward formulas or checklists. Instead, the court must weigh competing evidence and rely on its own knowledge of market behavior.[223] Predicting this behavior has become even more complex in dynamic markets that are constantly changing and inextricably integrated to the broader economy.[224] Technological advancements and globalization have made it harder to isolate markets and predict their effects, as the world has become increasingly interdependent.[225] But these difficulties should not lower antitrust standards. Instead, courts, enforcement agencies, and the legislature should review and adjust the ways in which they analyze markets to ensure that scrutiny of mergers within dynamic markets continues to promote the goals of antitrust.

American jurisprudence must make two distinct changes to correctly regulate and monitor dynamic markets. First, courts must expand their efficiency analysis to balance the tension between short-term market efficiencies and the long-term need for disruptive innovations. Second, courts must account for merger effects on both individual market consumers and the global economy, recognizing that shifts in a dynamic and integrated market

221. *Id.* at 187–88.

222. *See id.*

223. *See id.*

224. *See generally* Kwoka, *supra* note 181.

225. *See* Christine Lagarde, *The Interconnected Global Economy: Challenges and Opportunities for the United States—And the World*, INT'L MONETARY FUND (Sept. 19, 2013), https://www.imf.org/en/News/Articles/2015/09/28/04/53/sp091913 [https://perma.cc/97CQ-EDZB].

can have profound effects on numerous international markets. Courts should examine the various ways in which these changes will depart from the Chicago School as it currently exists. While the Chicago School (and antitrust generally) promotes necessary goals such as innovation, quantity, quality, and low prices for consumers, the efficiency analysis used to measure these objectives must be changed. Additionally, the goals themselves must be expanded to account for the effects of mergers on the broader economy, including long-term effects and connected markets.

These shifts align generally with the Neo-Brandeisian view of antitrust.[226] While this Note does not discuss the intricacies of the various Neo-Brandeisian arguments, *Deutsche Telekom* illustrates why Neo-Brandeisians' broader focus on political and social concerns—as well as the protection of non-consumer players in the market and the competitive process as a whole—should be applied to dynamic, integrated markets. For example, the implications of an American 5G market include labor concerns and international power relations in addition to consumer prices. These concerns are better analyzed under a Neo-Brandeisian framework.

Some might argue that it is futile to ask judges to consider broader economic impacts or longer-term efficiency analyses because these considerations might become overwhelming, unadministrable, and political.[227] However, judges have tools to help them conduct these broader investigations. First, judges should rely on documentary evidence, especially over witness testimony, as a signal for pertinent considerations in the market, given that many of these documents come from market experts who have special knowledge about trends and technologies.[228] Second, and as Judge Marrero himself stated, judges can use their "own skills and frontline experience[s]" to make determinations and parse out larger considerations.[229] Unlike relying on witnesses who might have varying rationales and hopes for an outcome, a judge can use her own skills, as well as amici documents, to determine the relevance and importance of a merger. While assessing every potential downstream effect from a merger would be unadministrable and imprecise, judges can use their discretion to focus on salient areas with broader impacts. For instance, the *Deutsche Telekom* court recognized, though it did not deeply analyze, the global economic impact in the telecommunications industry as

226. *See* Sallet, *supra* note 37, at 368.

227. *See generally* Thibault Schrepel, *Antitrust Without Romance*, 13 N.Y.U. J.L. & LIBERTY, 326, 326 (2020) (lamenting what the author deems to be the shortcomings of the revitalization movement of antitrust law, and describing the changes proposed to make antitrust more robust as "romanticizing" antitrust).

228. *See* Albert Foer, *Prediction and Antitrust*, 56 ANTITRUST BULLETIN 505, 520 (2001).

229. New York v. Deutsche Telekom AG, 439 F. Supp. 3d 179, 188 (S.D.N.Y. 2020).

jobs and national security.[230] Because judges can see the evidence simultaneously, they are in the best position to recognize the unique considerations of each case.

Courts cannot update this framework alone but must partner with governing agencies and regulatory bodies to bring about cohesive and unified change. There are several ways that the judicial and government branches of government can achieve this change. First, the Department of Justice should revise its merger guidelines to explicitly consider dynamic markets and the effects of disruptive innovation.[231] As Professors Gregory Sidak and David Teece argue, these guidelines must reflect an understanding that dynamic markets, unlike static markets, are directly driven by innovation and therefore require different analyses.[232] Second, Congress should pass a comprehensive bill updating the 1914 Clayton Act to reflect the differences in analyzing dynamic markets, such as the importance of protecting incentives to innovate. Most recently, Senator Klobuchar proposed a bill in 2017 that would tighten the standard for acquisitions in the Clayton Act.[233] While this bill did not pass, it illustrated the possibly of introducing these amendments in Congress. This adoption would take an adjustment period, but updating antitrust in dynamic markets will best prepare all parties to properly regulate the new dynamic markets of the twenty-first century.

Today's antitrust climate is ideally suited to advocate for structural updates, and the lessons learned from *Deutsche Telekom* are a perfect vehicle to drive these changes. There has been a renewed focus on antitrust law in the last few years, most notably in the form of bi-partisan congressional hearings examining potential antitrust violations by large technology companies.[234] In a rare move, the DOJ recently filed an antitrust suit against Google, a dominant market player in the dynamic search engine market.[235] Further, the FTC and multiple state attorneys general recently brought separate antitrust suits against social media giant Facebook, claiming that the company "illegally squash[ed]

230. *Id.* at 210.

231. Sidak & Teece, *supra* note 2, at 885.

232. *Id.*

233. Consolidation Prevention and Competition Promotion Act of 2017, S. 1812, 115th Cong. (2017).

234. *See* Shannon Bond, *Congress Holds Big Tech Antitrust Hearing*, NPR (July 29, 2020, 4:05 PM ET), https://www.npr.org/2020/07/29/896840093/congress-holds-big-tech-antitrust-hearing [https://perma.cc/EDG8-H7TW].

235. Cecilia Kang, David McCabe & Daisuke Wakabayashi., *U.S. Accuses Google of Illegally Protecting Monopoly*, N.Y. TIMES (Oct. 20, 2020), https://www.nytimes.com/2020/10/20/technology/google-antitrust.html [https://perma.cc/T7BT-DVPD].

competition" by "buying up its rivals."[236] They allege that Facebook's motivation for purchasing rivals was to "eliminate[] competition that could have one day challenged the company's dominance"[237] and that their actions "leave[] consumers with few choices for personal social networking."[238] Notably, these regulators have recognized that integrated and dynamic markets and platforms such as Facebook can "deny consumers the benefits of competition" and that regulators must step in to preserve innovation in the market.[239]

Antitrust regulators have started to make the incremental changes needed to properly protect individuals in dynamic markets, but these changes must be solidified uniformly through statutes, guidelines, and judicial opinions. The recent focus on antitrust can usher in a broader, more accurate view of the law that is built for the dynamic markets of the twenty-first century.

V. CONCLUSION

Modern technology markets are increasingly dynamic and integrated into society. Scholars and litigators must update the analytical framework of antitrust law to reflect these changes if they hope to effectively protect competition and consumers. Updates to antitrust may also incentivize different third parties, not just the federal government, to bring antitrust suits, as the state of New York did in *Deutsche Telekom*. While the act of predicting future market behavior will never reach perfection, this Note suggests that courts and regulatory agencies can better protect markets by studying long-term efficiencies and overall economic effects. This expands both the goals of antitrust and the means of achieving those goals.

236. Cecilia Kang & Mike Isaac, *U.S. and States Say Facebook Illegally Crushed Competition*, N.Y. TIMES (Dec. 9, 2020), https://www.nytimes.com/2020/12/09/technology/facebook-antitrust-monopoly.html [https://perma.cc/2WTB-F3LX].

237. *Id.*

238. Press Release, Fed. Trade Comm'n, FTC Sues Facebook For Illegal Monopolization (Dec. 9, 2020), https://www.ftc.gov/news-events/press-releases/2020/12/ftc-sues-facebook-illegal-monopolization [https://perma.cc/WJZ5-VMUT].

239. *See id.*

WRITING IN RACE: CULTURAL DEMOCRACY IN THE DIGITAL AGE

Natalie T. Crawford†

I. INTRODUCTION

Social media platforms change the social conditions of speech, allowing historically oppressed groups to mobilize in new ways. For instance, in 2013, the Black Lives Matter (BLM) Movement was formed online by Patrisse Khan-Cullors, Alicia Garza, and Opal Tometi via hashtags like #BlackLivesMatter following the killing of Trayvon Martin in 2012.[1] The movement pushed back on media's portrayal of young Black men as thugs, thereby "recoding" society's perception of Black death into a continuation of racism and oppression of the Black community. More recently, in 2020, BLM came back into the cultural forefront following the deaths of George Floyd, Ahmaud Arbery, and Breonna Taylor. Protests were ignited not only in the streets but also online in the form of social media blackouts,[2] brand boycotts,[3] and rallied support for Black artists, creators, and businesses.[4] In doing so, social media has become a vehicle for the cultural participation of the Black community as they voice not only the harms done to the Black community but also its successes. This highlights that as technology evolves, so do methods of meaning making, as well as the means of participation. In other words, not only can the meaning of Black death in America change, but it can be changed due to the increased participation of Black persons in the cultural ecosystem.

DOI: https://doi.org/10.15779/Z38MW28G08

† J.D. Candidate, 2022, University of California, Berkeley, School of Law.

1. *Black Lives Matter*, BRITANNICA (Aug. 13, 2020), https://www.britannica.com/topic/Black-Lives-Matter.

2. Sandra Gonzalez, *Music industry leaders vow to pause business for a day in observation of Blackout Tuesday*, CNN (Jun. 2, 2020), https://www.cnn.com/2020/06/01/entertainment/blackout-tuesday-the-show-must-be-paused/index.html.

3. Coral Murphy, *'Boycott for Black Lives': People plan to stop spending in companies that don't support BLM*, USA TODAY (Jun. 18, 2020), https://www.usatoday.com/story/money/2020/06/18/boycotts-people-plan-stop-spending-stores-dont-support-blm/3208170001/.

4. Alaa Elassar, *TikTokers stand in solidarity with black creators to protest censorship*, CNN (May 19, 2020), https://www.cnn.com/2020/05/19/us/tiktok-black-lives-matter-trnd/index.html; Tina Bellon, *Uber Eats features black-owned resturants on U.S., Canadian delivery platform*, REUTERS (Jun. 4, 2020), https://www.reuters.com/article/us-uber-eats-protests/uber-eats-features-black-owned-restaurants-on-u-s-canadian-delivery-platform-idUSKBN23B3DL.

Cultural democracy theory explores this long-established process of cultural production.[5] While utopic, cultural democracy "describe[s] a world where audiences freely and widely engage in the use of cultural symbols…A [cultural] democracy enables the audience…to 'resist,' 'subvert,' and 'recode' certain cultural symbols to express meanings that are different from the ones intended by their creators."[6] However, not all audiences are able to "freely and widely" engage in culture. While social media has amplified Black voices, it has also allowed those who seek to oppress the Black community to continue to do so in a larger and louder way than ever before.

For example, Facebook has become a haven for militia groups akin to modern day KKK chapters.[7] Following the shooting of Jacob Blake in Kenosha, Wisconsin, the Kenosha Guard Facebook group raised a call to arms against African Americans and their BLM allies.[8] As a result, social media's lack of geographical limitations worked against the Black community as many who answered the call, including Kyle Rittenhouse from Illinois, travelled to Kenosha to intimidate protestors. Once there, Rittenhouse killed Anthony M. Huber and Joseph Rosenbaum and wounded Gaige Grosskreutz.[9] Despite Facebook's Community Standards preventing "dangerous individuals and organizations" on its platform and the 455 complaints filed against the group prior to the violence in Kenosha, its presence on Facebook allowed racist rhetoric to fester and manifest into real world violence.[10]

---

5. JOHN FISKE, TELEVISION CULTURE 239 (Routledge 1999) (1987). In Fiske's theory, he uses the phrase "semiotic democracy," which is equivalent for cultural democracy in this Note.

6. Sonia K. Katyal, *Semiotic Disobedience*, 84 WASH. UNIV. L. REV. 489, 489-90 (2006).

7. Tyler Sonnemaker, *Facebook 'did nothing' about violent militia and hate groups for 5 years despite being warned at least 10 times, advocacy group says*, INSIDER (Sep. 17, 2020), https://www.businessinsider.com/facebook-ignored-warnings-violent-anti-muslim-militia-hate-groups-2015-2020-9; *see also* Ewan Palmer, *Facebook Sued Over Kenosha Shootings Using 19th Century Law to Combat KKK*, NEWSWEEK (Sep. 24, 2020), https://www.newsweek.com/facebook-lawsuit-kenosha-protest-kkk-1534003.

8. Ryan Mac and Craig Silverman, *How Facebook Failed Kenosha*, BUZZFEED NEWS (Sep. 3, 2020), https://www.buzzfeednews.com/article/ryanmac/facebook-failed-kenosha; *see also* Russell Brandom, *Facebook takes down 'call to arms' event after two shot dead in Kenosha*, VERGE (Aug. 26, 2020), https://www.theverge.com/2020/8/26/21402571/kenosha-guard-shooting-facebook-deplatforming-militia-violence.

9. Brad Parks and Artemis Moshtanghian, *Kenosha shooting suspect Kyle Rittenhouse posts $2 million bail*, CNN (Nov. 20, 2020), https://www.cnn.com/2020/11/20/us/kyle-rittenhouse-kenosha-shooting-bail-release/index.html.

10. Mac & Silverman, *supra* note 8' *Dangerous Individuals and Organizations*, Facebook: Community Standards, https://www.facebook.com/communitystandards/dangerous_individuals_organizations (last visited Sep. 30, 2022).

While racist speech may not always result in physical violence as in Kenosha, the harm is no less *real*. When confronted with racist speech, it affects someone at not only the level of individual dignity, but it is also felt on the level of the collective, as members of a group that has been historically oppressed.[11]

As a result, the digital age emphasizes two forms of speech: one that seeks to engage in cultural production and another that limits the cultural production of the former, intentionally or otherwise. Therefore, within a *race-conscious* cultural democracy, some control of racist speech is necessary in order to allow everyone to freely and widely engage.

Social media platforms are well-positioned to rethink the way in which racist speech is controlled because while the First Amendment shields some racist speech, it only prevents government encroachment. Platforms, as private actors, are therefore given breadth of discretion as to how they can conceptualize and control racist speech. Facebook in particular has repeatedly found itself at the intersection of race and speech online. In May 2020, the platform established an Oversight Board to help Facebook manage speech on its sites (Facebook and Instagram).[12] Two of the Board's first six cases concerned hate speech content.[13]

This Note uses Facebook and its Oversight Board to examine how a race-conscious cultural democracy might function online. It concludes that Facebook's current definition of hate speech stands open to abuse and that more transparency and methods of information dissemination would enable Facebook to better trust its users, thereby allowing users to more effectively regulate themselves. However, while community control is the most democratic, it stands open to the criticisms of cancel culture and an inability to effectuate substantive change against racism. Therefore, some form of corporate oversight is needed, and the Oversight Board is a promising method of such control.

Part II of this Note lays out the cultural democracy framework as dictated by current legal theorists and proposes how it should be reworked to account for inequalities between persons and ideas, particularly those of historically oppressed groups. It then establishes a theory of race-conscious cultural

---

11. Richard Delgado, *Words that Wound: A Tort Action For Racial Insults, Epithets, and Name-Calling*, 17 HARV. C.R. C.L. LAW REV. 133, 136 (1982).

12. Nick Clegg, *Welcoming the Oversight Board*, FACEBOOK: NEWSROOM (May 6, 2020), https://about.fb.com/news/2020/05/welcoming-the-oversight-board/.

13. *Announcing the Oversight Board's first cases and appointment of trustees*, OVERSITE BD. (Dec. 2020), https://www.oversightboard.com/news/719406882003532-announcing-the-oversight-board-s-first-cases-and-appointment-of-trustees/.

democracy, which recognizes that in order for all persons to participate in cultural production, racist speech must be controlled. Part III outlines the spectrum that racist speech functions on and establishes that if speech is a message (1) "of racial inferiority" (2) "directed against a historically oppressed group" (3) that is "persecutorial, hateful, and degrading," then it should be considered actionable racist speech that warrants control.[14] With this as a guide, this Note examines Facebook's own definition and finds it insufficient. Part IV introduces the hybrid model for speech control in the digital age: community control buttressed by corporate oversight. It examines the way in which community control works both in theory and already in practice on Facebook and then evaluates the benefits of corporate oversight. It then concludes that Facebook should institute more tools to allow it to better trust its users and that Facebook's Oversight Board is an effective method of corporate oversight that promotes a race-conscious cultural democracy.

II. THE CULTURAL DEMOCRACY FRAMEWORK

Culture is the glue of society and the result of collective meaning making.[15] It is a set of conditions that provide a baseline for conversation in society, and it influences everything from dating practices and Twitter trends to political systems and international relations.[16] It is produced in part through a process of democratization. Information, knowledge, and ideas are under constant production or are otherwise remixed (built upon and manipulated into something new).[17] This process maps onto the entire evolution of technology, and the intertwined histories of the printing press and the Bible provides a fitting example to give practical context to this process of democratization.

The printing press, while a centuries-old precursor to modern day social media platforms, emphasizes a history of culture production (and

---

14. Mari J. Matsuda, *Public Response to Racist Speech: Considering the Victim's Story*, 87 MICH. L. REV. 2320, 2357 n. 36 (1989).

15. YOCHAI BENKLER, THE WEALTH OF NETWORKS 279 (2006) ("It is difficult to specify how [culture] functions in terms readily amenable to a conception of individuals whose rationality and preferences for their own good are treated as though they preexist and are independent of society. A conception of culture requires some commonly held meaning among these individuals. Even the simplest intuitive conception of what culture might mean would treat this common frame of meaning as the result of social processes that preexist any individual, and partially structure what it is that individuals bring to the table as they negotiate their lives together, in society or in a polity.").

16. *Id.* at 282 ("[Culture] is a frame of meaning from within which we must inevitably function and speak to each other, and whose terms, constraints, and affordances we always negotiate. ").

17. *See generally* LAWRENCE LESSIG, REMIX: MAKING ART AND COMMERCE THRIVE IN THE HYBRID ECONOMY (2008).

reproduction). In comparison to the 30,000 books available in all of Europe in the early 1450s, Gutenberg managed to produce an astounding 180 Latin Vulgate Bibles of consistent quality in a matter of years.[18] Not even a hundred years after the Gutenberg press revolutionized printing, Martin Luther became the first "best-selling author" with his Bible translation seeing more than 40 editions.[19] Needless to say, the production potential of the printing press not only fueled Luther's individual success but also the Protestant Reformation as a whole. In Luther's own words, "Printing is the ultimate gift of God and the greatest one."[20]

Like Luther, the consumers of the Bible prescribe their own meanings onto the text. Then as producers, these once-consumers reproduce those meanings for others via new translations and versions, reformulating religious culture. With each iteration of this cycle, power is decentralized. While the Church was once relied on as the sole source of salvation, vernacular translations put the power of salvation in the hands of the consumers. As a result, as of 2019, there were around 670 translations of the Bible.[21]

This is the process of the cultural democratization—as means of production improve, the amount of cultural material in the marketplace increases and the reconfigurations of that material proliferate. In other words, ideas spread, information multiplies, and power decentralizes. A simple book burning of a single heretic no longer suffices to silence opposition to the Church. Books print faster than they are burned, and eBooks and PDFs fan the flames of a growing cultural democracy.

A. EXISTING CULTURAL DEMOCRACY THEORY

While cultural democracy, as the above Biblical example shows, is nothing new, the theory that revolves around it is a creature of media studies initially developed by Professor John Fiske in 1987.[22] At that time, consumers had access to more television networks and therefore more shows, and it was this televised information buffet from which Fiske excavated his theory of semiotic democracy. Under this theory, audiences were not blank canvasses, incapable of critical consumption; instead, according to Fiske, the identities of individual

---

18. *Id.*

19. Dave Roos, *7 Ways the Printing Press Changed the World*, History: History Stories (Aug. 28, 2019) https://www.history.com/news/printing-press-renaissance.

20. Pearl Goyal, *Gutenberg's Printing Press*, WOODWARD JOURNAL (May 21, 2020), http://www.woodwardjournal.com/gutenbergs-printing-press.

21. Daniel Esparza, *How many times has the Bible been translated?*, ALETEIA (Jan. 1, 2019), https://aleteia.org/2019/01/06/how-many-times-has-the-bible-been-translated/.

22. Fiske, *supra* note 5 at 239. John Fiske used the phrase "semiotic democracy," which was then interchanged with "cultural democracy" by other theorists.

consumers allow them to superimpose their own meanings onto cultural texts like television shows—meanings that might differ from those intended by show writers and producers—and then reinsert those new meanings into the marketplace.[23] As a result, meaning making is not necessarily a monopolistic hierarchy, but rather a dialogue between cultural producers and consumers in which the consumers become producers themselves.[24]

Jack M. Balkin and other legal scholars[25] brought Fiske's semiotic democracy into the legal realm to re-envision what rights are and, perhaps more importantly, what rights should be when it comes to speech.[26] The encroachment of the digital revolution only further disrupted the hierarchies that Fiske was concerned with a decade prior. In particular, the digital revolution further cut costs of the initial distribution of cultural materials, increased content flows across cultural and geographic borders, cut costs of remixing content, and as a result, further democratized culture.[27] For comparison, prior to Gutenburg, a book had to be transcribed by hand over days, if not months, and struggled to meet a growing demand.[28] However, in 2022, filming a TikTok, posting a Facebook post, or sending a Tweet takes a matter of seconds, is free, and is distributed worldwide. Therefore, as technology changes, so must the rights that secure and encourage participation; in other words, technology clarifies and alters social relations, requiring rights such as free speech to adapt as well.[29]

To this end, free speech is not integral to only political democracy but also cultural democracy. While free speech encourages political debate and the democratic political process, free speech is integral to the cultural democracy as well. According to Balkin, the Internet has several important traits related to free speech, which suggest that a conception of free speech as merely related to political democracy is too limited:

> Speech ranges over a wide variety of subjects, including not only politics but also popular culture. The speech of ordinary people is full of innovation and creativity. That creativity comes from building

23. *See generally* Fiske, *supra* note 5.

24. Katyal, *supra* note 6 at 490.

25. *See generally* Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 2 (2004); Katyal, *supra* note 6; Benkler, *supra* note 15; Lessig, *supra* note 17.

26. Balkin, *supra* note 25 at 52 ("Rights dynamism is the claim that the nature, scope, and boundaries of rights, and in particular fundamental rights like speech, are continually shifting with historical political, economic, and technological changes in the world.").

27. *Id.* at 6-9.

28. *See* Richard W. Clement, *Medieval and Renaissance books*, UTAH ST. U. LIBR. FAC. & STAFF PUBLICATIONS 1, 8 (1997).

29. Balkin, *supra* note 25 at 52.

> on what has come before. Speech is participatory and interactive as opposed to mere receipt of information. It merges the activities of reading and writing, of production and consumption. Finally, speech involves cultural participation and self-formation. The Internet reminds us how central and important these features are to speech generally. It reveals to us in a new way what has always been the case.[30]

Understanding free speech in this way—as central to both political and cultural democracy—allows democracy to be as much about culture as it is about governance. Free speech in the political sphere allows people to debate and create politics, laws, and public policy.[31] Free speech in the cultural sphere allows people to create and progress the culture via meaning making, which in turn produces people's identities.[32] This extends beyond consumerism and deciding which brand of cereal to buy. In reposting a Facebook post, information, knowledge, and/or ideas are not only shared but can be modified through the inclusion of a caption or primer that (re)frames how a future viewer might understand the content. A civil rights activist can share a racist Facebook post in order to educate and counteract the harm done by the original post by providing context and educational resources.[33] Free speech in the cultural democracy is the freedom to create. It is "an active engagement with the world."[34]

In addition, in a robust cultural democracy, speech "is interactive and appropriative."[35] Speech is interactive because it relies on the exchange between persons and is not a one-way street. As the speaker communicates, the receiver internalizes, digests, and responds. In doing so, one either assimilates to or rejects the culture created in that speech.[36] Speech is appropriative because to move culture forward, it remixes—builds on and manipulates—existing cultural materials. For this reason, speech is as collective

---

30. *Id.* at 32.

31. In December, 1791, James Madison wrote in the *National Gazette*, "Public opinion sets bounds to every government, and is the real sovereign in every free one." James Madison, *For the National Gazette*, FOUNDERS ONLINE (DEC. 19, 1791) https://founders.archives.gov/documents/Madison/01-14-02-0145.

32. Balkin, *supra* note 25 at 33.

33. The nuance that arises between racist speech on its own and racist speech used as a tool to inform and advocate creates a grey area of speech that is often difficult for platforms to sift through. Facebook even recognizes this in its Community Standards, and therefore, content that is meant to raise awareness must have its intent clearly indicated. *Hate Speech,* FACEBOOK: COMMUNITY STANDARDS, https://www.facebook.com/communitystandards/hate_speech (last visited Sep. 22, 2022).

34. Balkin, *supra* note 25 at 33.

35. *Id.* at 4.

36. *Id.*

as it is individual; it enacts a cultural system—"a network of people interacting with each other, agreeing and disagreeing, gossiping and shaming, criticizing and parodying, imitating and innovating, supporting and praising."[37]

The insurgence of social media has decreased the cost of participating in cultural democracy and increased access to means of cultural production to the benefit of historically oppressed groups. In the 18th century, slave codes criminalized gatherings of three or more slaves and prohibited slaves from riding horses.[38] By controlling bodies, slaveholders were able to effectively control the sharing of ideas, information, and knowledge—thereby controlling the cultural collective of Black Americans. However, in the digital age, space between bodies no longer affects the ability to congregate and engage in cultural production. The Black Lives Movement is a concrete example of this—multiple protests are able to erupt in various cities on the same day based solely on Facebook event pages. Black Twitter is an even more fluid example of the growing cultural participation of Black voices online. Through platform engagement via coded language and hashtags, Black Twitter users can find each other, share experiences, and communicate. One prominent focus of Black Twitter every year is the Oscars. In 2015, 86 percent of top films featured white actors and all 20 acting nominations were white, leading to an annual hashtag of #OscarsSoWhite.[39] When Black individuals speak as a collective via Black Twitter, they are able to garner more attention to the dearth of Black representation in film and work to reshape culture in a way that accounts for systematic racism. Nonetheless, it would be inaccurate to view pockets of cultural participation as equal participation—in 2022, the numbers have yet to improve with only one Black nominee.[40]

B. RACE-CONSCIOUS CULTURAL DEMOCRACY THEORY

In part due to its utopic vision of cultural production and reproduction, the implications and limitations of current cultural democracy theory are far reaching and conflict with the actual American cultural ecosystem. Not all speech is interactive—in particular, racist speech limits the interaction of historically oppressed persons in culture. Then, when historically oppressed persons do engage in cultural participation, the appropriation of their speech

37. *Id.*

38. Jill Lepore, *The Invention of the Police*, NEW YORKER (Jul. 20, 2020), https://www.newyorker.com/magazine/2020/07/20/the-invention-of-the-police.

39. Reggie Ugwu, *The Hashtag That Changed the Oscars: An Oral History*, N.Y. TIMES (Sep. 20, 2020), https://www.nytimes.com/2020/02/06/movies/oscarssowhite-history.html.

40. *See* Katey Rich, *Oscar Nominations 2022: Full List of Nominees Led by* The Power of the Dog, VANITY FAIR (Mar. 27, 2022), https://www.vanityfair.com/hollywood/2022/02/oscar-nominations-2022-list.

does not have the positive effect that Balkin espouses. Balkin approaches speech with the false assumption that culture is a singular expression despite its pluralistic nature. Therefore, a race-conscious cultural democracy must recognize and account for the harms of certain speech by instituting methods of control.

Balkin's conception of speech is as interactive disregards the harm that can result from some speech; in particular, within a cultural democracy, speech that is anti-productive fails to be interactive. Some believe that in order to protect the speech of oppressed persons, the speech of the oppressors (or historically dominant group) must also be protected, making counter speech the most effective way to prevent racist speech.[41] In other words, by protecting both sides' right to speech, a dialogue is created between the two.[42] However, this stance is unpersuasive because, when it comes to racist speech, dialogue is not productive and therefore is not interactive. For example, the speech espoused by the KKK is intended to incite fear and silence the voices of African Americans, and therefore it does not promote dialogue between members of society. Fearmongering impedes healthy dialogue and debate. In addition, the dialogue between the KKK and African Americans is not productive because the former seeks to not only delegitimize but also *dehumanize* the latter.

Furthermore, even unintentional racist speech has a capacity of harm that prevents it from being interactive as intended under Balkin's theory. A general influx in racist speech online has routinely been shown to result in an influx of real world violence, regardless of the intent of any one racist speaker.[43] Studies have also shown that repeated exposure to racist speech takes a mental toll on individual members of a historically oppressed group regardless of the intention of the original speaker.[44] In particular, historically oppressed persons

---

41. Matsuda, *supra* note 14, at 2326-27 n. 36.

42. *See id.* at 2357.

43. *See e.g.,* Katarzyna Bojarska, *The Dynamics of Hate Speech and Counter Speech in the Social Media: Summary of Scientific Research*, CTR. FOR INTERNET AND HUM. RTS., https://cihr.eu/wp-content/uploads/2018/10/The-dynamics-of-hate-speech-and-counter-speech-in-the-social-media_English-1.pdf (last visited Sep. 30, 2022); *Hate Speech on Twitter Predicts Frequency of Real-life Hate Crimes*, NYU (Jun. 24, 2019), https://www.nyu.edu/about/news-publications/news/2019/june/hate-speech-on-twitter-predicts-frequency-of-real-life-hate-crim.html; Rachel Hatzipanagos, *How online hate turns into real-life violence*, Wash. Post (Nov. 29, 2018), https://www.washingtonpost.com/nation/2018/11/30/how-online-hate-speech-is-fueling-real-life-violence/; *In Germany, online hate speech has real-world consequences*, ECONOMIST (Jan. 12, 2018), https://www.economist.com/graphic-detail/2018/01/12/in-germany-online-hate-speech-has-real-world-consequences; Melissa Fisher Paoni, *Understanding the harm in hate speech*, ILL. TIMES (Nov. 8, 2018), https://www.illinoistimes.com/springfield/understanding-the-harm-in-hate-speech/Content?oid=11447249.

44. Bojarska, *supra* note 43.

have to be constantly prepared to encounter racist speech, and because racist speech functions on a spectrum, in a variety of forms, such preparedness requires a heightened level of vigilance.[45] This translates to increased stress, known as "minority stress," which can in turn lead to adverse health conditions such as anxiety and depression.[46] Therefore, repeated engagement with racist speech is not only ineffective at controlling it but also increases the physical and mental harm done to historically oppressed persons. As a result, it fails to be interactive within the scope of Balkin's definition of speech in a cultural democracy.

Balkin's conception of speech as appropriative is equally harmful because it ignores the dynamic interplay between different cultures. When there is a power imbalance between the appropriator and those being appropriated and the appropriator profits from their actions, the harm done is incongruous with Balkin's conception of cultural democracy. For example, in 2014, fashion brand DKNY launched "Urban Fabulous" at New York Fashion Week, featuring white women with laid edges.[47] Not only was the use of "urban" coded language for Black, but the hairstyles themselves, historically deemed "ghetto," were reimagined to be a signifier of high fashion when donned by white women.

This is not to say all appropriation is bad; to the contrary, conversation between two cultures can enhance the lived experience of people in both and particularly those that live at their intersection. For example, in 2020, a white TikTok user posted a video asking Black users for methods of controlling her baby hairs in a way that did not appropriate the storied history behind laid edges.[48] Within the Black and Latina communities, laying edge has gone beyond a way of taming baby hairs and has been transformed into art—a form of speech between Black and Latina women, expressed through complex swirls and swoops.[49] The response from Black TikTok was entirely supportive,

---

45. *Id.*

46. *See e.g.*, *What is Minority Stress?*, CNTR. FOR COMMUNITY PRAC., https://www.urccp.org/article.cfm?ArticleNumber=69 (last visited Mar. 3, 2021); Michael Dentato, *The minority stress perspective*, AM. PSYCHO. ASS'N (APR. 2012), https://www.apa.org/pi/aids/resources/exchange/2012/04/minority-stress.

47. Kara Brown, *The problem with Baby Hairs, 'Urban' and the Fashion Industry,* JEZEBEL (Sep. 17, 2014), https://jezebel.com/the-problem-with-baby-hairs-urban-and-the-fashion-indu-1635947700.

48. Alex Lasker, *People are loving this TikToker who asked about a hairstyle before potentially appropriating it*, IN THE KNOW (Nov. 13, 2020), https://www.intheknow.com/2020/11/13/tiktok-baby-hairs/.

49. Ali King, *Four Celebrity Natural Hairstylists Share Their Go-To Baby Hair Tips and Tricks*, VOGUE (Aug. 6, 2019), https://www.vogue.com/article/celebrity-natural-hairstylists-baby-hairs-tips-yara-shahidi.

providing her with product suggestions and tips to manage her "shorty hairs."[50]

The cultural exchange in the DKNY example is invariably different from the TikTok example because (1) there is a greater power imbalance between the Black/Latina communities and the high fashion sphere that has routinely denied them a role in setting beauty standards and (2) DKNY sought to profit off their culture. As a result, while not all speech is appropriative, it can very easily become so when it profits off the cultures of historically oppressed groups.

If the foundation of cultural democracy is speech, then cultural democracy theory must account for the anti-interactive nature and harmful appropriation of certain speech. This is especially true when it comes to racist speech and appropriation across racial lines. A race-conscious approach to cultural democracy realizes this complex interplay between race and cultural production. Therefore, while it may seem anti-democratic and counterintuitive to control some speech to promote cultural democracy overall, this must be the case. By controlling racist speech, historically oppressed persons can more freely engage in their own cultural production without worry of harmful appropriation, and they are empowered to interact across cultures in a productive way.

III. DEFINING RACIST SPEECH

As cultural democracy theory recognizes, speech is the bread and butter of social interaction and thus embodies the complicated nature of social relationships. Race, ethnicity, gender, sexuality, religion, place of origin, etc., all affect social relationships and therefore are fundamental to speech. However, given the various histories that mold the identities of different historically oppressed groups, it is important to consider them with their own unique histories at the forefront. In other words, while a discussion on race can reciprocally inform a discussion on gender, the Venn diagram does not always overlap, and to cast a net wide enough to encompass all historically oppressed groups would necessarily dilute the particular experiences of any one group.[51] Therefore, in order to think about what a more equitable cultural democracy theory would look like, racist speech is taken as the proxy for this Note, but as Part V will discuss, once the framework of a race-conscious

50. Lasker, *supra* note 48.

51. *See* Kimberlé Crenshaw, *Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color*, 42 STANFORD L. REV. 1241, 1242 (1991) ("The problem with identity politics is not that it fails to transcend difference, as some critics charge, but rather the opposite—that it frequently conflates or ignores intragroup difference.").

cultural democracy has been established, it can be an effective tool for examining other types historically oppressed groups and hate speech directed toward them.

A. THE SPECTRUM OF RACIST SPEECH

Racist speech functions on a spectrum, making it difficult to define. A conceptualization of the spectrum of racist speech places clear threats of racial violence on one end and covert, sanitized speech on the other. In the middle lies an ambiguous zone with historically laden symbols, slurs, and overt microaggressions.

Table 1: Spectrum of Racist Speech

Direct Calls for Violence	Historically laden Symbols and Slurs	Overt Microaggressions	Covert, Sanitized Speech

Following the events in Kenosha, a lawsuit was filed against Facebook for the failure to uphold its Community Standards on militia groups on behalf of Hannah Gittings (life partner of Anthony Huber), Christopher McNeal (allegedly harassed by militia members), Carmen Palmer (allegedly threatened by militia members), and Nathan Peet (present at one of the killings).[52] Calls for violence—such as those used by the militia groups targeting African Americans and advocating for their and their allies' death—are the easiest to identify and therefore should be the easiest to control.

On the other side of the spectrum, is covert, sanitized speech, and the rhetoric of the CEO of Wells Fargo, Charles Scharf, is a salient example. Scharf stated in a memo that "while it might sound like an excuse, the unfortunate reality is that there is a very limited pool of black talent to recruit from."[53] While this rhetoric plays into and reinforces the stereotype of African Americans as uneducated and poor, it is too deep into nuanced waters for an enforcer to attempt to regulate it. This type of covert, sanitized speech is a harm commonly remedied by public accountability alone. In the case of Wells

52. Gettings v. Mathewson, No. 2:20-cv-1483 (E.D. Wis. filed Sept. 22, 2020); *see also* Jeff Richgels, *4 file federal lawsuit against Facebook, Kyle Rittenhouse, militia groups in wake of Kenosha protest shootings*, Wisc. St. J. (Sep. 24, 2020), https://madison.com/wsj/news/local/crime-and-courts/4-file-federal-lawsuit-against-facebook-kyle-rittenhouse-militia-groups-in-wake-of-kenosha-protest/article_646089ba-dbec-5972-9e8b-acc9ddac2559.html.

53. Imani Moise, Jessica DiNapoli, and Ross Kerber, *Exclusive: Wells Fargo CEO ruffles feathers with comments about diverse talent*, REUTERS (Sep. 22, 2020), https://www.reuters.com/article/us-global-race-wells-fargo-exclusive/exclusive-wells-fargo-ceo-ruffles-feathers-with-comments-about-diverse-talent-idUSKCN26D2IU.

Fargo, Scharf's speech resulted in boycotts and people withdrawing their money from the bank in favor of Wells Fargo's competitors, in particular Black-owned banks.[54]

However, there is a middle ground between a racially motivated militia group's call to action and a CEO's inability to understand the complexity of race in hiring pools. This middle ground is composed of historically laden racist symbols, slurs, and microaggressions. An example of a historically laden symbol is a burned cross left in the yard of an African American family's home.[55] An example of a microaggression is an African American interviewee being informed by her interviewer that her hair would be more professional if it were straightened.[56]

Even within these periphery categories of racist speech, there is often a blurring that reinforces the assertion that racist speech must be viewed as a spectrum rather than distinct buckets of types of speech. In particular, microaggressions often worm their way into covert, sanitized speech just as historically laden racist symbols and slurs meld into direct calls for violence. For example, an ad agency refusing to include African Americans in a campaign because their hair does not convey the idea of luxury would lie somewhere between a microaggression and covert, sanitized speech.

B. CRITICAL RACE THEORY DEFINITION OF RACIST SPEECH

There is no universal definition of hate speech under U.S. law,[57] nor a definition of racist speech. Commonly, racist speech is defined under an umbrella definition of hate speech, which is "speech, writing, or nonverbal communication that attacks, threatens, or insults a person or group on the basis of national origin, ethnicity, color, religion, gender, gender identity, sexual orientation, or disability."[58] However, a definition like this provides little

54. Isabella Jibilian, *People are calling for a boycott of Wells Fargo after CEO blames 'very limited pool of Black talent' for the bank's trouble hitting its diverse hiring goals*, Business Insider (Sep. 23, 2020), https://www.businessinsider.com/wells-fargo-boycott-ceo-charles-scharf-black-talent-hiring-comments-2020-9.

55. *See* Dennis Romero, *Virginia man accused of burning cross on Black neighbors' front lawn*, NBC NEWS (Jun. 27, 2020), https://www.nbcnews.com/news/us-news/virginia-man-accused-burning-cross-black-neighbors-front-lawn-n1232345.

56. *See* Chanté Griffin, *How Natural Black Hair at Work Became a Civil Rights Issue*, JSTORE DAILY (Jul. 3, 2019), https://daily.jstor.org/how-natural-black-hair-at-work-became-a-civil-rights-issue/.

57. "Hate Speech and Hate Crime," AM. LIBR. ASS'N, http://www.ala.org/advocacy/intfreedom/hate (last visited Sep. 30, 2022).

58. "Hate Speech," DICTIONARY.COM, https://www.dictionary.com/browse/hate-speech (last visited Sep. 30, 2022).

guidance in differentiating racist speech along the spectrum in a way that controls the most problematic, unwanted racist speech.

Nonetheless, if certain speech is to be systematically controlled as necessary within a race-conscious cultural democracy, then a clearer line must be drawn between racist speech that should be controlled and speech that, while still unacceptable, should not be controlled. In other words, actionable racist speech must be separated from unactionable racist speech. And while communities can still hold accountable speech that is unactionable (as in the case of the Wells Fargo CEO), racist speech rising to the level of actionable speech requires a heightened level of control. Such control can be brought by corporate oversight in addition to public accountability, as discussed later in Part V.[59]

Mari Matsuda provides two workable definitions for racist speech—one qualitative and the other engrained in distinct, value-laden elements.[60] The first definition Matsuda puts forward is that racist speech presents "an idea so historically untenable, so dangerous, and so tied to perpetuation of violence and degradation of the very classes of human beings who are least equipped to respond that it is properly treated as outside the realm of protected discourse."[61] However, this first understanding of racist speech does not easily cut through the contextual ambiguity that often shrouds racist speech. Matsuda's other definition of racist speech, however, has three well-defined characteristics: it is a message (1) "of racial inferiority" (2) "directed against a historically oppressed group" (3) that is "persecutorial, hateful, and degrading."[62] This second definition can be utilized to cut through the ambiguity of racist speech.

While this definitional line in the sand oversimplifies the interwovenness of racist speech outlined in Section III.A, it at least casts a net that lands between direct calls for violence and covert, sanitized speech. Furthermore, Matsuda's conception of racist speech in this way also recognizes the importance of context, as not all speech verbal or written. In the case of symbols and imagery especially, context is important, and in these cases, "if the historical message, known to both victim and perpetrator, is racist persecution, then the sign is properly treated as actionable racist speech."[63] By using Matsuda's definition of racist speech, speech that works against a race-conscious cultural democracy because it is anti-interactive and problematically

59. *Infra* Part V.
60. Matsuda, *supra* note 14 at 2357.
61. *Id.*
62. *Id.*
63. *Id.* at 2365-66.

appropriative can be controlled, allowing historically oppressed persons to more freely and widely engage in cultural production.

Furthermore, under Matsuda's conception of actionable racist speech, expressions of hatred directed toward the oppressor—such as when Malcolm X spoke of the white devil—would not be restricted.[64] This language only arises as a byproduct of racism against African Americans, and while offensive, those targeted are a part of a historically non-oppressed group and are able to more easily retreat and reaffirm their personhood.[65] Conversely, the rhetoric espoused by neo-Nazi leader David Duke would fall within prohibited speech. In other words, so long as there is an unequal distribution of power along race lines, the harm done by the speech of the oppressed will never rival the harm done by the speech of the oppressor, and by limiting the oppressor's racist speech, the cultural participation of historically oppressed can not only increase but prosper.[66]

C. FACEBOOK'S DEFINITION OF RACIST SPEECH

Facebook has used its Terms of Service (ToS) and Community Standards to delineate the types of speech unacceptable on its platform. The platform defines racist speech under the broader umbrella of hate speech and provides two main elements for objectionable speech.[67] First, it must be a "direct attack," defined as "violent or dehumanizing speech, harmful stereotypes, statements of inferiority . . . calls for exclusion or segregation."[68] Second, it must target people with "protected characteristics," with race being one such characteristic.[69] A direct attack on persons with protected characteristics aligns

---

64. *Id.* at 2362.

65. *Id.*

66. Furthermore, returning to the DKNY example from earlier, the coded language of "urban fabulous" would not be considered racist speech under Matsuda's definition because of a failure to meet the second prong (directed against a historically oppressed group). "Urban fabulous" teeters too far over the line into cultural appropriation. This is harmful because it effectively erases the cultural origin, history, and meaning behind laid edges, thereby negatively affecting the cultural engagement and production of African Americans. Cultural appropriation is typically driven by ignorance and is often more focused on the benefit for the appropriating culture rather than the harm done to the appropriated culture. *What is cultural appropriation and how can you spot it?*, THE WEEK (Jun. 1, 2020), https://www.theweek.co.uk/cultural-appropriation. As a result, this type of cultural speech lies closer toward covert, sanitized speech on the spectrum of racist speech, making it far harder to systematically control.

67. PART III: OBJECTIONABLE CONTENT, FACEBOOK COMMUNITY STANDARDS, https://www.facebook.com/communitystandards/objectionable_content (last visited Mar. 3, 2021).

68. *Id.*

69. *Id.*

with the first and third elements of Matsuda's definition—a message of racial inferiority that is persecutorial, hateful, and disregarding.[70]

However, without that second element (directed against a historically oppressed group), Facebook's definition stands open to abuse and fails to rectify the biggest concern of race-conscious cultural democracy: that racist speech silences historically oppressed persons, thereby limiting their cultural participation. Returning to the example of Malcolm X versus David Duke, if Facebook does not clarify the difference between the speech of the oppressed from the speech of the oppressor, then both Malcolm X and David Duke's speech would be removed from Facebook. This would bring to fruition the fear that attempts to control speech will be disproportionately affect the people they are meant to protect.[71] By specifying the group that a definition of racist speech is intended to protect, this type of coopting can be remedied. Therefore, Facebook should amend its definition to read "we define hate speech as a direct attack on people *of a historically oppressed group* based on what we call protected characteristics."

IV. A HYBRID MODEL OF SPEECH CONTROL FOR THE DIGITAL AGE

Reexamining how Facebook defines actionable racist speech is only the first step in working toward a race-conscious cultural democracy where all members can freely and widely engage in cultural production. A system must be put in place to ensure racist speech is controlled. To that end, the community, as the most democratic option, must be the first line of defense. However, relying solely on the community to govern speech inevitably stands open to criticism of cancel culture and inability to effectuate substantive change. Therefore, there needs to be a second layer of protection like corporate oversight; and social media platforms like Facebook are well positioned to control racist speech online.

This Part examines how the community functions to control speech—historically and online today—and then looks to how Facebook already allows its community to manage racist speech. It then turns in Section IV.B to how corporate oversight can reinforce the community in order to support a race-conscious cultural democracy online. And finally, it uses two of Facebook's

70. Matsuda, *supra* note 14 at 2357.

71. Cindy Cohn, *When Limiting Online Speech to Curb Violence, We Should be Careful*, WIRED (Aug. 9, 2019), https://www.wired.com/story/when-limiting-online-speech-to-curb-violence-we-should-be-careful/. Often, policies aimed at controlling racist speech disproportionately affect the speech of historically oppressed persons.

Oversight Board's first decisions to examine how corporate control might successfully work in practice.

A. ON THE FRONTLINE: THE ROLE OF THE COMMUNITY

The most democratic means of speech control is relying on the community, and when laws are not enough to adequately control racist speech—such as in the United States under the First Amendment—the community must take a larger role in controlling speech online. And given the vast amount of speech posted online, it is a matter of practicality to first rely on the on-the-ground readers and watchers to determine what speech is unwanted. This way, it becomes easier for an overseeing body to institute systematic change based upon the community response. Furthermore, this is something communities are already doing under different methods of public accountability, and it has the power to positively affect the lives of historically oppressed persons.

1. In Theory: Benefits and Detriments of Community Speech Control

The benefits of public accountability are often overshadowed by cancel culture rhetoric as a modern tyranny of the majority.[72] Unsurprisingly though, the term "cancelling" itself was an internet meme on Black Twitter that was coopted by journalists, politicians, celebrities, and the like to misconstrue community accountability into feared censorship as a means of "delegitimiz[ing] the dissension that echoes from society's margins."[73] As a result of this cancel culture rhetoric, while public accountability is an important mechanism for racist speech control online, its success is limited, and its effectiveness must be amplified by additional measures like that of corporate control as discussed in Section IV.B.

Public accountability is nothing new. Historically, public accountability manifested as public shaming. For example, in the medieval era, there was the pillory, a wooden framework with three holes intended to restrain someone's head and wrists, and those most often subjected to the pillory include perjurers, forgers, counterfeiters, and blasphemers.[74] While someone might not be displayed on the pillory for more than a couple hours, it was located in

---

72. Edwin J. Feulner, *Preventing "The Tyranny of the Majority,"* THE HERITAGE FOUNDATION (May 7, 2018), heritage.org/conservatism/commentary/preventing-the-tyranny-the-majority.

73. Meredith D. Clark, *DRAG THEM: A brief etymology of so-called "cancel culture"*, 5 (3-4) COMM AND THE PUB. (COMMENT. FOR SPECIAL F.) 88, 89 (2020).

74. *Pillory*, ENCYLOPEDIA.COM, https://www.encyclopedia.com/history/modern-europe/british-and-irish-history/pillory (last visited Mar. 23, 2021).

a central, busy part of town to allow for the most visibility.[75] While often the result of some judicial proceeding, the crux of the pillory's effectiveness was audience engagement. People would often throw whatever they had on hand at those subjected to the pillory, from rotten food to rocks. The effects of being sentenced to the pillory went beyond the couple hours someone spent berated by passersby.[76] Especially in small towns, where reputation was everything, a pillory sentence could result in job loss and crushed marriage prospects—effectively, a social cancellation.[77]

For modern comparison, the drama within YouTube's beauty influencer, or gurus, community is riddled with prime examples of public accountability. There have been numerous "[enter name] is over parties," where an influencer's racist comments are exposed—typically on Twitter—followed by a massive unfollowing spree.[78] In the case of "Dramageddon I" in 2018, numerous beauty gurus were exposed for their previous racist tweets.[79] For example, Gabriel Zamora, Nikita Dragun, and Laura Lee all had racist tweets resurface from 2012.[80] Public uproar akin to the medieval pillory proceeded, minus the literal rotten food and rocks. Lee in particular lost half a million of her 4.5 million subscribers within the month following the exposé. She lost an estimated $25,000–$65,000 in income from viewership alone, in addition to losses in lucrative partnerships and brand deals.[81]

Dramageddon I highlights two main concerns critics have with "cancel culture." First, it can happen to anyone at any time for anything. In particular, cancel culture revolves around the belief that no material is too old; any previous tweet can become the basis for a current cancellation, and the idiom "the Internet never forgets" stands to be abundantly true. It raises the question if some things are better left forgotten, as the cultural climate of 2012, some argue, was different than that of 2018.[82] As a result, cancel culture can create a

---

75. Mathew Green, *A Grim and Gruesome History of Public Shaming in London: Part 1*, LONDONIST (Jan. 19, 2017), https://londonist.com/2015/12/publicshaming1.

76. *Pillory*, BRITANNICA (Oct. 1, 2022) https://www.britannica.com/topic/pillory-penology.

77. *Id.*

78. Lilly Gelman, *Twitter Explained: Who is Over and Where's the Party?*, MOMENT (Aug. 15, 2015), https://momentmag.com/twitter-explained-who-is-over-and-wheres-the-party/.

79. Alex Abad-Santos, *Laura Lee, Jeffree Star, and the racism scandal upending the YouTube Beauty community, explained*, VOX (Aug. 31, 2018), https://www.vox.com/2018/8/28/17769996/laura-lee-jeffree-star-racism-subscriber-count.

80. *Id.*

81. *Id.*

82. It is worth noting in 2012, racism was as unacceptable as it was in 2018. 2012, the publication year of most of the racist tweets concerned with in Dramageddon I, is the same year a national uproar occurred following the death of Trayvon Martin. Therefore, it was not

sense of panic because anything said in the past can be brought up as problematic in the present and because anything said now might become a point of contention in the future.[83]

Second, in addition to the fear incited by cancel culture, there are concerns that it is not an effective method of changing racist behavior. For example, Nikita Dragun, one of the influencers exposed during Dramageddon I, has repeatedly been called out for racially problematic behavior since 2018—blackfishing in particular.[84] Yet despite flooded comment sections and trending on Twitter numerous times for her behavior, Dragun's Instagram is still composed of an array of artificial skin tones. In fact, she has even leaned into the criticism by tweeting "which race is nikita gonna be today ?" in October 2020.[85]

For another example of cancel culture's perceived ineffectiveness, in 2019, Shane Gillis was fired from *Saturday Night Live (SNL)* over previous racist comments.[86] Then, in a statement Gillis released soon after his removal from *SNL*, Gillis failed to reckon with why the public sought to hold him accountable, equating racism with pushing boundaries and taking risks.[87] Like Laura Lee's, Gillis's cancellation is largely reminiscent of the effects of the

---

racism that changed but people's ability to put into action effective counters for that speech. People's ability to use social media as a tool of collective action to shed light on racism immensely improved over those six years.

83. Austin M. Hooks, Cancel Culture: Posthuman Hauntologies in Digital Rhetoric and the Latent Values of Virtual Community Networks 16 (Aug. 2020) (M.A. thesis, University of Tennessee at Chattanooga). Furthermore, the panic connected with cancel culture has been problematically associated with a fear of actual harm, "adding a neologic twist on the origin of the practice by associating it with an unfounded fear of censorship and silencing." Clark, *supra* note 73 at 89.

84. Blackfishing is when non-black women use fake tanners, dark makeup, and sometimes wigs to appear black or at least racial ambiguous. Emma Nolan, *Who is Nikita Dragun? YouTuber Under Fire For Race Remarks*, NEWSWEEK (Oct. 5, 2020), https://www.newsweek.com/nikita-dragun-youtuber-racist-tweets-instagram-blackfishing-1536416/. Even though blackfishing, like the DKNY example earlier in this Note, would not meet Matsuda's definition for actionable racist speech, social media users' inability to alter Nikita's behavior is still a salient example of how cancel culture is viewed as ineffective.

85. Sophia Thompson, *Nikita Dragun called out for insensitive tweet mocking her blackfishing past*, PopBuzz (Oct. 6, 2020), https://www.popbuzz.com/internet/youtubers/nikita-dragun-blackfishing-tweet-racism/.

86. Jacob Stolworthy, *SNL star Shane Gillis' firing critised as 'cancel culture' by Bill Burr and Jim Jeffries*, INDEPENDENT (Sep. 17, 2019), https://www.independent.co.uk/arts-entertainment/tv/news/shane-gillis-snl-saturday-night-live-fired-bill-burr-jim-jeffries-a9108071.html.

87. Following his removal from SNL, Gillis tweeted a statement saying, "I'm a comedian who pushes boundaries. I sometimes miss. If you go through my 10 years of comedy, most of it is bad, you're going to find a lot of bad misses. I'm happy to apologize to anyone who's actually offended by anything I've said. My intention is never to hurt anyone but I am trying to be the best comedian I can be and sometimes that requires risk." *Id.*

pillory—he suffered immediate societal attention and even financial loss. However, when Gillis and other members of the comedy profession do not confront the harm done by the racism they perpetuate, is social condemnation and stifled streams of income a solution or no more than a temporary, individualized Band-Aid?

As a result, the end goal of public accountability is sometimes dubious. If it is to remove someone's platform in order to mitigate the damage done by their racism, there is nothing stopping them from finding new avenues, such as creating new accounts or switching social media platforms entirely, or simply reoffending as in the case of Nikita Dragun. If it is to cut off income streams as a form of retribution, it is unlikely to elicit changed behavior absent conversation on why the language or behavior was problematic in the first place, as in the case of Shane Gillis.

However, public accountability cannot be so easily dismissed as an effective tool of controlling racist speech online. Rather, if the end goal of public accountability is refocused from changing the offender to minimizing the impact felt by those affected, then it can be a positive tool in controlling racist speech online. In particular, the act of holding someone accountable can allow for those in a historically oppressed group to reclaim power, and this is evident in displays of public accountability within Black, Latinx, and LGBT communities.

For instance, "reading," the open critique of someone's appearance, within drag culture is a prime example of productive public accountability. Reading dates back to queens of color in the 1950s but was mainstreamed by the film *Paris is Burning*, a documentary that illuminated the New York City drag scene and house culture in the 80s.[88] It has also become a reoccurring feature on the popular television show *RuPaul's Drag Race*.[89] Reading is usually done in a dramatic, rhythmic way; and yet, at the heart of reading is unconditional community love and awareness.[90] It is less of an insult and more of an act of displaying each other's flaws before the community and proclaiming love regardless of those flaws for not just the individual being read but the community as a whole. In formal sessions, as in the case of *RuPaul's Drag Race*, an announcer begins by "opening the library," making reading a communal

88. PARIS IS BURNING (Jennie Livingston and Barr Swimar 1991).

89. *RuPaul's Drag Race* (VH1).

90. Ted Mentele, *The history of drag queen slang*, READER'S DIGEST, https://www.readersdigest.co.uk/inspire/life/the-history-of-drag-queen-sang (last visited Oct. 1, 2022).

act, where even the person read is in on the activity.[91] Reading highlights the potential for accountability to strengthen a community overall.

Public accountability was also a primary driver for boycotts in the Civil Rights Era.[92] In December of 1955, the Montgomery Improvement Association with Martin Luther King, Jr. at its helm led the Montgomery bus boycott, which led to *Bowder v. Gayle*, which in turn ended segregation on buses.[93] In King's memoir *Strides Toward Freedom*, King highlights the importance of the "growing self-respect" within the Black community that came with the boycotts.[94] Over 60 years later, communities still rely on boycotts as a means of public accountability as a reflection of self-respect—as was the case for Goya in 2020. When Robert Unanue, Goya's CEO, praised former president Donald Trump, Goya's largely Latino consumer base withdrew support given "everything…going on with [the Trump] administration and the border, the family separations and DACA."[95]

Effectively, the tools that historically oppressed groups have been using to assert their own voice have remained the same—the read, the boycott, the cancel. Public accountability is driven by the collective experience of targeted historically oppressed persons as they discuss, morally evaluate, and prescribe remedies for harm done unto them.[96] Furthermore, as put by Professor Meredith Clark, "Not every critique can come wrapped up in niceties and polite speech. Nor should it. Sometimes, the urgency and weight of oppression requires us to immediately cry out."[97]

Nonetheless, the potential for cancel culture to run awry has led some to support an alternative form of public accountability: the *call-in*. The call-in is "a call-out done with love."[98] It recognizes that some conflicts can be resolved privately, rather than in the public pillory or in the Twitter trending page.

---

91. *Id.*; Gary Hartley, *A Beginner's Guide to Drag Terminology*, CAPE TOWN MAGAZINE, https://www.capetownmagazine.com/arts-culture/a-beginners-guide-to-drag-terminology/104_22_19320#:~:text=A%20subtler%20form%20of%20reading,%2C%E2%80%9D%20explains%20Manila%20von%20Teez (last visited Oct. 1, 2022).

92. *Id.*

93. *Montgomery Bus Boycott*, STANFORD: THE MARTIN LUTHER KING, JR. RESEARCH AND EDUCATION INSTITUTE, https://kinginstitute.stanford.edu/encyclopedia/montgomery-bus-boycott (last visited Mar. 25, 2021).

94. *Id.*

95. Raul A. Reyes, *Latinos boycotting Goya say it's not about politics. It's about standing against Trump's 'hate'*, ABC NEWS (Jul. 18, 2020), https://www.nbcnews.com/news/latino/latinos-boycotting-goya-say-it-s-not-about-politics-it-n1234052.

96. Clark, *supra* note 73 at 89.

97. *Id.*

98. Loretta Ross, *I'm a Black Feminist. I Think Call-Out Culture Is Toxic.*, N.Y. TIMES (Aug. 17, 2021), https://www.nytimes.com/2019/08/17/opinion/sunday/cancel-culture-call-out.html.

Furthermore, it supports the belief that open conversations and debate allow people the opportunity for growth and change.[99] This argument is familiar—it is the same one used to justify robust free speech rather than control of racist speech.[100] As a result, the counter remains the same: the harm caused by racist speech is such that it dissuades historically oppressed persons from engaging in culture production. The harm done is felt physically, as in the case of Kenosha, but it is also felt psychologically as evident through studies on minority stress.[101] Therefore, while open conversation, especially between private individuals, is one effective method of accountability for racist speech, it is not the only one, nor should it be. Public accountability not only tells the individual wrongdoer that their racist speech is unwanted but it tells others who might favor similar speech that that speech is unwelcome within the cultural ecosystem.

2. *In Practice: Facebook's Current Community Speech Control Regime*

Facebook users are the producer-consumers of culture on the platform and are therefore best suited to be the first line of defense when it comes to evaluating racist speech that warrants control. Utilizing its own definition of racist speech outlined in Section III.C, Facebook uses multiple approaches to monitor speech: user reporting, artificial intelligence, and third-party fact-checking. This Section discusses the first approach (user reporting), leaving the latter two for evaluation in Section IV.B.2. User reporting is the bread and butter of community control of racist speech; however, the crux of Facebook's current regime falls on the platform's lack of trust in its own community.

When a user encounters offensive speech on Facebook, there are three ways that the content can be removed. First, a user can directly delete a racist comment on their post. However, when the user deletes the comment, they are not presented with an opportunity to also report it to Facebook. In addition, the user can only report the offender via the offender's profile for "harassment and bullying" or "posting inappropriate things."[102] At no point is the user allowed to flag the offender for racism or expand on their justification for reporting besides these broad categories. After deleting the comment or reporting the offending user via their profile page, the user can either block, unfollow, or unfriend the offending user.

---

99. *Id.*

100. *Id.*

101. *See* note 52 and accompanying text.

102. Reporting Form, FACEBOOK, https://www.facebook.com/ (The reporting form can be located following the steps described within the text of this Note.) (last visited Oct. 1, 2022).

The second way racist content can be removed is if a third-party user sees a racist comment on another's post and reports it for review by Facebook. On the report screen, the user is presented with various problem options, including hate speech. The user is only allowed to select one problem, so if the content in question is both harassment and hate speech (as is common), the reporting user must decide which problem to file the content under. If the user selects hate speech, the user is able to clarify that the speech in question is related to race. Then, the reporting user is presented with an abridged version of Facebook's Community Standards, stating:

> We only remove content that directly attacks people based on certain protected characteristics. Direct attacks include things like:
>
> - Violent or dehumanizing speech
>
> For example, comparing all people of a certain race to insects or animals
>
> - Statements of inferiority, disgust or contempt
>
> For example, suggesting that all people of a certain gender are disgusting
>
> - Calls for exclusion or segregation
>
> For example, saying that people of a certain religion shouldn't be allowed to vote[103]

Notably, while Facebook includes harmful stereotypes in its Community Standards as a direct attack, it is left out of the above list presented to users upon reporting. After reviewing the definition provided of a direct attack, the user can either select "I don't know. I'd like to see other steps I can take." or "Yes, I'd like to continue filing this report."[104] The first option will lead the user to other steps they can take, including blocking the offending user or hiding all their posts. The second option will file the report with Facebook.

The third way racist content can be removed on Facebook is through group administrators. Within Facebook groups, users can report racist speech directly to the administrators rather than going through the corporation itself. Once the content is reported, an admin can remove it and the offender from the group. A user still has the opportunity to report the content via Facebook's reporting system outlined above. However, reporting to a group administrator will likely receive a quicker response than reporting to Facebook itself. In addition, a reporting user is likely more interested in the speedy removal of the

103. *Id.*
104. *Id.*

particularly offensive content and the offender than in improving the overall environment on the platform. Unless a user or administrator reports directly to Facebook, however, the offender is still capable of joining other groups and committing similar offenses, all while staying below Facebook's radar.

In order for the community to have a stronger systematic effect, Facebook must (1) improve the toolkit it provides users and (2) modify its current definition of racist speech to align with critical race theory, as discussed in Section III.B. Doing so would allow Facebook to have more trust in the decisions made by the community, which in turn would result in a more race-conscious environment on the platform overall.

Facebook should institute better tracking of users that are removed from groups over Community Standards violations like racist speech. This way, an offending user cannot repeatedly jump groups and commit similar violations without Facebook being notified. Then, to provide heightened transparency, the offending user should be notified of how many particular strikes they have against them; as of now, Facebook does not preemptively notify users when their content has been removed, let alone why.[105]

However, Facebook administrators cannot be the only vehicle for community control given the repeated issues Facebook has had with groups themselves that violate community guidelines. When 455 users reported the Kenosha Guard to Facebook, it slipped through Facebook's control mechanisms, despite the group blatantly violating Community Standards.[106] Moreover, the Kenosha Guard incident also came a year after Facebook claimed it was clamping down on groups by removing the option to create secret groups and limiting groups to be either public or private.[107] Therefore, more work needs to be done to properly monitor groups. For example, Facebook should institute a screening page when someone seeks to create a group, reminding them of the Community Standards and prohibition on certain kinds of groups like militias.

Facebook should also modify its current definition of racist speech to ensure that only historically oppressed persons are explicitly protected as

---

105. This is also a policy recommendation that the Board includes in its decision on case 2020-003-FB-UA. The case and the Board's policy recommendation is discussed in depth in the following Section.

106. Mac & Silverman, *supra* note 8; *Dangerous Individuals and Organizations*, FACEBOOK: COMMUNITY STANDARDS, https://www.facebook.com/communitystandards/dangerous_individuals_organizations (last visited Oct. 1, 2022).

107. Brandy Zadronzy, *Facebook has doubled down on groups–now it's looking to clean them up*, NBC NEWS (Aug. 14, 2019), https://www.nbcnews.com/tech/social-media/facebook-has-doubled-down-groups-now-it-s-looking-clean-n1042336. This article was published August of 2019 with the Kenosha shooting occurring August of 2020.

discussed in Part III. Without Matsuda's element that actionable racist speech be "directed against a historically oppressed group," Facebook's current definition stands open to abuse, protecting both historically oppressed groups *and* historically oppressive groups.

Furthermore, while the current reporting system is robust, it would benefit from providing its users more information. Facebook should add "harmful stereotypes" to its list of justifications for removal in order to more accurately reflect how its Community Standards define a direct attack.[108] It would also be beneficial to have a link to the full Community Standards upon reporting. Facebook utilizes a three-tier system to further define attacks,[109] and many users might not be familiar Facebook's robust definition unless they are presented with it upon reporting.

Another method of increasing community self-regulation would be by increasing the scenarios in which a user can report racist content. For example, when an administrator goes to remove someone from a group, they should be able to simultaneously report that user to Facebook, or when a user deletes a comment from their post, they should also be able to simultaneously report that comment to Facebook.

By implementing methods of increased transparency and by providing more access to information, Facebook can give its users the tools needed to better hold themselves accountable. Doing so would be the most democratic means of controlling speech online. However, it is inevitable that some incidents will slip through the cracks. Having a secondary level of review can help alleviate the fear that some have that they are being silenced for simply being in opposition. In addition, this secondary level of review can also work to reaffirm the community's belief that what is deemed racist is in fact so. This can help reinforce the benefit of public accountability: not only is the reported user made aware that their speech is unwanted, but so are others who might be of similar mind.

B. REINFORCEMENT: THE ROLE OF CORPORATE OVERSIGHT

Platforms are best suited to be the second level of review for content that slips through the cracks of public accountability because of the flexibility they have under the First Amendment and Section 230; and platforms alone—without the assistance of the community—cannot sift through the vast amount of content uploaded online every day. Therefore, only when platforms work with the community, implementing community decisions systematically and

108. *Hate Speech*, FACEBOOK: COMMUNITY STANDARDS, https://www.facebook.com/communitystandards/hate_speech (last visited Oct. 1, 2022).

109. *Id.*

ensuring that content is moderated in alignment with critical race theory, can race-conscious cultural democracy online prosper.

1. *In Theory: Corporate Oversight of Speech Control*

First Amendment jurisprudence does not apply to social media platforms because platforms are private actors. Unless a private actor is found to be exercising "a function traditionally exclusively reserved to the State," it is not obligated to observe another's freedom of speech.[110] For instance, in *Manhattan Cmty. Access Corp. v. Halleck*, the Court held that simply because a public access channel on a cable system provides a space for the speech of the others, it does not mean it is a state actor under the state-action doctrine.[111]

The government could reach users' content on social media platforms by extending the state-action doctrine. In doing so, social media platforms would be regarded as stepping into the shoes of the government, thereby becoming incapable of limiting the speech that occurs on their sites. To some, this would be an easy extension, especially after the strong language Justice Kennedy used in his opinion in *Packingham v. North Carolina*.[112] Justice Kennedy stated:

> A fundamental First Amendment principle is that all persons have access to places where they can speak and listen, and then, after reflection, speak and listen once more. Today, one of the most important places to exchange views is cyberspace, particularly social media, which offers "relatively unlimited, low-cost capacity for communication of all kinds,"…to users engaged in a wide array of protected First Amendment activity on any number of diverse topics. The Internet's forces and directions are so new, so protean, and so far reaching that courts must be conscious that what they say today may be obsolete tomorrow.[113]

However, Justice Alito in his concurrence chalks up Kennedy's above language to "loose rhetoric,"[114] and this language has yet to sway courts in how they approach speech on the Internet. Furthermore, in *Nyabwa v. Facebook*, a federal

---

110. Manhattan Cmty. Access Corp. v. Halleck, 139 S. Ct. 1921, 1926 (2019) (internal quotation omitted). *But see* Marsh v. Alabama, 326 U.S. 501, 507 (1946) (holding that "[private] ownership [of a town] does not always mean absolute dominion" and that "[s]ince [the] facilities are built and operated primarily to benefit the public and since their operation is essentially a public function, it is subject to state regulation.").

111. 139 S. Ct. at 1926.

112. 137 S. Ct. 1730, 1738 (2017) (holding that North Carolina violated a sex offender's right to free speech by preventing them from accessing sites like Facebook and Twitter that permit minors to become members).

113. *Id.* at 1732 (internal citations omitted).

114. *Id.* at 1743 (Alito, J. concurring).

district court in Texas directly held that Facebook was not covered by the state-action doctrine.[115]

In addition to the lack of a state-action doctrine extension, Section 230 of the Communications Decency Act shields social media platforms from liability based on third-party speech. Section 230 states "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[116] As a result, platforms like Facebook cannot be held liable for the content shared on their platform. Without Section 230, platforms would be forced to either not host content at all or to employ heavy censorship in order to shield themselves from liability.[117] Therefore, with Section 230 in place, platforms are able to navigate their own approaches to speech without the fear of liability that would result if First Amendment obligations were extended to online platforms. This breadth of freedom is what would allow a platform like Facebook to take innovative steps to rethink controlling racist speech.

2. *In Practice: Facebook's Approach to Corporate Oversight of Speech Control*

Given the flexibility Facebook has to moderate content, the platform is best fit to adapt to and control racist speech in a way that promotes a race-conscious cultural democracy. Facebook's Oversight Board has the potential to be an important vehicle for promoting this race-conscious cultural democracy on the platform, so long as Facebook gives the Board the leeway to make meaningful change. Of the first six cases the Board reviewed upon creation, two dealt with Facebook's Community Standard section on hate speech.[118] The first case, Case 2020-002-FB-UA, reached the right decision, but it did so for the wrong reasons and has the potential to hinder the flourishment of a race-conscious cultural democracy online. However, Case 2020-003-FB-UA presents a workable rationale that best supports a race-

115. Nyabwa v. Facebook, No. 2:17-CV-24, 2018 U.S. Dist. LEXIS 13981 *2 (S.D. TX Jan. 26, 2018).

116. 47 U.S.C. § 230.

117. CDA 230: THE MOST IMPORTANT LAW PROTECTING INTERNET SPEECH, EEF, https://www.eff.org/issues/cda230 (last visited Mar. 3, 2021).

118. Thus far, the Board has not reviewed any cases on racist speech in particular. The two hate speech cases involved hate speech based on religion and ethnicity. However, the race-conscious culture democracy framework can still be applied to ask whether the right kind of speech is being controlled using Matsuda's three factors of actionable racist speech: a message (1) "of racial inferiority" (2) "directed against a historically oppressed group" (3) that is "persecutorial, hateful, and degrading." This definition can be utilized to cut through the ambiguity of racist speech. Matsuda, *supra* note 14 at 2357. Matsuda herself uses this definition in other contexts such as Anti-Semitism. *See id.* at 2363. The definition is merely changed by inserting "religion" or "ethnicity" in the first element.

conscious cultural democracy online. As a result, so long as the Board continues to approach questions of hate speech in alignment with Case 2020-003-FB-UA, then the Board is a beneficial method of racist speech control in the digital age; it sets an example for other platforms in how to approach speech.

In Case 2020-002-FB-UA, the Board overturned Facebook's decision to remove a Burmese user's allegedly Islamophobic post.[119] In doing so, the Board reached the right conclusion under the race-conscious cultural democracy framework, but it utilized a definition of hate speech different than that advocated in this Note and even different from Facebook's own definition in its Community Standards.

For context, the offensive content included two infamous photos from 2015 of a Syrian toddler who drowned in an effort to reach Europe. Attached to the photo was rhetoric on the psychology of Muslims. The Burmese poster criticized the Muslim response to the treatment of Uyghur Muslims in China in comparison to the attacks that followed the cartoon depictions of the Prophet Muhammad in France. As a result, a driving force of Facebook's decision to originally remove the content was the history of intense Islamophobia in Myanmar. Facebook cited to its Community Standards as still prohibiting hate speech even if it does not advocate or incite violence given "the capacity [of hate speech] to trigger acts of discrimination, violence, or hatred, particularly if distributed widely, virally, or in context with severe human rights risks."[120]

To reach its decision, the Board utilized three "bodies of law": Facebook's Community Standards, Facebook's espoused values, and international law. Under its Community Standards, Facebook acknowledges that hate speech can, *but does not always*, result in real world violence and, as discussed in Part III, defines hate speech as an attack based on protected characteristics.[121] An attack includes harmful stereotypes and "generalizations that state inferiority (in written or visual form) [such as] mental deficiencies[, which] are defined as

119. OVERSIGHT BD., CASE DECISION 2020-002-FB-UA (Jan. 28, 2021), https://web.archive.org/web/20210128145029if_/https://oversightboard.com/decision/FB-I2T6526K/.

120. *Introduction*, FACEBOOK: COMMUNITY STANDARDS, https://www.facebook.com/communitystandards/introduction (last visited May 4, 2021). In particular, the Introduction states: "We are committed to making Facebook a safe place. Expression that threatens people has the potential to intimidate, exclude or silence others and isn't allowed on Facebook."

121. *Hate Speech*, FACEBOOK: COMMUNITY STANDARDS, https://www.facebook.com/communitystandards/hate_speech (last visited May 4, 2021). As noted previously, while Facebook does a lot of work to define and explain its approach to hate speech, it would benefit from requiring that hate speech be directed at historically oppressed persons to ensure that the definition is not abused and used against the people it is intended to protect.

those about: intellectual capacity…education…mental health."[122] However, the Introduction to Facebook's Community Standards highlights that "Voice" is an important value of the platform, which must be balanced against Facebook's value of "Safety."[123] Furthermore, the Board also utilized the UN Guiding Principles on Business and Human Rights, a human rights framework intended for private businesses, and used the international human rights standards for the right to freedom of expression, the right to non-discrimination, and the right to life and security to support its decision.[124]

In reaching its decision, the Board rightfully considered the content in its full context and used the full breadth of its tools as outlined in its bylaws. Using context is imperative when approaching hate speech given the nuances of different cultures. This is especially true for Facebook, which is an American corporation with users all around the world. Such a context-driven approach promotes a race-conscious cultural democracy. In this case, the Board consulted with translators and cultural experts and even took public comments into consideration when reaching its decision.[125] What the case notably highlighted was the importance of translation—while Facebook translated the text as "[i]t's indeed something's wrong with Muslims psychologically," the Board's translators suggested that the text actually read "[t]hose male Muslims have something wrong in their mindset."[126] Though a slight change in language, the Board believed that the latter translation was not derogatory or violent.

In addition, when consulting cultural experts on Myanmar, they found comments on the mental state of Muslims was not a part of typical Islamophobia in the country. Furthermore, the post was within a group that was intended for intellectual and philosophical discussion and meant to be a "commentary pointing to the apparent inconsistency between Muslims' reactions to events in France and in China."[127] Therefore, given this context,

---

122. *Id.*

123. *Supra* note 120.

124. *See* International Covenant on Civil and Political Rights art. 19, Dec. 19, 1966, 999 U.N.T.S. 171, 175 (entered into force Mar. 23, 1976) [hereinafter ICCPR]; ICCPR Human Rights Comm., General Comment No. 34 on Article 19, U.N. Doc. CCPR/C/GC/34 (2011); United Nations High Commissioner for Human Rights, *Annual Report of the United Nations High Commissioner for Human Rights*, U.N. Doc A/HRC/22/17/Add.4 (Oct. 05, 2012); David Kaye (Special Rapporteur), Promotion and Protection of the Right to Freedom of Opinion and Expression, U.N. Doc. A/74/486 (2019); ICCPR art. 6 and art. 9.

125. OVERSIGHT BD., CASE DECISION 2020-002-FB-UA (Jan. 28, 2021), https://web.archive.org/web/20210128145029if_/https://oversightboard.com/decision/FB-I2T6526K/.

126. *Id.*

127. *Id.*

the Board decided that the post was an expression of opinion and not hate speech.

When examining the post under Matsuda's conception of racist speech as a message (1) "of racial inferiority" (2) "directed against a historically oppressed group" (3) that is "persecutorial, hateful, and degrading,"[128] the Board's decision was rightfully made. While the comment was directed against a historically oppressed group (Muslims), it was neither a message of inferiority nor was it persecutorial, hateful, or degrading. Instead, when considered in full context, the post intended to draw attention to how Muslims reacted to different situations: one in which there was a cartoon and one in which there was a threat to Muslim lives. Therefore, by allowing this speech on Facebook, race-conscious cultural democracy is supported as others can engage with the poster's ideas and either support or reject them, thereby continuing the cycle of cultural production.

However, the Board's logic diverges from Matsuda's definition when it weighed human rights standards on freedom of expression, citing Article 20, para. 2 of the International Covenant on Civil and Political Rights (ICCPR) .[129] In particular, the Board decided that while the content was offensive, it did not rise to the level of "advocat[ing] hatred or intentionally incit[ing] any form of imminent harm."[130] As a result, the Board established a high bar for actionable racist speech that is actually above the one established in Facebook's own Community Standards and also above the one advocated in this Note.

The language "intentionally" is not located in Facebook's Community Standards and does not work to support a race-conscious cultural democracy online. The Community Standards explicitly state that "[i]f intention is unclear, we may remove content."[131] Therefore, intent is not a requirement for removal. In addition, as explored in Part III, intent is not a defining factor of racist speech under Mari Matsuda's definition, which only establishes a bar of

128. Matsuda, *supra* note 14, at 2357.

129. ICCPR, *supra* note 124, art. 20(2).

130. *Oversight Board overturns Facebook decision: Case 2020-002-FB-UA*, OVERSIGHT BD. (Jan. 2021), https://oversightboard.com/news/773985406543781-oversight-board-overturns-facebook-decision-case-2020-002-fb-ua/.

131. *Hate Speech,* FACEBOOK: COMMUNITY STANDARDS, https://www.facebook.com/communitystandards/hate_speech (last visited May 4, 2021). This section on intent is in reference to content that is perceived as racist when in fact the poster intends for the racist content to be used for educational purposes—a different situation entirely from the one in this case. In other words, the Community Standards require intent to be made clear when the intention is to *not* be racist, while the intent standard used by the Oversight Board is an intent to be racist.

"persecutorial, hateful, and degrading" speech.[132] Often, intent is not present at all in racist speech but that does not lessen its harm.[133]

Furthermore, an imminent harm requirement is also not located in Facebook's definition of hate speech, which hinges on the speech being a direct attack. An attack is "violent or dehumanizing speech, harmful stereotypes, statements of inferiority, expressions of contempt, disgust or dismissal, cursing, and calls for exclusion or segregation."[134] This is a far different standard than that used to evaluate the content in Case 2020-002-FB-UA, and it even directly conflicts with the rationale of Case 2020-003-FB-UA, which recognizes that Facebook is allowed to "prohibit some discriminatory expressions…absent any requirement that the expression incite violent or discriminatory acts."[135]

While citing to international law lends legitimacy to the Board's decision making, it also sets a dangerous precedent for future cases given that hate speech is not always driven by intent and does not always result in imminent harm. Nonetheless, harm is still associated with hate speech, and this harm hinders the cultural participation of those affected, thereby suppressing a robust cultural democracy online.

In Case 2020-003-FB-UA, the Board upheld Facebook's removal of a post that included a Russian slur regarding Azerbaijanis.[136] Similar to Case 2020-002-FB-UA, the Board based its decision on Facebook's Community Standard, Facebook's values, and relevant international law. However, unlike the previous case, the Board acknowledged Facebook is not under the same obligations as a government when it comes to controlling racist speech. Therefore, it focused its analysis on whether the limitation on freedom of expression was justified under ICCPR Article 19, para. 3, which includes a three-part test: legality, legitimacy, and necessity and proportionality.[137] This standard, as opposed by the one in the previous case, allows the Board more flexibility to promote a race-conscious cultural democracy online.

---

132. Matsuda, *supra* note 14, at 2357.

133. *See supra* note 43–46 and accompanying text.

134. *Hate Speech*, FACEBOOK: COMMUNITY STANDARDS, https://www.facebook.com/communitystandards/hate_speech (last visited May 4, 2021).

135. OVERSIGHT BD., CASE DECISION 2020-003-FB-UA (2021), https://oversightboard.com/decision/FB-QBJDASCV/ .

136. *Id.*

137. *Id.* ("Facebook's Hate Speech Community Standard prohibits some discriminatory expression, including slurs, absent any requirement that the expression incite violent or discriminatory acts. While such prohibitions would raise concerns if imposed by a Government at a broader level…,the Special Rapporteur indicates that entitles engaged in content moderation like Facebook can regulate such speech….")

The content under question was the term "taziks" to refer to Azerbaijanis. In Russian, the term refers to "wash bowls," but it is also a play on the word "aziks," a derogatory term for Azerbaijanis that was on Facebook's list of prohibited slurs. As a result, Facebook removed the post as a violation of its Community Standards because it was meant to dehumanize Azerbaijanis.

When examining the post in question, the Board agreed that the slur was a violation of Facebook's Community Standards on hate speech. It also believed that the post violated Facebook's values of "Dignity" and "Safety" to such an extent as to outweigh Facebook's countervailing commitment to "Voice."[138]

The main question that the Board confronted was whether Facebook's restriction met three requirements: legality, legitimacy, and necessity and proportionality.[139] To answer the question of legality, the Board asked if the restriction on speech was clear and accessible to user. Because Facebook has an extensive Community Standard section on hate speech that explicitly lists slurs as prohibited, this requirement was met. In order to meet the legitimacy requirement, the Board asked if the restriction on slurs served a legitimate aim. In this case, the restriction on slurs protects "people's right to equality and non-discrimination" under Article 2, para. 1 of the ICCPR.[140] Finally, in order to meet the necessity and proportionality requirement, the restriction must be in response to a threat to the rights of others and proportional. The Board acknowledged that while international law does not prohibit "insults, ridicule or slander," Facebook, being a company and not a government, is allowed to make such restrictions.[141]

In its decision, the Board highlighted the importance of context. The Board consulted cultural experts in the region as to the meaning and use of the slur, determining that it was hateful and dehumanizing. Furthermore, the post was shared in an area that is in the middle of an armed conflict. Therefore, even if offline harm resulting from online speech might not have been imminent or known, the Board concluded that the hate speech risked such a result. The Board also held that Facebook's response was proportional; mere warning labels would not have sufficed, and Facebook did not look to take

138. *Id.*; *Introduction*, FACEBOOK: COMMUNITY STANDARDS, https://www.facebook.com/communitystandards/introduction (last visited May 4, 2021).

139. *Supra* note 135.

140. ICCPR, *supra* note 121, art. 2(1).

141. OVERSIGHT BD., CASE DECISION 2020-003-FB-UA (2021), https://oversightboard.com/decision/FB-QBJDASCV/ .

more severe measures such as removing the offender from the platform entirely.[142]

Case 2020-03-FB-UA was also the only case of the first six to come with a policy recommendation. The Board suggested that the platform increase its transparency by providing users with reasons for why their content was restricted. In particular, the Board highlighted that "Facebook's lack of transparency left its decision susceptible to the mistaken belief that it had removed the post because the user was addressing a controversial subject or expressing a viewpoint Facebook disagreed with."[143] Additional transparency by Facebook can only help the platform as it would allow users to instill more trust in the platform, and in turn, allow the platform to instill more trust in its users, an important aspect of race-conscious cultural democracy.

This decision fits squarely in line with race-conscious cultural democracy. Under Matsuda's definition, the slur in this case was a message of inferiority directed against Azerbaijanis that was persecutorial, hateful, and degrading, making it an easy case of actionable hate speech. The Board relied on similar reasoning under both the legitimacy analysis and the necessity and proportionality analysis. The Board also emphasized the harmful effect that hate speech can have regardless of whether it is experienced offline in the form of violence—a key tenant of why hate speech must be controlled within a race-conscious cultural democracy. It also recognized that removal of hate speech is sometimes the best course of action, leaving no room for the argument that open debate is the best solution to hate speech. Importantly, the Board acknowledged that the offender was still able to engage in cultural production as they were "free to engage in discussions on the same issues within the boundaries of the Community Standards."[144] Therefore, not only is the offender not exiled from culture in the way that cancel culture rhetoric threatens, but those harmed by the slur are also able to engage in a space free from hate, thereby promoting race-conscious cultural democracy.

---

142. *Id.* Notably, however, the Board acknowledged that the offender in question had previously posted similar hate speech on the platform, raising the question of how many times someone must get reported for individual incidents before they are removed from the platform entirely. The potential for reoffending is a setback for community accountability as discussed in Part IV.D and is also a driving criticism that Facebook received following the 450 reports placed against the Kenosha Guard. Mac & Silverman, *supra* note 8. The Board's decision does not specify how many times this user was reported, nor does it suggest how many infractions would be necessary rise to the level of platform removal. This is a question that will inevitably need to be resolved by the Board in future cases.

143. *Id.*

144. Oversight Bd., Case Decision 2020-003-FB-UA (2021), https://oversightboard.com/decision/FB-QBJDASCV/ .

As a result, while Case 2020-002-FB-UA stands on shaky grounds in regards to both race-conscious cultural democracy theory and Facebook's own Community Standards, Case 2020-003-FB-UA establishes a rationale that is supported by both. As more cases on similar issues come before the Board, it will have to look at its case precedent and decide which approach to follow when analyzing hate speech. Only the latter approach, however, is tenable under race-conscious cultural democracy theory.

V. CONCLUSION

Finding a balance between too much and not enough control over speech has always been difficult. Racist speech works on a spectrum with direct call for violence on one end and covert, sanitized speech on the other. In between is a zone of ambiguity that is colored with context that is often difficult to disentangle from a perspective outside the community affected by the racist speech. Critical race theory can help explain where to draw the line between actionable and non-actionable racist speech. Further, cultural democracy theory can help elucidate how to put that line into practice by emphasizing cultural participation from all members of society. In particular, speech that is a message (1) "of racial inferiority" (2) "directed against a historically oppressed group" (3) that is "persecutorial, hateful, and degrading" is actionable racist speech that should be controlled.[145]

However, current cultural democracy theory neglects the harm that racist speech can have on the cultural participation of historically oppressed persons. Therefore, a race-conscious cultural democracy recognizes the need to control some speech to promote the cultural participation of all persons. To that end, the first line of defense against racist speech should be the community—the most democratic form of control. But with community accountability, there is often a fear of cancel culture and a worry that it does not do enough to remedy racism. Therefore, social media platforms like Facebook must step up to ensure that not only that their users have the proper tools to regulate themselves, but also that the correct speech is being controlled.

There is still much work to be done before a race-conscious cultural democracy can be fully realized, and it will inevitably take trial and error. Communities themselves are still learning the best way to use their collective power, as shown by the constant turmoil in YouTube's beauty guru community. Platforms as well are still learning how to craft ToS and Community Standards that capture the right kind of unwanted speech.

145. Matsuda, *supra* note 14, at 2357.

In the case of Facebook, their Community Standards should be amended to state, "we define hate speech as a direct attack on people *of a historically oppressed group* based on what we call protected characteristics." This will more explicitly distinguish the speech of the oppressor from the oppressed and therefore prevent the hijacking of the section by those who seek to weaponize it against the people it is meant to protect.

In addition, Facebook must trust its users and remove reported content—especially when numerous users report the same content. This is an issue that went unresolved following Kenosha and that has been acknowledged but unremedied by the Oversight Board.[146] Facebook can increase its trust in its users by enhancing the users' toolkit with more opportunities to report, providing them with more information on Facebook's definition of racist speech, and increasing the transparency when content is removed.

Facebook's Oversight Board is a promising mode of corporate oversight that works to buttress community control and promote race-conscious cultural democracy in the digital age. While the decision in Case 2020-002-FB-UA has potentially detrimental reasoning, it reaches the right conclusion and is balanced by the rationales underpinning Case 2020-003-FB-UA. Therefore, going forward, if the Board seeks to create a space on the platform that allows all persons to freely and widely engage in cultural production, it should follow the latter case's reasoning. However, as of yet, it is unclear what role precedent cases will hold in the Board's future decision making as it has only recently selected its next batch.[147]

Regardless, the installation of the Oversight Board should spur other platforms to consider how they too can better manage speech on their sites. They should note the benefits that such a Board can have when it comes to promoting the ever-growing cultural democracy online.

---

146. *See* Mac & Silverman, *supra* note 8; OVERSIGHT BD., CASE DECISION 2020-002-FB-FBR (2021), https://web.archive.org/web/20210128145029if_/https://oversightboard.com/decision/FB-I2T6526K/.

147. *Announcing the Oversight Board's next cases*, OVERSIGHT BD. (Jan. 2021), https://oversightboard.com/news/175638774325447-announcing-the-oversight-board-s-next-cases/.

FIXING TRADEMARK LAW WITH *ROMAG FASTENERS*: A NEW THEORY OF UNFAIR COMPETITION

Cole Gingrich†

I. INTRODUCTION

Why do luxury companies burn their unsold products at the end of the year?[1] The practice seems absurd. The products obviously have value; why not sell them at a discount and make some money? A surprising part of the answer is trademark law.

Trademark law protects consumers and businesses from harm resulting from consumer confusion.[2] Consumer confusion occurs when consumers look at a product bearing a mark, and the mark causes them to mistake or be unsure about certain aspects of the product's manufacture.[3] Consumer confusion can cause harm in multiple ways. For instance, a watch bearing the Rolex trademark carries certain associations for consumers. A consumer purchasing a Rolex watch in a store is likely willing to pay a premium based on the manufacturer's reputation for quality.[4] If the consumer purchases the watch and later finds out it was a knockoff, then her harm is the amount she overpaid.[5] Additionally, the Rolex watch manufacturer has suffered harm

---

DOI: https://doi.org/10.15779/Z38H708182

† J.D., 2022, University of California, Berkeley, School of Law.

1. *See, e.g.*, Matthew Dalton, *Why Luxury Brands Burn Their Own Goods*, WALL ST. J. (Sept. 6, 2018), https://www.wsj.com/articles/burning-luxury-goods-goes-out-of-style-at-burberry-1536238351 [https://perma.cc/8RPE-AQ4R].

2. *See* 1 J. THOMAS MCCARTHY, MCCARTHY ON TRADEMARKS AND UNFAIR COMPETITION § 2:1 (5th ed. 2020) (explaining the historical "dual goals of trademark law").

3. *See* Robert G. Bone, *Taking the Confusion Out of Likelihood of Confusion: Toward a More Sensible Approach to Trademark Infringement*, 106 NW. U. L. REV. 1307, 1338–39 (2012) (listing origin confusion, licensing confusion, post-sale confusion, and initial-interest confusion).

4. *See* Mark A. Lemley, *The Modern Lanham Act and the Death of Common Sense*, 108 YALE L.J. 1687, 1690 (1999) (explaining that the ability of trademarks is particularly important for characteristics "that consumers cannot readily verify except by buying the product").

5. *See* Mark P. McKenna, *The Normative Foundations of Trademark Law*, 82 NOTRE DAME L. REV. 1839, 1844 (2007) (describing the "narrow sense" of trademarks' benefits to consumers—"protecting them from being deceived into buying products they do not want").

because, absent the knockoff, it would have made a sale.[6] Its harm is the amount of profit lost.

Under the Lanham Act, courts have expanded trademark protection to prevent more abstract forms of harm, including to a trademark owner's reputation.[7] The Lanham Act is the statute that defines trademark infringement and specifies remedies.[8] Unfortunately, it does not define "confusion."[9] In the Rolex example, a more comprehensive definition of the confusion that occurred is confusion as to the origin of a good at the point-of-sale. The consumer assumed that the watch she was buying was manufactured by, or originated from, the well-known Rolex watch manufacturer, when in reality it was a knock-off. That is the very first type of confusion that trademark law ever protected against.[10] However, modern courts have expanded the definition of confusion. The Lanham Act specifies the types of actionable confusion as those to the "affiliation, connection or association of [one] person with another person, or as to the origin, sponsorship, or approval of his or her goods, services, or commercial activities . . . ."[11] Also, courts have found that actionable confusion can occur even if the consumer is initially confused but the confusion is cleared up before purchase,[12] or if a third party is confused after observing a product post-sale.[13] Often, courts will characterize these less-direct forms of harm as harm to a trademark owner's reputation, or "goodwill."[14] By strongly protecting reputational value, trademark law incentivizes trademark owners to engage in activities that increase that value, such as creating artificial scarcity by burning unsold designer clothes.[15]

---

6. *See* Mark A. Thurmon, *Confusion Codified: Why Trademark Remedies Make No Sense*, 17 J. INTELL. PROP. L. 245, 260 (2010) ("[E]ach sale made by the defendant apparently was a sale the plaintiff would have made but for the infringement . . . .").

7. *See id.* (explaining that the poor quality of the defendant's infringing product could also harm the plaintiff's reputation, "likely [leading] to additional lost sales").

8. Lanham Act § 43(a), 15 U.S.C. § 1125(a) (infringement); Lanham Act § 35(a), 15 U.S.C. § 1117(a) (monetary remedies).

9. Bone, *supra* note 3, at 1338.

10. *See* McKenna, *supra* note 5, at 1850–51 (describing the "clothier's case," often cited as the first English trademark case).

11. Lanham Act § 43(a), 15 U.S.C. § 1125(a).

12. *See* Bone, *supra* note 3, at 1339 (describing initial-interest confusion).

13. *Id.*

14. *See, e.g.*, Thurmon, *supra* note 6, at 267 ("part of the injury is the damage done to the plaintiff's reputation").

15. *See* Deven R. Desai, *The Chicago School Trap in Trademark: The Co-Evolution of Corporate, Antitrust, and Trademark Law*, 37 CARDOZO L. REV. 551, 603–04 (2015) (describing how current trademark protection facilitates branding strategies that permit trademark owners "to turn a mark into a thing with value that is desired on its own").

Some courts, driven by concern that trademark protection was too strong, developed a rule that limited one available remedy, leading to a split in the U.S. Courts of Appeals. The remedy at issue was an award of the infringer's profits: some circuits held that a profits remedy would only be available in cases of willful infringement.[16] Because the profits remedy often overcompensated the plaintiff for actual harm, courts imposing this rule reasoned that equity only permitted it when the infringer acted with a more culpable mental state.[17] Other circuits, noting that the Lanham Act did not impose an explicit mental state requirement, held that the defendant's mental state was only one factor to consider when determining an appropriate remedy.[18]

The U.S. Supreme Court in *Romag Fasteners v. Fossil* held that willful infringement was not a prerequisite to the profits remedy.[19] The majority opinion stated that while it "might be possible" to read the statute as requiring willfulness if courts of equity had historically imposed such a requirement, courts of equity had not consistently done so.[20] Therefore, a willfulness prerequisite failed to meet the majority's definition of a "principle of equity," as the Lanham Act uses the phrase. The majority explained that principles of equity are usually "fundamental rules that apply more systematically across claims and practice areas."[21]

Courts should use *Romag* and the remedial discretion provided by the Lanham Act to rein in trademark protection by utilizing a new theory of unfair competition law. Scholars have explained that trademark law and unfair competition law are related because "[t]rademark law is part of the branch of tort law known as unfair competition."[22] Unfortunately, existing theories as to what constitutes "unfair," as opposed to legitimate, competition all inevitably lead to unlimited protection, increasingly characterizing once-legitimate competition as unfair. This Note proposes a new theory—the Consumer

16. *See, e.g.*, George Basch Co. v. Blue Coral, 968 F.2d 1532, 1540 (2d Cir. 1992), *superseded in relevant part by statute*, Lanham Act § 35(a), 15 U.S.C. § 1117(a) ("[A] plaintiff must prove that an infringer acted with willful deception before the infringer's profits are recoverable by way of an accounting."). An accounting is another name for the profits remedy.

17. *See id.* ("[A]n accounting may overcompensate for a plaintiff's actual injury and create a windfall judgment at the defendant's expense.").

18. *See*, *e.g.*, Roulo v. Russ Berrie & Co., 886 F.2d 931, 941 (7th Cir. 1989) ("Other than general equitable considerations, there is no express requirement . . . that the infringer willfully infringe . . . to justify an award of profits.").

19. Romag Fasteners, Inc. v. Fossil, Inc., 140 S. Ct. 1492, 1497 (2020).

20. *Id.* at 1496–97.

21. *See id.* at 1496.

22. William M. Landes & Richard A. Posner, *The Economics of Trademark Law*, 78 TRADEMARK REP. 267, 267 (1988).

Approach—to separate unfair from legitimate competition and explains how courts can implement it under the current Lanham Act.

The Consumer Approach defines an act of unfair competition as a business practice that is anticompetitive, wherein anticompetitive means that the business practice harms consumers more than it benefits them. The relevant consumer benefits are lower prices or higher quality goods. By conducting an economic analysis, courts can use consumer benefit to limit the situations in which harm, either business or consumer, is actionable. In unfair competition cases where harm seems very abstract, the consumer benefits arising out of the defendant's activities might well outweigh the harm.

In trademark cases, courts can apply the Consumer Approach at the remedy stage. Although the Lanham Act currently requires courts to use the likelihood-of-confusion test to determine infringement,[23] courts have significant discretion to adjust remedies according to the "principles of equity,"[24] as defined in *Romag*. One such principle that meets the *Romag* majority's definition is promoting competition.[25] Based on this principle, courts following the Consumer Approach should simply deny the plaintiff a remedy if the defendant's activities are procompetitive, meaning the consumer benefits they produce outweigh the consumer harm.

To manage the administrative costs of determining whether a defendant's activities are procompetitive, courts should utilize the burden shifting already present in the Lanham Act. For the damages remedy, plaintiffs currently bear the burden of proving actual harm.[26] For the profits remedy, the plaintiff must only prove the defendant's sales; the defendant then has the burden to prove deductions not due to the infringement.[27] Under the Consumer Approach, courts would determine the appropriate remedy based on the defendant's

---

23. *See* Lanham Act § 43(a), 15 U.S.C. § 1125(a) (imposing liability for acts that are "likely to cause confusion").

24. *See id.* § 35(a) (permitting courts to adjust damages and profits awards "according to the circumstances of the case").

25. *See, e.g.*, 1 LOUIS ALTMAN & MALLA POLLACK, CALLMAN ON UNFAIR COMPETITION, TRADEMARKS AND MONOPOLIES § 4:1 (4th ed. 2020) ("It is the primary function of the law of unfair competition to safeguard the competitive community . . . . The antitrust laws are equally designed to maintain the competitive order . . . ."); Eastman Co. v. Reichenbach, 20 N.Y.S. 110, 116 (N.Y. Sup. Ct. 1892), *aff'd sub nom.* Eastman Kodak Co. v. Reighenbach, 29 N.Y.S. 1143 (N.Y. Gen. Term 1894) ("[I]t is always the policy of the law to foster and encourage [legitimate competition.]").

26. *See* Mark A. Thurmon, *Federal Trademark Remedies: A Proposal for Reform*, 5 AKRON INTELL. PROP. J. 137, 149 (2011) ("[T]o recover actual damages a prevailing trademark owner must be able to accurately quantity its monetary injury and show a causal connection between that injury and the infringement.").

27. Lanham Act § 35(a), 15 U.S.C. § 1117(a).

mental state. If the defendant's infringement is unintentional, then the court could award damages. To realize the award, the plaintiff would have the burden to prove that consumer harm outweighs benefit. If the defendant's infringement is at least negligent, then the court could award profits. To prevent the award, the defendant would have the burden to prove that consumer benefit outweighs harm. Because quantifying harm and benefit is difficult, shifting the burden to the defendant is likely to increase the ultimate monetary award.[28] If the infringement is unintentional, then the increase is not appropriate because the only purpose of the monetary award is to provide compensation to the plaintiff.[29] However, if the infringement is at least negligent, then the increase is appropriate because the monetary award serves the additional purpose of deterring future infringers.[30]

This Note has four additional Parts. Part II describes how the Lanham Act came to provide its current, expansive trademark protection and how the circuit split over the profits remedy developed. Part III explains *Romag Fasteners v. Fossil*. Part IV proposes the Consumer Approach to identifying unfair competition and applies it to trademark law. Part V concludes.

II. THE LANHAM ACT'S PROPERTY-LIKE PROTECTION

Although modern law does not consider trademarks to be a form of property in themselves akin to patents and copyrights,[31] judicial expansion of trademark infringement under the Lanham Act has created a nearly-equivalent level of protection.[32] Nominally, trademark protection serves a dual role of protecting consumers and the trademark owner's underlying business from harm.[33] However, courts have interpreted the infringement provision of the Lanham Act broadly, protecting trademark owners from very abstract forms of business harm.[34] Eventually, some courts grew concerned that the high level

28. *See* Thurmon, *supra* note 6, at 313 ("[A]ctual damages are hard to quantify in most modern trademark cases . . . .").

29. *See* Oren Bracha & Talha Syed, *The Wrongs of Copyright's Statutory Damages*, 98 TEX. L. REV. 1219, 1236–37 (2020) (explaining the absence of a deterrence rationale in innocent infringement cases).

30. *See id.* (explaining the presence of a deterrence rationale "[o]nce the negligence line is crossed").

31. *See* Lemley, *supra* note 4, at 1337 ("The justifications for trademark law are different from those for other forms of intellectual property.").

32. *See* Bone, *supra* note 3, at 1337 ("Over the past thirty years or so, courts have given the likelihood of confusion test very broad interpretations that risk high social cost without clear trademark-related benefits.").

33. *See* MCCARTHY, *supra* note 2, at § 2:1.

34. *See* Bone, *supra* note 3, at 1338 ("The likelihood of confusion test merely instructs courts to determine whether 'confusion' is likely [without defining 'confusion']. Many judges

of protection may be overburdening some defendants and imposed a rule limiting an award of the infringer's profits.[35] Disagreement over the rule led to a circuit split that the Supreme Court eventually resolved in *Romag Fasteners v. Fossil*.[36]

This Part explains the three candidate interests that courts have historically proposed as the interests that trademark law protects, the potential harms to those interests, and how one—the property interest—was ultimately rejected by the Supreme Court. Then, it explains how courts, despite the Supreme Court's rejection, have interpreted the Lanham Act's infringement and remedy provisions to essentially protect the property interest in some cases by preventing all consumer confusion. Finally, it describes the evolution of a circuit split over the profits remedy.

A. THREE CANDIDATE PROTECTED INTERESTS

Courts have developed three candidates for the interest that trademark law protects: the consumer interest, the business interest, and the property interest. Consumer confusion potentially harms all three. However, a division exists between the consumer and business interests on one side and the property interest on the other. Harm from consumer confusion to the consumer and business interest is monetary and theoretically measurable. Harm to the property interest is not measurable. Although some courts throughout history have intermittently protected the property interest,[37] modern trademark law nominally protects only the consumer and business interests.

The consumer interest is customers' confidence in the accuracy of their past experiences with a trademark.[38] Trademarks label products, either goods or services. Consumers form associations with trademarks based on their past experiences with those products.[39] Trademarks therefore provide consumers with information.[40] By protecting the consumer interest, trademark law

---

have followed this instruction quite literally by broadly giving credit to any possible way that consumers might be confused.").

35. *See, e.g.*, George Basch Co. v. Blue Coral, 968 F.2d 1532, 1540 (2d Cir. 1992), *superseded in relevant part by statute*, Lanham Act § 35(a), 15 U.S.C. § 1117(a) ("[W]e believe that this [willful infringement] requirement is necessary to avoid the conceivably draconian impact that a profits remedy might have in some cases.").

36. *See* Romag Fasteners, Inc. v. Fossil, Inc., 140 S. Ct. 1492, 1494 (2020) (noting that "some circuits" had adopted a willful infringement prerequisite to an award of profits and holding that no such prerequisite existed).

37. *See* McKenna, *supra* note 5, at 1882 ("[C]ourts sometimes said that exclusive use of a mark made it the producer's property.").

38. Desai, *supra* note 15, at 602.

39. *Id.*

40. *Id.*

provides consumers with confidence that they are associating those past experiences accurately.

Infringers can harm the consumer interest in two ways. The first method is by causing a consumer to overpay.[41] If a consumer's past experiences have led her to form a positive association with a trademark, then she will perceive a future purchase of a product bearing the trademark as less risky.[42] Her possibility of dissatisfaction is lower because of her positive past experiences. Therefore, she is willing to pay a higher price for a product with the trademark than she would be for a product with which she has no prior experiences. An infringer, by appropriating the trademark, can also appropriate the consumer's willingness to pay a higher price. The measure of this harm is the difference between what a consumer pays for the infringer's product with the trademark and what she would have paid for the infringer's product without the trademark.

The second method of harm to the consumer interest is an increase in search costs.[43] Trademarks provide consumers with a shortcut. Without them, consumers would spend more time researching and inspecting a product before purchase.[44] This harm is more difficult to quantify for two reasons. First, some trademarks are more effective than others, and an effective trademark causes a greater reduction in search costs.[45] Second, time value differs among consumers. Increased search costs cause less harm to a consumer with a plethora of free time.

The business interest is the trademark owner's future profits due to consumers' associations with the trademark. Once consumers form positive associations with a trademark, the trademark owner can charge a higher price because the purchase carries less risk of disappointment for the consumer and because the consumer's search costs have decreased.[46] Consumer associations comprise the trademark owner's reputation, or "goodwill."[47] By protecting the

41. *See* McKenna, *supra* note 5, at 1844.

42. *See id.*

43. *See* Landes & Posner, *supra* note 22, at 277 (asserting that a consumer may be willing to pay more for the assurance that a product is of sufficient quality).

44. *See id.* at 271 ("Rather than investigating the attributes of all goods in order to determine which one [to purchase], the consumer may find it less costly to search by identifying the relevant trademark . . . .").

45. *Id.* at 278.

46. *Id.*

47. *See* Robert G. Bone, *Hunting Goodwill: A History of the Concept of Goodwill in Trademark Law*, 86 B.U. L. REV. 547, 549 (2006) ("Goodwill . . . denotes the special value that attaches to a mark when the sellers' advertising and investments in quality generate consumer loyalty . . . .").

business interest, trademark law ensures that trademark owners reap the rewards of building a favorable reputation.[48]

Infringers can harm the business interest by causing the trademark owner to lose future profits. Infringers can cause this harm directly through trade diversion. Trade diversion can occur if the trademark owner and the infringer operate in the same market, meaning they sell substitute products in the same geographical area.[49] Substitute products are those that can easily replace each other such that a consumer would likely buy one or the other.[50] Trade diversion happens when consumers intend to buy the trademark owner's product but mistakenly purchase an infringer's product instead.[51] The measure of the harm is the amount of profit the trademark owner would have made on the sale. An underlying assumption is that the trademark owner had sufficient inventory to make the sale.

Infringers can also harm the business interest indirectly by tarnishing the trademark owner's reputation.[52] This harm can occur even if the trademark owner does not have sufficient inventory to make the infringer's sale and even if the trademark owner and the infringer do not operate in the same market. An infringer tarnishes a trademark owner's reputation by selling inferior products bearing the trademark. Consumers' experiences with the inferior product negatively impact their associations with the trademark, leading them to perceive a greater risk of dissatisfaction when purchasing future products bearing the trademark.[53] Because consumers perceive a greater risk, the trademark owner will either sell fewer products or have to lower the product price, decreasing future profits.

The property interest is the trademark itself as the owner's exclusive property,[54] and it shares some similarities with the business interest. For both, positive consumer associations create a reputation for the trademark that allows the trademark owner to increase profit.[55] Also for both, protecting

48. *See id.* ("[T]rademark law prevents others from appropriating [goodwill] by using a similar mark.").

49. *See* Thurmon, *supra* note 6, at 260.

50. *See* Mark P. McKenna, *Testing Modern Trademark Law's Theory of Harm*, 95 IOWA L. REV. 63, 99–100 (2009) ("Substitute products tend to have common applications and use contexts such that one product could replace the other in usage and satisfy the same needs.").

51. *See* Thurmon, *supra* note 6, at 260.

52. *See id.* (explaining that the poor quality of the defendant's infringing product could also harm the plaintiff's reputation, "likely [leading] to additional lost sales").

53. *See* Landes & Posner, *supra* note 22, at 277 (asserting that a consumer may be willing to pay more for the assurance that a product is of sufficient quality).

54. *See* McKenna, *supra* note 5, at 1882 ("A . . . possible reading of the cases . . . is that courts sought to protect a property right in the trademark itself.").

55. *See* Landes & Posner, *supra* note 22, at 277.

trademarks allows trademark owners to reap the rewards due to the favorable reputation.[56]

However, the property interest differs from the business interest in terms of the harm that infringers inflict. Infringers harm the property interest through unauthorized use.[57] Similar to a trespass upon real property, this harm occurs regardless of any actual damage the infringer causes.[58] The harm exists simply because the trademark owner has lost exclusive control of the trademark's reputation.

Some courts throughout American history have intermittently protected the property interest, but the Supreme Court rejected it in favor of the business interest. In *Oakes v. Tonsmierre*, the now-abolished U.S. Circuit Court for the Southern District of Alabama found infringement "on account of the unauthorized . . . use" by the defendant because "trade-marks are protected . . . on the ground of property."[59] However, the Supreme Court in *Hanover Star Milling Co. v. Metcalf* situated trademark law in unfair competition, not property, law.[60] The Court further specified the extent of a trademark owner's property right, stating that the trademark owner has no property right "except as appurtenant to an established business or trade in connection with which the mark is used."[61] The Seventh Circuit's earlier opinion in the same case was even clearer: "It is not the trade-mark, but the trade, the business reputation and good will, that is injured . . . ."[62] These opinions describe trademark law as protecting the business interest, even though couched in property terms.

Modern trademark law protects both the consumer and business interests. Because consumer confusion harms both interests, the modern trademark

---

56. *See* Bone, *supra* note 47, at 549 ("[T]rademark law prevents others from appropriating [goodwill] by using a similar mark.").

57. *See* James M. Koelemay Jr., *Monetary Relief for Trademark Infringement Under the Lanham Act*, 72 TRADEMARK REP. 458, 469–70 (1982) (describing an early case in which the court "spoke of the plaintiff having 'title to the marks in question' and a right of 'exclusive use'").

58. *See id.* at 476 ("[A] definite trend developed toward awarding damages and an accounting as a matter of right in cases of technical trademarks . . . because of the defendant's 'trespass' on the plaintiff's absolute right of property.").

59. Oakes v. Tonsmierre, 49 F. 447, 453 (C.C.S.D. Ala. 1883).

60. Hanover Star Milling Co. v. Metcalf, 240 U.S. 403, 413 (1916), *superseded by statute*, Lanham Act § 22, 15 U.S.C. § 1072, *as recognized in* Park 'N Fly, Inc. v. Dollar Park and Fly, Inc., 469 U.S. 189 (1985).

61. *Id.* at 414.

62. Hanover Star Milling Co. v. Allen & Wheeler Co., 208 F. 513, 516 (7th Cir. 1913), *aff'd*, 240 U.S. 403 (1916), *superseded by statute*, Lanham Act § 22, 15 U.S.C. § 1072, *as recognized in* Park 'N Fly, Inc. v. Dollar Park and Fly, Inc., 469 U.S. 189 (1985).

statute—the Lanham Act[63]—protects both by simply preventing consumer confusion.[64] However, as the next Section explains, judicial expansion of the types of actionable "confusion" has caused trademark law to effectively re-protect the property interest under some circumstances.

B. THE LANHAM ACT PROVIDES EXPANSIVE PROTECTION

Congress drafted the infringement and monetary remedy provision of the Lanham Act broadly, and courts have interpreted the provisions to provide extensive protection.[65] The likelihood-of-confusion test for trademark infringement now protects trademark owners against abstract forms of reputational injury.[66] Increased protection at the infringement stage of a case also leads to increased protection at the remedy stage by affecting the remedies available. The damages remedy places the burden of proving harm on the plaintiff,[67] but as the injuries recognized become more abstract, they also become harder for trademark owners to prove.[68] This leads courts to place more emphasis on the profits remedy, which shifts the burden of proof onto the defendant.[69] However, the profits remedy tends to overcompensate the plaintiff in relation to actual harm suffered.[70] Thus, current trademark protection has expanded to both recognize additional types of injury and overcompensate for the injuries recognized. This Section describes the infringement and remedy provisions of the Lanham Act and details how courts have interpreted them to expand trademark protection.

1. *Trademark Infringement*

> Any person who, on or in connection with any goods or services . . . uses in commerce any word, term, name, symbol, or device . . . which is **likely to cause confusion**, or to cause mistake, or to deceive as to the **affiliation, connection, or association** of

63. Pub. L. No. 79-489, 60 Stat. 427 (codified as amended in scattered sections of 15 U.S.C.).

64. *See* Lanham Act § 43(a), 15 U.S.C. § 1125(a) (imposing liability for defendant trademark uses that are "likely to cause confusion").

65. *See* Bone, *supra* note 3, at 1338.

66. *See, e.g.*, *id.* at 1339 ("The court worried that third parties viewing purchased jeans worn in public might [mistakenly] infer from the stitching pattern that they were Levi's jeans.").

67. Thurmon, *supra* note 26, at 149.

68. *See*, *e.g.*, *id.* at 138 n.2 ("It may be difficult, even impossible, to quantify the monetary damage resulting from a particular instance of trademark infringement, but if the goodwill is harmed, then it follows that economic harm has occurred.").

69. *See id.* at 169 ("Under the current rules, when profits are awarded, defendants bear the burden of proving up any deductions from gross revenue . . . .").

70. *Id.*

> such person with another person, or as to the **origin, sponsorship, or approval** of his or her goods, services, or commercial activities by another person . . . shall be liable in a civil action . . . .[71]

Courts determine infringement based on a likelihood-of-confusion test that equates confusion with harm.[72] The test is multi-factor. Although the specific factors differ between the circuits, there is some overlap.[73] For example, the Ninth Circuit uses eight non-exhaustive factors in cases involving related goods:[74]

1. strength of the mark;
2. proximity of the goods;
3. similarity of the marks;
4. evidence of actual confusion;
5. marketing channels used;
6. type of goods and the degree of care likely to be exercised by the purchaser;
7. defendant's intent in selecting the mark; and
8. likelihood of expansion of the product lines.[75]

Actual harm is not a listed factor. The closest factor to actual harm is "evidence of actual confusion." However, although actual harm is logically impossible without actual confusion, "[e]very circuit . . . has made clear that evidence of actual confusion, though probative, is not required [to find infringement]."[76] Because proving actual harm and actual confusion is very difficult, it makes sense that the lack of such proof should not preclude finding infringement where the other factors make confusion likely. Especially important is the defendant's intent—"[t]he facile assumption, evidently quite pervasive among the courts, [is] that if the defendant intended to confuse, then it succeeded in doing so . . . ."[77]

---

71. Lanham Act § 43(a)(1)–(A), 15 U.S.C. § 1125(a)(1)–(A) (emphasis added).

72. McKenna, *supra* note 50, at 67.

73. *See* Barton Beebe, *An Empirical Study of the Multifactor Tests for Trademark Infringement*, 94 CALIF. L. REV. 1581, 1582–83, 1589–90 (2006) (describing the disparities between the Circuits and "five core factors").

74. "Related goods are those 'products which would be reasonably thought by the buying public to come from the same source if sold under the same mark.'" AMF Inc. v. Sleekcraft Boats, 599 F.2d 341, 348 n.10 (9th Cir. 1979) (quoting Standard Brands, Inc. v. Smidler, 151 F.2d 34, 37 (2d Cir. 1945)). Related goods contrast with competitive goods, those that "compete for sales with [the goods] of the trademark owner." *Id.* at 348.

75. *Id.* at 348–49.

76. McKenna, *supra* note 50, at 67 n.5.

77. *See* Beebe, *supra* note 73, at 1631.

The Lanham Act does not define "confusion," and courts have expanded the term to further divorce infringement from actual harm.[78] Traditionally, actionable confusion meant a customer confused at the point-of-sale.[79] Today, however, consumer confusion both before and after the sale is also actionable.[80] For instance, the Second Circuit found infringement of the distinctive back-pocket stitching on Levi's jeans, not because a customer might be confused at the point-of-sale, but under the theory that a third party might observe someone wearing the defendant's jeans with similar stitching and mistake the defendant's jeans for Levi's.[81] The court reasoned that even though the similarity was unintentional, the defendant may have benefitted from "bootstrapping" to Levi's goodwill in making the sale (if the customer purchased the defendant's jeans hoping others would mistake them for Levi's), or that the third party's future buying decisions might somehow be affected by the mistaken observation.[82] In either case, harm to the business or consumer interest is possible. The customer's hope that others would mistake the defendant's jeans for Levi's could lead her to pay more than she otherwise would, harming the consumer interest. Harm to the business interest would result if, absent the defendant's similar stitching, the customer would instead have bought Levi's. However, those harms are much more abstract than the harms resulting from the customer herself being confused at the point-of-sale.

The expansion of actionable confusion has caused a similar expansion of "intent."[83] Logically, expanding confusion to encompass more abstract forms of harm means that intent should also broaden to encompass those abstract forms of harm. For instance, if the confusion caused by a customer buying non-Levi's jeans hoping that third party observers will confuse them for Levi's is actionable, then it makes sense for courts to hold that a defendant that intends for such "bootstrapping" to occur has bad faith intent. Some courts

78. *See* Bone, *supra* note 3, at 1338.

79. *See id.* at 1315 ("'Source confusion' lies at the heart of trademark law. It exists when the defendant's use of a similar mark leads consumers to believe that the [defendant] is actually selling the [plaintiff's] products.").

80. *Id.* at 1339.

81. *Id.* (citing Lois Sportswear, U.S.A., Inc. v. Levi Strauss & Co., 799 F.2d 867 (2d Cir. 1986)).

82. *See Lois Sportswear*, 799 F.2d at 872–73, 875 (explaining the possible types of post-sale confusion present in the case even though the evidence indicated that the defendants "happened on the stitching pattern serendipitously").

83. *See* Bone, *supra* note 3, at 1337 ("In the nineteenth century, the relevant intent was limited to subjective intent to deceive, but over time it has been construed much more broadly.").

have gone even further and said that simple knowledge of the plaintiff's mark coupled with adoption of a similar one is enough to infer intent.[84]

The related goods doctrine further enlarges trademark protection. Related goods are "products which would be reasonably thought by the buying public to come from the same source if sold under the same mark."[85] Historically, only a direct competitor could infringe a trademark.[86] This limitation made sense before nationwide marketing was possible, as firms usually stayed within a single product or service category.[87] However, the limitation stopped making sense in the early twentieth century when firms began to use one trademark across many categories.[88] For instance, in the case of *Aunt Jemima Mills Co. v. Rigney & Co.*, the plaintiff used the Aunt Jemima's mark for flour.[89] The defendant, with knowledge of the plaintiff's prior use, used the Aunt Jemima's mark for maple pancake syrup.[90] The Eastern District of New York upheld the direct competitor limitation, stating "[n]o one desiring to purchase flour would purchase syrup without knowing the difference. That is the test."[91] The Second Circuit rejected the direct competitor limitation, making it one of the earliest courts to do so,[92] reasoning that the goods were related because they were both food products commonly used together.[93] Such a relationship would lead "the public, or a large part of it, seeing this trade-mark on a syrup, [to] conclude that it was made by the complainant."[94] In related goods cases, consumer confusion resulting from the defendant's use of a similar mark can harm both the consumer and business interests, but the harm is less direct than the trade diversion that occurs when the parties are direct competitors.

The Lanham Act's expansive protection essentially protects the property interest for well-known trademarks. One study found that if a court finds both similarity between the marks and bad faith intent by the defendant, then the

84. *Id.*

85. AMF Inc. v. Sleekcraft Boats, 599 F.2d 341, 348 n.10 (9th Cir. 1979).

86. *See* McKenna, *supra* note 50, at 74 (explaining that courts only found a defendant's use infringing if it led to trade diversion and that trade diversion was impossible "absent direct competition between the parties").

87. *Id.*

88. *See id.* ("[M]ark owners wanted broader trademark rights that would allow them to prevent certain noncompetitive uses.").

89. Aunt Jemima Mills Co. v. Rigney & Co., 247 F. 407, 408 (2d Cir. 1917).

90. *Id.*

91. Aunt Jemima Mills Co. v. Rigney & Co., 234 F. 804, 808 (E.D.N.Y. 1916), *rev'd*, 247 F. 407 (2d Cir. 1917).

92. Bone, *supra* note 3, at 1316 n.51.

93. *Aunt Jemima Mills*, 247 F. at 409–10.

94. *Id.* at 410.

court finds a likelihood of confusion, regardless of any other factors.[95] Therefore, because a defendant will presumably be aware of well-known trademarks, the defendant's simple adoption of a mark that a later court concludes is similar will likely lead to the court inferring a bad faith intent.[96] Then, the similarity of the marks and the bad faith intent will likely lead the court to find a likelihood of confusion, and therefore infringement.[97] Recall that if trademark law protects the property interest, then it protects against a defendant's unauthorized use, regardless of any actual harm.[98] For well-known trademarks, this seems to be an apt description of current trademark protection under the Lanham Act.

2. *Trademark Monetary Remedies*

> When a violation . . . under section 1125(a) or (d) of this title, or a willful violation under section 1125(c) of this title, shall have been established . . . the plaintiff shall be entitled . . . **subject to the principles of equity**, to recover (1) defendant's profits, (2) any damages sustained by the plaintiff, and (3) the costs of the action.[99]

After finding infringement, courts can award three monetary remedies to the plaintiff "subject to the principles of equity."[100] Similar to § 1125(a), the infringement provision, Congress drafted § 1117(a), the remedies provision, expansively. The first sentence of the provision, provided above, seems to direct courts to award all three remedies for every violation.[101] However, the limitation, "subject to the principles of equity," means that the provision only makes the three remedies technically available and leaves to courts the task of determining whether any or all the remedies are appropriate based on the circumstances of the case.[102] The rest of § 1117(a) clarifies the vastness of the court's discretion:

---

95. *See* Beebe, *supra* note 73, at 1607 ("[T]wo findings—that the similarity of the marks factor favors a likelihood of confusion and the defendant's intent factor also favors a likelihood of confusion—are together sufficient to trigger an overall finding of a likelihood of confusion . . . regardless of the outcomes of any other factors.").

96. *See* Bone, *supra* note 3, at 1337 ("Intent to deceive can now be inferred merely from the defendant's adoption of a similar mark with knowledge of the plaintiff's prior use.").

97. *See* Beebe, *supra* note 73, at 1607.

98. *See* Koelemay, *supra* note 57, at 476 ("[A] definite trend developed toward awarding damages and an accounting as a matter of right in cases of technical trademarks . . . because of the defendant's 'trespass' on the plaintiff's absolute right of property.").

99. Lanham Act § 35(a), 15 U.S.C. § 1117(a) (emphasis added).

100. *Id.*

101. *See id.* ("[T]he plaintiff *shall be entitled* . . . to recover [profits, damages, and costs]." (emphasis added)).

102. *See* Koelemay, *supra* note 57, at 486 ("It may be hypothesized, therefore, that the purpose of the [principles of equity] phrase was to affirm the principle, established in prior

> In assessing **profits** the plaintiff shall be required to prove the defendant's sales only; defendant must prove all elements of cost or deduction claimed. In assessing **damages** the court may enter judgment, according to the circumstances of the case, for any sum above the amount found as actual damages, not exceeding three times such amount. If the court shall find that the amount of recovery based on **profits** is either inadequate or excessive the court may in its discretion enter judgment for such sum as the court shall find just, according to the circumstances of the case. Such sum . . . shall constitute compensation and not a penalty. The court in exceptional cases may away **reasonable attorney fees** to the prevailing party.[103]

The provision requires a clear formula to determine the profits award but then allows a court to increase or decrease the award "in its discretion."[104] Additionally, it permits a court to treble the actual damages assessed.[105] In neither case does the provision detail *how* a court should exercise its discretion, except vaguely as "subject to the principles of equity" and to compensate rather than penalize.[106]

The sweeping discretion of the remedies provision reflects the reality that harm from trademark infringement is difficult to prove. Because any monetary remedy is meant to be "compensation and not a penalty,"[107] courts understandably focus on harm to the plaintiff's business interest, meaning the ways in which the infringement has caused the plaintiff to directly or indirectly lose sales.[108] Even in the most direct case, where the parties are direct competitors and customers are confused at the point-of-sale, any remedy involves assumptions and approximations.[109] Business sales can increase or decrease for a multitude of reasons, so separating any impact specifically due to the infringement becomes more challenging as the harm becomes more

---

case law, that monetary relief was not automatic but was governed by equitable principles; this hypothesis is consistent with the natural import of the amending language.").

103. Lanham Act § 35(a), 15 U.S.C. § 1117(a) (emphasis added).

104. *Id.*

105. *Id.*

106. *Id.*

107. *Id.*

108. *See* Koelemay, *supra* note 57, at 487 ("Throughout the 1800s and early 1900s the purpose of monetary relief in trademark cases, both at law and in equity, was to compensate the trademark owner for loss of sales and good will . . . .").

109. *See* Thurmon, *supra* note 6, at 314 (describing the correlation between an infringer's profits and a plaintiff's lost profits in historical cases as merely "reasonably close").

indirect.[110] Remedial discretion allows courts to address these difficulties by tailoring monetary awards to the individual facts of each case.[111]

The Lanham Act also uses burden shifting to mitigate the difficulties of proving harm. For a damages remedy, the plaintiff bears the burden of proving harm, such as by showing a decrease in sales during the period of infringement.[112] However, even if the plaintiff can show a decrease, the plaintiff must also prove that the decrease resulted from the infringement, as opposed to any other reason.[113] For a profits remedy, the roles are reversed: "The plaintiff shall be required to prove the defendant's sales only; defendant must prove all elements of cost or deduction claimed."[114] Said another way, the defendant must prove that its sales did *not* result from the infringement.

As courts have expanded infringement to protect against more abstract forms of harm, the profits remedy has become increasingly important in compensating plaintiffs.[115] The simple reason is that as the harms become more abstract, they become more difficult to prove.[116] Returning to the earlier example regarding Levi's jeans, to receive damages, Levi's would have to prove that consumers purchased the infringer's jeans because of the similar back pocket stitching and further that, had the infringer used a different back pocket stitching, the consumers would have purchased Levi's jeans instead. If Levi's cannot meet that burden (which is likely), the damages remedy is unavailable. Therefore, the profits remedy effectively becomes the only option for courts to compensate for abstract harm like Levi's.[117] However, the profits remedy often overcompensates for harms because the defendant now bears the burden.[118] In the Levi's case, it becomes the defendant's task to prove that its sales were not due to consumers purchasing its jeans as imitation Levi's due to

110. *See id.* ("[A] reasonably close correlation between the defendant's profits as a result of the infringement and the plaintiff's lost profits . . . will seldom exist today, given the many complexities of modern markets.").

111. *See* Koelemay, *supra* note 57, at 480 (noting that the legislative history of the 1905 trademark statute indicated a congressional intent to provide courts with discretion due to "[t]he difficulty of providing exact damages in cases of this character").

112. *See* Thurmon, *supra* note 26, at 149.

113. *Id.*

114. Lanham Act § 35(a), 15 U.S.C. § 1117(a).

115. *See* Thurmon, *supra* note 6, at 313 ("The defendant's profits remedy has become the primary monetary remedy in Lanham Act cases.").

116. *See id.* ("[A]ctual damages are hard to quantify in most modern trademark cases . . . .").

117. *See id.* ("An accounting of defendant's profits is, as a practical matter, the sole remedy available to compensate injured plaintiffs . . . .").

118. Lanham Act § 35(a), 15 U.S.C. § 1117(a) ("[D]efendant must prove all elements of cost or deduction claimed.").

the similar back pocket stitching. The profits from any sales for which the defendant cannot meet this heavy burden go to the plaintiff.

Increasing reliance on the profits remedy increases the protection afforded by the infringement provision, creating a potentially very harsh doctrine. After finding infringement, the natural pressure is to award some monetary remedy to compensate. It seems unfair to say, on one hand, that the defendant's use of a mark has harmed the plaintiff in a way that trademark law recognizes and, on the other hand, withhold compensation for that harm.[119] However, as courts have increasingly divorced the likelihood-of-confusion from actual harm, the plaintiff's burden of proving actual harm to establish a damages remedy has commensurately increased. Because the increased plaintiff's burden means a damages remedy is often unavailable, courts turn to the only alternative: the often-overcompensating profits remedy.[120] The end result is a form of protection that incentivizes a new business to be *very* careful when selecting its trademark to avoid any similarity to any known, preexisting trademarks, at risk of losing its profits based on assumed harm to the trademark owner.

C. CIRCUIT SPLIT OVER THE PROFITS REMEDY

Courts generally agreed that a profits remedy was only available if the infringer acted with a culpable mental state, but the exact threshold varied.[121] Recognizing that a profits remedy often overcompensated the plaintiff, courts almost never awarded it in cases of "good-faith" or "innocent" infringement, where the infringer adopted its mark without being aware of the plaintiff's and the similarity between them occurred by chance.[122] At the other end of the spectrum, courts agreed that "willful" infringement, where an infringer purposefully imitates a plaintiff's mark hoping to create confusion, supported a profits remedy[123] because the remedy served unjust enrichment and deterrence purposes in addition to compensating the plaintiff for harm to its business interest.[124] Because the willful infringer acts deliberately, ensuring she does not profit from her infringement deters future harm to both the business

119. *See, e.g.*, Thurmon, *supra* note 6, at 248 (describing the Lanham Act's remedial scheme as "seriously flawed" in part because "most victims of infringement receive no monetary compensation").

120. *See id.* ("Those lucky few [plaintiffs] who do obtain money judgments often receive windfalls that far exceed their injury.").

121. *See* Koelemay, *supra* note 57, at 498–99 (describing the "clash" of decisions over the appropriate mental state to support a profits remedy).

122. *See id.* ("[A]n unknowing or accidental infringement warrants only injunctive relief.").

123. *Id.* at 498.

124. *See id.* at 487–94 (explaining how courts adopted unjust enrichment and deterrence rationales to justify a profits remedy when the compensation rationale alone was insufficient).

and consumer interests.[125] Similarly, because the willful infringer acts with a culpable mental state, disgorging profits prevents unjust enrichment.[126] The difficult questions arise when the infringer's mental state is somewhere between "innocent" and "willful," such as when the infringer is aware of the plaintiff's mark but thinks either that the plaintiff's mark is invalid or that there will be no confusion.[127]

Some circuits imposed a willfulness prerequisite to rein in trademark protection on equitable grounds. A willfulness prerequisite was necessary "to avoid the conceivably draconian impact [of] . . . overcompensat[ing] for a plaintiff's actual injury and creat[ing] a windfall judgment at the defendant's expense."[128] Although this prerequisite does nothing to decrease protection at the infringement stage, it does limit protection in real terms, as it often leaves plaintiffs without any monetary remedy.[129] Courts grounded this rule in the flexible language of the Lanham Act—"every award is 'subject to equitable principles' and should be determined 'according to the circumstances of the case.'"[130]

Circuits rejecting a willfulness prerequisite simply pointed out that the remedies statute lacked any mention of intent.[131] Originally, § 1117(a), the remedies provision, did not contain the clause, "or a willful violation under section 1125(c) of this title."[132] Congress added this language in 1999, explained below. Therefore, at the time of the initial circuit split, § 1117(a) specified the available remedies without any mention of the defendant's mental state. Some circuits reasoned that the omission meant that a defendant's intent

125. *See id.* at 491 ("The appellate panel noted that one purpose of the Lanham Act was to protect the public from deception, and found that if the infringer could retain his unjustly earned profits, an injunction alone would not adequately deter willful infringements . . . .").

126. *See* Romag Fasteners, Inc. v. Fossil, Inc., 140 S. Ct. 1492, 1498 (2020) (Sotomayor, J., concurring in the judgment) ("[L]ongstanding equitable principles . . . seek to deprive only wrongdoers of their gains from misconduct.").

127. Koelemay, *supra* note 57, at 498.

128. George Basch Co. v. Blue Coral, 968 F.2d 1532, 1540 (2d Cir. 1992), *superseded in relevant part by statute*, Lanham Act § 35(a), 15 U.S.C. § 1117(a).

129. *See* Thurmon, *supra* note 6, at 247 ("Trademark owners . . . seldom recover money when they win.").

130. *George Basch*, 968 F.2d at 1540–41 (quoting Lanham Act § 35(a), 15 U.S.C. § 1117(a)).

131. *See, e.g.*, Roulo v. Russ Berrie & Co., 886 F.2d 931, 941 (7th Cir. 1989) ("The Lanham Act specifically provides for the awarding of profits in the discretion of the judge subject only to principles of equity.").

132. Romag Fasteners, Inc. v. Fossil, Inc., 817 F.3d 782, 787 (Fed. Cir. 2016), *vacated in part*, 140 S. Ct. 1492 (2020).

was "simply a 'relevant factor[ ] to the court's determination of whether an award of profits is appropriate.'"[133]

Considering the defendant's mental state as only a factor instead of a strict prerequisite has an equitable appeal. Even if equitable principles should bar a profits remedy against an innocent infringer, a willfulness prerequisite goes further and also bars a profits remedy for infringers acting with mental states between innocent and willful, such as knowing of the plaintiff's mark but thinking the mark is invalid.[134] In those cases, the equities are less clear. While a profits remedy might overcompensate the plaintiff, it also seems inequitable to deny the plaintiff all compensation where Congress had chosen to label the defendant's actions infringing (assuming the plaintiff could not prove damages).[135]

The 1999 amendment to the Lanham Act made the willfulness prerequisite more difficult to maintain. Three years earlier, Congress had created a new cause of action for trademark dilution under § 1125(c).[136] However, Congress failed to also amend the remedies provision to actually provide a monetary remedy in trademark dilution cases.[137] The 1999 amendment fixed the oversight and provided that monetary remedies were available for "a willful violation under section 1125(c) of this title."[138] This language had a potentially unintended consequence. For courts simply reading the remedies provision after the amendment, the natural, negative inference was that Congress, by adding a willfulness prerequisite for trademark dilution under § 1125(c), specifically chose not to impose a willfulness prerequisite for trademark infringement under § 1125(a).[139] That inference heavily influenced the Supreme Court's *Romag Fasteners v. Fossil* opinion resolving the circuit split.

133. *Id.* at 786 (quoting Pebble Beach Co. v. Tour 18 I Ltd., 155 F.3d 526, 554 (5th Cir. 1998), *abrogated on other grounds by* TrafFix Devices, Inc. v. Mktg. Displays, Inc., 532 U.S. 23 (2001)) (alterations in original).

134. *See* Koelemay, *supra* note 57, at 498–99 (describing how some courts have found less-culpable mental states sufficient for a profits remedy).

135. *See, e.g.*, Thurmon, *supra* note 6, at 313 ("Any gain obtained through trademark infringement is unjustified, even if the infringement is unintentional.").

136. *Romag Fasteners*, 817 F.3d at 787.

137. *Id.*

138. *Id.* at 787–88.

139. *See id.* at 788 (describing the contention that with the 1999 amendment Congress "made plain that it did not intend willful infringement to be a prerequisite to monetary recovery" for violations other than trademark dilution).

III. *ROMAG FASTENERS V. FOSSIL*

The Supreme Court's decision in *Romag Fasteners v. Fossil* will likely further increase trademark protection. The issue in the case was whether the "principles of equity" demanded that courts only award a profits remedy when the defendant infringed willfully.[140] The Supreme Court held that willful infringement was not a necessary prerequisite to a profits remedy because the language of the remedies provision made it unlikely that Congress intended such a prerequisite and because, historically, courts of equity were not unanimous in requiring willfulness.[141] The decision re-emphasizes the enormous discretion that courts have to adjust trademark remedies within equitable bounds. Most likely, *Romag* will increase the availability of the profits remedy, increasing trademark protection in real terms by overcompensating plaintiffs.

This Part describes the initial facts, procedural history, and Supreme Court decision in *Romag Fasteners v. Fossil*. Specifically, it compares the majority and concurring opinions and concludes that Justice Sotomayor's concurrence was more correct. Finally, it predicts that *Romag* will likely increase trademark protection but also suggests that such an outcome is not inevitable.

A. FACTS & PROCEDURAL HISTORY

The primary parties were Romag Fasteners, Inc. ("Romag") and Fossil, Inc. ("Fossil").[142] Romag sells its patented magnetic snap fasteners under its registered ROMAG trademark.[143] The fasteners are used in various products, such as wallets, purses, and handbags.[144] Fossil designs and sells various fashion accessories, including jewelry and handbags.[145] Both parties contracted with separate business entities in China to manufacture their respective products.[146]

In 2002, Fossil contracted with Romag to use ROMAG fasteners in Fossil's products, and Fossil instructed its manufacturer in China to purchase

140. *See* Romag Fasteners, Inc. v. Fossil, Inc., 140 S. Ct. 1492, 1495 (2020) (explaining Fossil's argument that the Lanham Act required willfulness).

141. *Id.* at 1495–97.

142. *Romag Fasteners*, 817 F.3d at 783.

143. *Id.*

144. *Id.*

145. Romag Fasteners, Inc. v. Fossil, Inc., 29 F. Supp. 3d 85, 93 (D. Conn. 2014), *aff'd*, 817 F.3d 782 (Fed. Cir. 2016), *vacated in part*, 140 S. Ct. 1492 (2020) ("[Fossil's manufacturer] informed Fossil that it had purchased ROMAG snaps from a manufacturer that was 'not the authorized licensee of Romag.'").

146. *Id.* at 92.

the fasteners from Romag's manufacturer in China.[147] This arrangement continued in relative harmony for multiple years, but eventually, Fossil's manufacturer began to purchase counterfeit ROMAG fasteners.[148] In 2010, Romag sued Fossil for, *inter alia*, trademark infringement pursuant to 15 U.S.C. § 1125(a).[149]

A jury found Fossil liable for trademark infringement under 15 U.S.C. § 1125(a) and awarded a profits remedy.[150] However, the jury did not find willful infringement.[151] The district court set aside the profits award because Second Circuit precedent only allowed disgorgement of profits in cases of willful infringement.[152]

The Federal Circuit affirmed, holding that the Second Circuit had endorsed a willfulness requirement for a profits remedy after the 1999 amendment to the remedies provision.[153] Although required to follow Second Circuit precedent, the Federal Circuit went further and added four reasons why it agreed that a willfulness prerequisite was appropriate.[154] First, the legislative history indicated that Congress did not intend the 1999 amendment to change the available remedies for trademark infringement, as Congress did not even acknowledge the circuit split.[155] Second, Congress could not have ratified a consistent interpretation of whether the "principles of equity" required willfulness because there was no consistent judicial consensus.[156] Third, as the statutory provision requiring willfulness for trademark dilution was enacted separately from the rest of the statute, it could not have created a negative presumption that "willfulness is always required in dilution cases but never for infringement."[157] Finally, equitable principles did not require willfulness to award *damages*. Therefore, an explicit willfulness requirement for violations of § 1125(c) was not superfluous; it was necessary to limit *damage* awards to willful

147. *Id.*

148. *See id.* ("From 2002 through 2008, [Fossil's manufacturer] purchased tens of thousands of ROMAG snaps . . . . However, between August 2008 and the commencement of this action, [Fossil's manufacturer] purchased on a few thousand ROMAG snaps . . . .").

149. *Romag Fasteners*, 817 F.3d at 783.

150. *Romag Fasteners*, 29 F. Supp. 3d at 90.

151. *Id.* Interestingly, the district court also imposed sanctions on Romag and found that Fossil had established an equitable defense of laches. *Id.* at 112. The court based both conclusions on Romag's "unmistakable pattern of relying on the pressure point of the holiday season . . . to orchestrate a strategic advantage and improperly obtain emergency injunctive relief . . . ." *Id.* at 102, 106.

152. *Id.* at 111.

153. *Romag Fasteners*, 817 F.3d at 789.

154. *Id.* at 789–91.

155. *Id.* at 789–90.

156. *Id.* at 790.

157. *Id.*

violations, even if equitable principles already imposed the same limit on a profits remedy.[158]

The Supreme Court granted *certiorari* to resolve the circuit split.[159] At the time, five circuits had imposed a willfulness prerequisite, six circuits had held that a profits remedy was available without willful infringement, and one circuit had issued no opinion.[160]

B. JUSTICE GORSUCH FOR THE COURT

Justice Gorsuch, writing for a majority of the Court, held that willfulness is not a prerequisite for an award of profits.[161] He began with the plain language and structure of the statute.[162] He noted that the text of the remedies provision lacked an explicit precondition of willfulness for a monetary remedy for trademark infringement.[163] The Act did, however, explicitly contain such a requirement for trademark dilution.[164] Justice Gorsuch stated, "[T]his Court usually [does not] read into statutes words that aren't there," especially when the term is included elsewhere in the same provision.[165] The fact that the rest of the Lanham Act also used direct terms to describe mental state requirements, the majority concluded, made it even less likely that Congress included one implicitly.[166]

The majority did not reject outright Fossil's argument that equity could limit the profits remedy to cases of willful infringement. Even though principles of equity are usually "fundamental rules that apply more systematically across claims and practice areas"[167] and that definition seemed unlikely to encompass a rule about a single remedy confined to trademark law,[168] the majority admitted such a reading "might be possible."[169] However, even assuming that a willfulness prerequisite could amount to a principle of

158. *Id.* at 790–91.

159. Romag Fasteners, Inc. v. Fossil, Inc., 140 S. Ct. 1492, 1494 (2020).

160. *See Romag Fasteners*, 817 F.3d at 785–86, 788–89 (discussing circuit court rulings both before and after the 1999 amendment to the Lanham Act remedies provision).

161. *Romag Fasteners*, 140 S. Ct. at 1497.

162. *Id.* at 1495.

163. *Id.*

164. *Id.*

165. *Id.*

166. *Id.*

167. *Id.*

168. *Id.*

169. *Id.* at 1496.

equity, whether trademark historically required such a showing was a separate question.[170]

The Court ultimately concluded that the case history was not clear enough to require willfulness.[171] Here, the majority's analysis stumbled. The opinion cited three cases that, it said, "expressly rejected any such rule."[172] However, of the three, only *Oakes v. Tonsmierre* actually authorized a profits award for unintentional infringement.[173] The statements in the other two cases were dicta because both cases involved willful infringement.[174] Further, the *Oakes* court awarded the profits under a property theory that, as mentioned previously, the Supreme Court later rejected in *Hanover Star Milling Co. v. Metcalf*.[175] Despite this slight distortion, the majority correctly recognized that courts historically were very confused as to what mental state justified a profits remedy and, therefore, did not reach a clear consensus.[176] Ultimately, this led the majority to reject any strict prerequisite; it held that while a defendant's mental state is "highly important," willfulness is not required to award a defendant's profits.[177]

C. JUSTICE SOTOMAYOR, CONCURRING IN THE JUDGMENT

While Justice Sotomayor agreed that a profits remedy lacked a willfulness prerequisite, she would have denied the remedy in cases of innocent infringement.[178] She correctly characterized the equitable case law as using the "willfulness" label for "a range of *culpable* mental states."[179] Justice Sotomayor's reasoning was somewhat ambiguous on whether profits should be available

---

170. *See id.* ("But even if we were to spot Fossil that first essential premise of its argument, the next has problems too.").

171. *See id.* ("[I]t's far from clear whether trademark law historically required a showing of willfulness before allowing a profits remedy.").

172. *Id.*

173. *See* Oakes v. Tonsmierre, 49 F. 447, 453 (C.C.S.D. Ala. 1883) ("[T]he case may go to a master for an account of gains and profits, on account of the unauthorized, though not intentional and fraudulent, use by respondents of the name of Peter Oakes.").

174. *See* Stonebraker v. Stonebraker, 33 Md. 252, 268–69 (1870) (describing the defendant's trademark use as "well calculated to produce, upon the public mind, the impression" that the defendant's goods were those of the plaintiff); Lawrence-Williams Co. v. Societe Enfants Gombault et Cie, 52 F.2d 774, 778 (6th Cir. 1931) ("[T]he case does not sufficiently show that the infringement was not willful.").

175. *See* Hanover Star Milling Co. v. Metcalf, 240 U.S. 403, 414 (1916) ("[T]he trade-mark is treated as merely a protection for the good-will, and not the subject of property except in connection with an existing business."), *superseded by statute*, Lanham Act § 22, 15 U.S.C. § 1072, *as recognized in* Park 'N Fly, Inc. v. Dollar Park and Fly, Inc., 469 U.S. 189 (1985).

176. *See* Koelemay, *supra* note 57, at 498–99 (describing the "clash" of decisions over the appropriate mental state to support a profits remedy).

177. *Romag Fasteners*, 140 S. Ct. at 1497.

178. *Id.* at 1498 (Sotomayor, J., concurring in the judgment).

179. *Id.*; *see* Koelemay, *supra* note 57, at 498–99.

for negligent infringement, but her conclusion clarified that it should be. She first stated that historical definition of willfulness excluded negligence, but then confined the rest of her analysis to "innocent or good-faith trademark infringement."[180] Her conclusion, however, was clear: "[A] district court's award of profits for innocent or good-faith trademark infringement would not be consonant with the 'principles of equity' . . . ."[181] The implication by omission was that the principles of equity would permit an award of profits for negligent infringement.

Justice Sotomayor was correct on all counts. Historically, courts of equity were near unanimous in denying a profits remedy in cases of innocent trademark infringement.[182] Justice Sotomayor seemed to primarily disagree with the majority on how the principles of equity affect the case analysis. The majority analyzed whether a willfulness prerequisite could itself be a principle of equity.[183] Justice Sotomayor recognized that equitable principles could dictate one rule for trademark law and different ones for "patent infringement cases and other arguably analogous suits,"[184] even if the rules themselves were not principles of equity.[185] Trademark law, as part of unfair competition law, is meant to protect trademarks in a way that promotes competition.[186] Patent law, in contrast, protects patents through explicit constitutional power to reduce competition.[187] Far from being analogous, trademark and patent law are in some ways opposites. Therefore, equitable principles should affect them differently.

D. RESULT: BROAD PROTECTION, FEW RULES, AND GREAT DISCRETION

The *Romag* decision will most likely further increase trademark protection because willfulness is now not required in any circuit. Further, although some

---

180. *Romag*, 140 S. Ct. at 1498 (Sotomayor, J., concurring in the judgment).

181. *Id.*

182. *See* Koelemay, *supra* note 57, at 498–99 ("Courts have generally agreed that . . . an unknowing or accidental infringement warrants only injunctive relief.").

183. *See Romag*, 140 S. Ct. at 1495 (majority opinion) ("But, Fossil says, trademark is different. There alone, a willfulness requirement was so long and universally recognized that today it rises to the level of a 'principle of equity' the Lanham Act carries forward.").

184. *Id.*

185. *See id.* at 1498 (Sotomayor, J., concurring in the judgment) ("Nor would [disgorging an innocent infringer's profits] seem to be consistent with longstanding equitable principles which, after all, seek to deprive only wrongdoers of their gains from misconduct.").

186. *See* Glynn S. Lunney Jr., *Trademark Monopolies*, 48 EMORY L.J. 367, 370–71 (1999) (describing the need for balance in trademark law because "trademark protection can both advance and disserve the development of an efficient and desirably competitive market").

187. *See* U.S. CONST. art. I, § 8, cl. 8 ("The Congress shall have Power . . . To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries . . . .").

courts may agree with Justice Sotomayor and continue to deny a profits remedy in innocent infringement cases, the profits remedy is now unambiguously available in cases where the defendant's mental state is somewhere between innocent and willful. For example, a defendant that adopted a trademark knowing of a plaintiff's prior use but believing the plaintiff's trademark to be invalid now faces an increased risk of profits disgorgement if the plaintiff's trademark is ultimately valid. Other courts may go further still. The Lanham Act does not contain any express mental state requirement, and Justice Sotomayor's concurrence could prove counterproductive to her goal of preventing a profits award against innocent infringers. In cases of innocent infringement, a plaintiff can legitimately argue that it would be inequitable to deny a profits remedy where damages are unavailable because doing so would leave the plaintiff uncompensated for harms that trademark law protects against.[188] And after *Romag*, the plaintiff could use Justice Sotomayor's statement that "the majority is agnostic about awarding profits for both 'willful' and innocent infringement"[189] for the further argument that the *Romag* majority believed that a profits remedy could be appropriate in such cases. The argument would be that the majority could have agreed with Justice Sotomayor and held that a profits remedy was never available in innocent infringement case. It declined to do so. Therefore, through its silence, the profits remedy must be available in at least some innocent infringement cases.

However, *Romag* also provides courts with a new opportunity to rein in trademark protection. The key is to align the majority's definition of an equitable principle with the purpose of trademark law. As the majority stated, "in the context of this statute, [the phrase 'principles of equity'] more naturally suggests fundamental rules that apply more systematically across claims and practice areas."[190] Trademark law's purpose is to protect trademarks in a way that promotes competition.[191] It would be difficult to find a more fundamental rule in any area of law relating to business than promoting robust competition.[192] As the next Part argues, courts should, absent congressional action, use the fundamental rule of promoting competition and their equitable discretion to adopt a new limit on trademark remedies that focuses on the

188. *See, e.g.*, Thurmon, *supra* note 6, at 313.

189. *Romag Fasteners*, 140 S. Ct. at 1498 (Sotomayor, J., concurring in the judgment).

190. *Id.* at 1496 (majority opinion).

191. *See* Lunney, *supra* note 186, at 370–71 (describing the need for balance in trademark law because "trademark protection can both advance and disserve the development of an efficient and desirably competitive market").

192. *See, e.g.*, ALTMAN & POLLACK, *supra* note 25, § 4:1; Eastman Co. v. Reichenbach, 20 N.Y.S. 110, 116 (N.Y. Sup. Ct. 1892), *aff'd sub nom.* Eastman Kodak Co. v. Reighenbach, 29 N.Y.S. 1143 (N.Y. Gen. Term 1894).

impact of the infringement on consumers to rein in trademark protection in practical terms.

IV. FIXING TRADEMARK OVERPROTECTION

Courts can and should rein in trademark overprotection using the equitable discretion provided by the Lanham Act and re-emphasized in *Romag Fasteners v. Fossil*. Trademark law and unfair competition law are linked because "[t]rademark law is part of the branch of tort law known as unfair competition."[193] Unfair competition law is meant to promote competition, but current theories lead to expansive protection that stifle it. Courts can limit protection by adopting a new theory of unfair competition law—the Consumer Approach—focused on how business practices affect consumers. For trademark law, while Congress would ideally amend the infringement provision of the Lanham Act to align with the Consumer Approach, courts can fully implement it within the current language of the remedies statute.

This Part first introduces the Consumer Approach and supports it by explaining how the three current alternative theories each lead to unlimited protection. Then, it outlines how courts could implement the Consumer Approach under current trademark law using remedies and without increasing administrative costs.

A. DEFINING UNFAIR COMPETITION

> "One has only to read a few judicial opinions to recognize the semantic quagmire that awaits the lawyer or judge who attempts to give a precise definition to that evasive concept called 'unfair competition.'"[194]

Despite past difficulties, the law should define an act of unfair competition as a business practice that is anticompetitive, wherein anticompetitive means that the business practice harms consumers more than it benefits them. Consumer benefit should be defined in the traditional sense: lower prices and higher quality goods.[195] The law should adopt these definitions because the only alternatives lead to unfair competition laws providing protection with no theoretical limit. Because unfair competition laws prohibit certain business practices, unlimited protection erases the line between unfair and legitimate competition.

---

193. Landes & Posner, *supra* note 22, at 267.

194. MCCARTHY, *supra* note 2, § 1.9.

195. *See, e.g.*, N. Pac. Ry. Co. v. United States, 356 U.S. 1, 4 (1958) ("[T]he unrestrained interaction of competitive forces will yield the best allocation of our economic resources, the lowest prices, the highest quality and the greatest material progress . . . .").

The first alternative is to define unfair competition as "conduct that 'shocks the judicial sensibilities.'"[196] As judicial sensibilities are based on current norms, the fact that this test provides no real limit to protection is almost self-evident. Different judges will be shocked in different ways, and current norms will evolve.[197] For instance, the common law once limited trademark infringement to cases involving direct competitors.[198] When that level of protection no longer seemed adequate, courts expanded it.[199]

The second alternative is to limit unfair competition law by confining it to certain categories, such as "passing off one's goods as those of another."[200] However, this alternative provides no theory to explain why only those categories are appropriate. It at least recognizes that some limit on protection is needed, but it provides no mechanism to limit protection within the categories that already exist. And the categories themselves are expandable. The clearest example of both the lack of a limiting mechanism and the categorical expandability is the enlargement of trademark protection to cover trademark dilution. Dilution allows "holders of famous marks [to] sue junior users 'regardless of the presence or absence of actual or likely confusion, of competition, or of actual economic injury.'"[201]

The third alternative is to define an act of unfair competition as a business practice that is anticompetitive, wherein anticompetitive means that the business practice harms existing businesses more than it benefits them. The theory is that "trademark law, like tort law in general (and trademark law is part of the branch of tort law known as unfair competition), can best be explained on the hypothesis that the law is trying to promote economic efficiency."[202] Economic efficiency here means overall wealth maximization.[203] Therefore, the definition of an act of unfair competition becomes any act that decreases overall wealth.

The law should reject the wealth maximation theory because it too results in unlimited protection. If an unfair business practice is one that decreases

---

196. MCCARTHY, *supra* note 2, § 1.9.

197. *See id.* ("[P]recedent that defines unfair competition as conduct that 'shocks the judicial sensibilities' gives no guidance either to a businessperson or a judge.").

198. *See* McKenna, *supra* note 50, at 74.

199. *See* Koelemay, *supra* note 57, at 488 ("The advent of nation-wide advertising and marketing altered the commercial realities on which [the direct competition] requirement was based, however, and courts began expanding the rights of trademark owners accordingly.").

200. MCCARTHY, *supra* note 2, § 1.9.

201. Desai, *supra* note 15, at 605 (citation omitted).

202. Landes & Posner, *supra* note 22, at 267.

203. *See* Desai, *supra* note 15, at 601 ("Modern trademark law operates as a subset of the Chicago School's approach to firms and competition, and in that approach trademark law's true goal is to enable the firm to maximize wealth.").

overall wealth, then fair business practices must be those that increase overall wealth. Theoretically then, the law should not prevent a defendant's activity that increases overall wealth, even if it falls into one of the traditional categories of unfair competition and causes the plaintiff some harm. Unfortunately, that theoretical possibility can never occur in any individual case because courts are not omniscient. Courts simply cannot predict how any individual defendant's activity will affect overall wealth. The wealth maximization theory recognizes that and resolves the issue through a generalizing assumption: firms are best placed to maximize wealth.[204] Therefore, any act that harms the plaintiff's firm value becomes an act of unfair competition. Given that all new firms harm existing firms in the same market in the sense that the new firms increase supply amid consistent demand and therefore reduce price and profits, a full application of this theory would support preventing all new firms from forming. Because that result is extreme, at least according to current norms, this theory simply collapses back into the second alternative, providing theoretically unlimited protection within existing, though expandable, categories of unfair competition.

Instead of maximizing wealth, unfair competition law should limit protection by focusing on consumers directly. It can do this through an economic analysis comparing the consumer benefit and harm of the defendant's activity. Under this theory, consumer benefit limits the harm the law recognizes. To use a trademark example, assume that the defendant's trademark use causes some consumer confusion, however defined. That consumer confusion causes some harm to consumers by causing them to overpay.[205] However, the defendant's use also causes some consumer benefit by increasing competition in the market, which lowers prices and increases quality.[206] If the benefits created are greater than the harm caused, then the defendant's use is a fair business activity and not an infringement. Fully enacting this change into trademark law would require Congress to amend § 1125. However, the next Section provides an example of how courts can implement this Consumer Approach under current trademark law using remedies.

B. THE CONSUMER APPROACH TO LIMITING TRADEMARK PROTECTION

Under the Consumer Approach, courts adopt a consumer harm-benefit analysis when determining remedies. After determining infringement, courts

204. *See id.* ("Under [the Chicago School] view, firms are better placed to manage resources and increase wealth . . . .").

205. *See* McKenna, *supra* note 5, at 1844.

206. *See, e.g.*, N. Pac. Ry. Co. v. United States, 356 U.S. 1, 4 (1958).

analyze the consumer harm and consumer benefit of the defendant's trademark use. If harm outweighs benefit, then the difference becomes the amount of the monetary award. If benefit outweighs harm, then the defendant's use is a fair business activity, and the court grants no award.

The key justification for the Consumer Approach is the Supreme Court's definition of a principle of equity in *Romag Fasteners v. Fossil*.[207] According to the majority, "[i]n the context of [§ 1117, the phrase 'principles of equity'] more naturally suggests fundamental rules that apply more systematically across claims and practice areas."[208] A fundamental rule that applies systematically at least across all areas of business law is to promote competition.[209] By recognizing the promotion of competition as a principle of equity, courts can legitimately adjust trademark remedies in service of that goal within the considerable discretion the Lanham Act currently provides.

As an example of how the Consumer Approach would operate in practice, consider *Romag* with some altered facts. Romag is still the plaintiff using the ROMAG mark for magnetic snap fasteners. However, assume a different defendant starts using the ROMAG mark to sell zippers. Magnetic snap fasteners and zippers are in different markets—no one specifically looking for a magnetic snap fastener is going to mistakenly buy a zipper.[210] However, some overlap exists: a handbag manufacturer like Fossil could use either as a sealing mechanism for its products.[211] The overlap is likely enough for a court to find a likelihood of consumer confusion and therefore a violation of § 1125(a), especially if the defendant used the ROMAG mark willfully, hoping to benefit from Romag's goodwill.[212]

---

207. Romag Fasteners, Inc. v. Fossil, Inc., 140 S. Ct. 1492 (2020).

208. *Id.* at 1496.

209. *See, e.g.*, ALTMAN & POLLACK, *supra* note 25, § 4:1; Eastman Co. v. Reichenbach, 20 N.Y.S. 110, 116 (N.Y. Sup. Ct. 1892), *aff'd sub nom.* Eastman Kodak Co. v. Reighenbach, 29 N.Y.S. 1143 (N.Y. Gen. Term 1894).

210. *See* AMF Inc. v. Sleekcraft Boats, 599 F.2d 341, 348 (9th Cir. 1979) (agreeing that two boat lines were not competitive because they "appeal[ed] to separate sub-markets" within the boating community).

211. *See id.* at 348 n.10 ("Related goods are those 'products which would be reasonably thought by the buying public to come from the same source if sold under the same mark.'" (citation omitted)).

212. *See* Beebe, *supra* note 73, at 1607 ("[T]wo findings—that the similarity of the marks factor favors a likelihood of confusion and the defendant's intent factor also favors a likelihood of confusion—are together sufficient to trigger an overall finding of a likelihood of confusion . . . regardless of the outcomes of any other factors."); *see also* Fleischmann Distilling Corp. v. Maier Brewing Co., 314 F.2d 149, 155 (9th Cir. 1963) (finding a likelihood of confusion where scotch and beer sellers used the same mark).

The first step at the remedy stage is for the court to examine any harm the defendant caused. Here, the court would compare the defendant's zipper pricing and sales to those of the average zipper market entrant. If the defendant's pricing and sales were lower or similar, the court could conclude that the likely consumer confusion had not ripened into actual consumer confusion and that no harm had occurred, either to consumers or to Romag's goodwill. If the defendant's pricing and sales were higher, the court would need to assess any possible reasons for the increase independent of the likely consumer confusion. If no reasons existed, or if they only provided a partial explanation, then the remaining difference would become the measure of consumer harm. The harm would be calculated as:

> (consumer price paid – consumer price without confusion) * (quantity of zippers sold by the defendant)[213]

The second step is for the court to examine any benefit the defendant caused. Here, the court would examine the market before and after the defendant's entrance. Presumably, the defendant's entrance provides additional supply and decreases the price of zippers overall. The benefit would be calculated as:

> (average market price before defendant – average market price after defendant) * (quantity of zippers sold by all sellers)

The final step is for the court to compare benefit and harm. Because harm and benefit are directly comparable, this is a simple subtraction. There is still some misalignment between consumers harmed and consumers benefitted because those harmed are the defendant's customers while those benefitted are all customers in the overall market. However, this misalignment is likely insignificant compared to the larger imprecisions caused by the administrative difficulties of measuring benefit and harm.

To handle the administrative difficulties inherent in measuring harm and benefit, courts should use the burden shift between the damages and profits remedies that the Lanham Act already provides. This Note has already established both the difficulties of measuring business harm from trademark infringement and the Lanham Act's solution of placing the burden on the

---

213. This is simplified. Dennis Corgill has proposed a more sophisticated method of measuring the ways in which a junior user could benefit from decreased market entry costs over the course of the infringing product's life cycle. *See generally* Dennis S. Corgill, *Measuring the Gains of Trademark Infringement*, 65 FORDHAM L. REV. 1909 (1997) (arguing for a monetary remedy based on the difference between an infringer's gross revenue over the entire product lifecycle and what that gross revenue would have been without the benefits from the infringement).

plaintiff to prove damages and on the defendant to prove deductions from profits.[214] The Consumer Approach likely imposes even greater administrative difficulties because a court must determine both consumer harm and consumer benefit, but the solution to those difficulties is a similar burden shift.

The Consumer Approach requires four determinations:

1. That the defendant's sales and/or prices were higher than those of the normal market entrant;
2. The defendant's mental state;
3. Any reasons for the defendant's higher sales and/or prices independent from the consumer confusion; and
4. A comparison of the consumer harm and benefit caused by the defendant's activities.

The plaintiff always bears the burden of showing (1) and (2), but (2) is already part of the likelihood-of-confusion test at the infringement stage.[215] If a damages remedy is appropriate, then the plaintiff must also show (3) that the defendant's higher sales and/or prices were due to the infringement, and (4) that the defendant's activities caused more consumer harm than benefit. On the other hand, if the profits remedy is appropriate, then the defendant must show (3) any sales that were not due to the infringement, and (4) that the defendant's activities caused more consumer benefit than harm. Because the measurements of consumer benefit and harm derive from economic analysis, courts should require proof based on a reasonable, not absolute, certainty.[216]

Courts should follow *Romag* and use the defendant's mental state as a "highly important consideration"[217] in determining the appropriate remedy. Specifically, courts should follow Justice Sotomayor's concurring opinion and award the damages remedy in cases of innocent infringement and the profits remedy when the defendant is at least negligent.[218] Because a defendant can only act with one mental state, this means the two remedies are mutually exclusive. A plaintiff can pray for both, but only one will still be available after the defendant's mental state has been determined. Because of the inherent administrative difficulties, the identity of the party bearing the burden is likely

214. *See supra* Section II.B.2.

215. *See* AMF Inc. v. Sleekcraft Boats, 599 F.2d 341, 349 (9th Cir. 1979) (identifying "defendant's intent in selecting the mark" as one factor in the likelihood of confusion test).

216. *See* Thurmon, *supra* note 26, at 166 (arguing that courts should adopt a "more flexible, pragmatic approach" to measuring trademark damages).

217. Romag Fasteners, Inc. v. Fossil, Inc., 140 S. Ct. 1492, 1497 (2020).

218. *See id.* at 1498 (Sotomayor, J., concurring in the judgment) ("[A] district court's award of profits for innocent or good-faith trademark infringement would not be consonant with the 'principle of equity' . . . .").

to greatly affect the amount of the award.[219] A defendant's wrongful mental state justifies creating a larger award by placing the burden on the defendant because the award further benefits consumers by deterring future infringers.[220]

The Consumer Approach aligns with the requirements of the Lanham Act. Section 1117(a) contains two commands relevant to the Consumer Approach's four determinations. First, § 1117(a) provides that "[i]n assessing profits the plaintiff shall be required to prove defendant's sales only; defendant must prove all elements of cost or deduction claimed."[221] In comparison, for a profits remedy under the Consumer Approach, the plaintiff has the burden to show (1) and (2), and the defendant must show (3) and (4). Because the defendant's mental state is part of the likelihood-of-confusion test for infringement, the plaintiff will in fact only need to prove (1) at the remedy stage while the defendant will still need to prove (3) and (4), exactly as the Lanham Act already commands. Second, the Lanham Act requires that the amount awarded "shall constitute compensation and not a penalty."[222] Under the Consumer Approach, courts can still characterize any award as compensation for harm to the business interest caused by the consumer confusion. That business harm, whether it results directly through trade diversion or indirectly through reputational injury, still exists. The only difference from current practice is that the Consumer Approach imposes a new limit on the compensation based on the equitable principle of promoting competition, which also aligns with the Lanham Act's command that courts award remedies "subject to the principles of equity."[223]

In addition to providing a theory that allows courts to limit trademark protection, the Consumer Approach provides two further benefits. First, it should make the damages remedy more available. Because business harm is often abstract and current practice imposes a high burden of proof on the plaintiff, courts rarely award damages.[224] The Consumer Approach, in contrast, would require the plaintiff to measure consumer harm and benefit only to a reasonable certainty. This lower burden means that a plaintiff will likely be able

219. *See* Thurmon, *supra* note 26, at 138 (noting that under the current Lanham Act burden shift "[p]revailing trademark owners usually get no monetary relief, but those few who do sometimes hit the jackpot").

220. *See* Koelemay, *supra* note 57, at 491 ("[I]f the infringer could retain his injustly earned profits, an injunction alone would not adequately deter willful infringements . . . .").

221. Lanham Act § 35(a), 15 U.S.C. § 1117(a).

222. *Id.*

223. *Id.*

224. *See* Thurmon, *supra* note 26, at 150 ("It is quite difficult to quantify the economic harm caused by trademark infringement or dilution, and for that reason, few prevailing trademark owners obtain actual damages.").

to recover damages where the infringement, though innocent, causes high levels of consumer confusion. Second, the amount awarded should naturally adjust under the Consumer Approach based on the defendant's mental state. This adjustment reflects the assumption that a negligent infringer will usually cause less harm than an intentional infringer. The negligent infringer's wrongful behavior, by definition, is only the failure to do adequate diligence.[225] The negligent infringer, presumably, fully intends to run a legitimate business and had no expectation of benefit from consumer confusion. The trademark pirate, on the other hand, presumably intends to benefit as much as possible from any consumer confusion. A reasonable assumption is that, on average, the trademark pirate will cause more consumer harm and provide less consumer benefit than the negligent infringer in otherwise-equivalent circumstances. Therefore, although the Consumer Approach grants a profits remedy in each case, the amount awarded against the trademark pirate will likely be greater.

V. CONCLUSION

The Supreme Court's decision in *Romag Fasteners v. Fossil* provides courts with the definition of a "principle of equity" as used by the Lanham Act.[226] With that definition and a new theory for determining which business practices constitute unfair competition, courts should adopt the Consumer Approach and use trademark remedies to end the modern overprotection of trademarks.

The path leading to the current trademark overprotection is clear. First, the dominant theory of unfair competition strives towards a goal of overall wealth maximization and considers existing businesses (trademark owners) as best placed to maximize overall wealth. This theory enables trademark owners to successfully push the broad language of the Lanham Act to create new types of actionable confusion. However, because the harms from these new types of actionable confusion are more difficult to prove, the damages remedy is often unavailable. This increased the importance of the profits remedy. But the profits remedy often overcompensates the plaintiff, so some courts adopted a willfulness prerequisite to limit it. *Romag* then eliminates the willfulness prerequisite, becoming the latest judicial decision to align with the wealth maximization theory.

---

225. *See* Restatement (Second) of Torts § 282 (Am. L. Inst. 1965) ("[N]egligence is conduct which falls below the standard established by law for the protection of others against unreasonable risk of harm.").

226. *See* Romag Fasteners, Inc. v. Fossil, Inc., 140 S. Ct. 1492, 1496 (2020) ("[I]n the context of this statute, [the phrase 'principles of equity'] more naturally suggests fundamental rules that apply more systematically across claims and practice areas.").

The Consumer Approach shifts the focus of unfair competition law from businesses to consumers. Maximizing overall wealth should not be the goal of unfair competition law because that goal is unmeasurable. Lacking a crystal ball, courts have no way to determine in many cases whether the defendant's business activity that the plaintiff alleges is unfair will or will not increase overall wealth if allowed to continue. Instead, courts should compare any harm and benefit that the defendant's activity has already caused. Performing this comparison requires using consumer, not business, harm and benefit because competition from new businesses will inevitably harm existing businesses. If business harm is the metric, then all competition will be unfair. The Consumer Approach empowers courts to preserve a zone of legitimate competition by permitting business activities that benefit consumers.